DRIFT: DIRECTED DEVELOPMENTAL GEOMETRY FOR SPATIOTEMPORAL TRANSCRIPTOMIC ALIGNMENT

Anonymous authors
Paper under double-blind review

ABSTRACT

Aligning unpaired spatial transcriptomic snapshots across developmental time is essential for reconstructing tissue dynamics, yet existing approaches primarily rely on transcriptomic similarity and spatial organization, which may yield developmentally implausible correspondences. We introduce DRIFT, a framework that incorporates directed developmental geometry into spatiotemporal transcriptomic alignment. DRIFT infers cell-type developmental reachability from velocity-informed single-cell dynamics and lifts this geometry to heterogeneous spatial spots through composition-aware optimal transport. The resulting developmental distance is integrated with transcriptomic and spatial information in a unified transport objective for cross-time alignment. Across four human spatiotemporal transcriptomics datasets, DRIFT produces correspondences that are more consistent with reference lineage structures and exhibit lower developmental deviation than representative transport-based baselines. These improved alignments also support more accurate held-out stage reconstruction under both linear transport and flow matching, with strong performance in both interpolation and extrapolation settings. Our code is publicly available at https://github.com/xxx/DRIFT.

1 INTRODUCTION

Inferring how cellular populations and tissue organization evolve over time is central to deciphering developmental programs (Liberali & Schier, 2024). Temporal spatial transcriptomics (ST) captures molecular and spatial snapshots across developmental stages, revealing changes in tissue composition and organization (Zhou et al., 2023; Liu et al., 2024). Because transcriptomic profiling is destructive, different time points typically sample distinct tissue sections. These snapshots are therefore unaligned, leaving cross-time correspondences between spatial states unobserved (Schiebinger et al., 2019; Rathod et al., 2025). Establishing reliable alignment across developmental stages is therefore essential for tracing how spatial tissue states evolve over time.

Methods for spatiotemporal trajectory inference from spatial transcriptomics can be broadly grouped into two alignment paradigms. *Transport-based methods*, such as Spateo (Qiu et al., 2022), spaTrack (Shen et al., 2025), and MOSCOT (Klein et al., 2025) infer cross-time couplings by jointly considering transcriptomic similarity and spatial structure. In contrast, *representation-based methods*, including STAligner (Zhou et al., 2023) and SLAT (Xia et al., 2023), learn shared latent embeddings that integrate molecular and spatial information for cross-snapshot matching. Despite their different formulations, both paradigms rely on transcriptomic similarity and may align states that appear similar but are not developmentally reachable (Schiebinger et al., 2019; Bryan et al., 2026).

Specifically, as illustrated in Fig. 1(a), similarity-based alignment matches spots mainly by their aggregated expression profiles, although each spot contains a mixture of cell types (Cable et al., 2022). Consequently, spots with distinct cellular compositions may appear similar in expression space. For example, in the developing brain, an early spot enriched for **radial glia** (□) and **intermediate progenitors** (□) can plausibly align with a later **neuron-rich spot** (■), whereas a **microglia-rich spot** (□) with similar aggregate expression should not align with the same **neuron-rich state** (△) (Demmings et al., 2025). Thus, transcriptomic similarity can yield correspondences that are close in expression space yet incompatible with developmental lineage, motivating explicit modeling of developmental transitions between the cellular populations underlying each spot.

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.

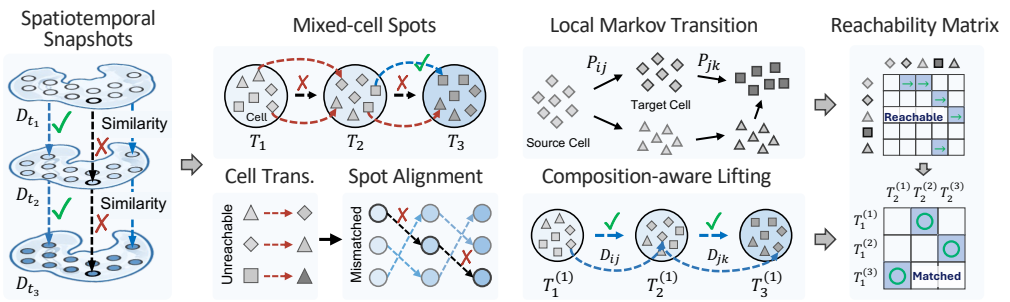


Figure 1: **Comparison of temporal spatial transcriptomics alignment strategies.** (a) Similarity-based alignment may produce developmentally incompatible correspondences. (b) DRIFT uses directed reachability and composition-aware transport for cross-time alignment. ✗ denotes a mismatched spot correspondence, and ⇢ a developmentally unreachable cell-type transition.

To address these limitations, we propose **DRIFT** (**D**irected **R**eachability for **I**nferring temporal **T**ransitions), which incorporates developmental reachability into cross-time spatial alignment, as illustrated in Fig. 1(b). The key idea is to define developmental compatibility by whether one cellular state can be reached from another along the underlying developmental flow, rather than by transcriptomic similarity alone. DRIFT first uses RNA velocity to infer local directional transitions between single-cell states, and aggregates these transitions into a directed developmental reachability matrix. It then lifts cell-level reachability to heterogeneous spatial spots through composition-aware optimal transport, yielding a spot-level developmental distance. Finally, this distance is integrated with transcriptomic similarity and spatial structure to infer cross-time correspondences that are both spatially coherent and developmentally plausible. In summary, the contributions of this paper are four folds:

- We propose a reachability-aware framework for spatiotemporal transcriptomic alignment that models cross-time correspondence through directed developmental geometry, moving beyond conventional alignment based primarily on transcriptomic and spatial similarity.
- We develop a velocity-induced developmental reachability strategy that transforms local single-cell dynamics into global, directed developmental relationships through a Markov process and first-hitting probabilities over multi-step developmental paths.
- We introduce a composition-aware developmental distance that lifts cell-type reachability to heterogeneous spatial spots through optimal transport over cell-type compositions, and integrates it with transcriptomic similarity and spatial structure for cross-time alignment.
- We conduct comprehensive experiments across four spatiotemporal transcriptomic datasets, demonstrating improved recovery of biologically plausible cross-time correspondences and developmental directionality compared with representative similarity-based alignment methods.

2 Related Work

Cellular Dynamics from Snapshots A broad class of methods reconstructs cellular dynamics from destructive single-cell measurements. Classical trajectory inference recovers developmental orderings or branching structures from static expression manifolds (Haghverdi et al., 2016; Qiu et al., 2017; Street et al., 2018; Wolf et al., 2019; Setty et al., 2019). RNA velocity (La Manno et al., 2018; Bergen et al., 2020) instead estimates local transcriptional motion from spliced and unspliced counts, while CellRank (Lange et al., 2022) combines velocity with transcriptomic similarity in a Markov process to infer fate probabilities. Continuous generative approaches include TrajectoryNet (Tong et al., 2020), which connects neural ODEs with dynamic optimal transport, PRESCIENT (Yeo et al., 2021), which learns stochastic differentiation dynamics, OT-CFM (Tong et al., 2023), which uses optimal-transport couplings for conditional flow matching, and Metric Flow Matching (Kapuśniak et al., 2024), which models flows along the data manifold. These methods improve dynamic modeling, but cross-snapshot correspondence remains a central modeling choice.

Spatiotemporal Alignment Waddington-OT (Schiebinger et al., 2019) introduced population-level optimal transport for inferring ancestor descendant relationships across temporal scRNA-seq

snapshots. For spatial transcriptomics, PASTE (Zeira et al., 2022) and PASTE2 (Liu et al., 2023) use fused and partial fused Gromov-Wasserstein formulations, while MOSCOT (Klein et al., 2025) provides scalable transport across temporal, spatial, and multimodal settings. Other spatial alignment approaches include STAligner (Zhou et al., 2023), SLAT (Xia et al., 2023), GPSA (Jones et al., 2023), and SpaTrack (Shen et al., 2025). Recent spatiotemporal methods further model developmental structure. DeST-OT (Halmos et al., 2025) accounts for growth and death using semi-relaxed OT, stVCR (Peng et al., 2026) jointly models differentiation, proliferation, migration, and spatial alignment, STORIES (Huizing et al., 2026) learns spatially informed differentiation potentials with FGW, and SOCS (Bryan et al., 2026) preserves contiguous tissue structures across time.

Biology-informed Correspondence A closely related line of work explicitly incorporates biological information into cross-snapshot correspondence. ContextFlow (Rathod et al., 2025) encodes local tissue organization and ligand–receptor communication into a transition-plausibility matrix that regularizes OT for flow matching. CellBRIDGE (Estévez et al., 2026) augments feature-based OT with directed interaction costs derived from ligand–receptor activity. Complementary methods further incorporate lineage or population dynamics. LineageOT (Forrow & Schiebinger, 2021) integrates lineage tracing with optimal transport, CoSpar (Wang et al., 2022) combines lineage and transcriptomic information, StationaryOT (Zhang et al., 2021) models growth-aware trajectories in steady-state systems, and TIGON (Tong et al., 2023) jointly reconstructs trajectories and population growth with dynamic unbalanced OT to improve biological consistency.

Our work instead defines a directed developmental distance between heterogeneous spatial spots by inferring developmental geometry from single-cell dynamics and lifting it through mixed cell-type compositions. The resulting distance measures whether one spatial state can plausibly develop into another, rather than merely whether they are similar.

3 Preliminaries and Problem Setup

RNA Velocity Modeling RNA velocity provides a local estimate of transcriptional dynamics from spliced and unspliced RNA abundances. For gene g, the standard transcription–splicing model is

$$\frac{du_g}{dt} = \alpha_g - \beta_g u_g, \qquad \frac{ds_g}{dt} = \beta_g u_g - \gamma_g s_g, \tag{1}$$

where u_g and s_g denote unspliced and spliced abundances, and α_g, β_g, and γ_g are the transcription, splicing, and degradation rates, respectively (La Manno et al., 2018; Bergen et al., 2020). Accordingly, the gene-wise velocity of cell i is estimated as

$$\hat{v}_{ig} = \hat{\beta}_g u_{ig} - \hat{\gamma}_g s_{ig}, \tag{2}$$

where $\hat{v}_{ig}$ represents the estimated transcriptional velocity of gene g in cell i. Stacking velocities across genes yields a transcriptional velocity vector, which can be projected into the same low-dimensional space as the cell state. The resulting cell state and velocity are denoted by $\mathbf{z}_i, \mathbf{v}_i \in \mathbb{R}^d$, respectively, representing the current transcriptional state and its local direction of change.

Spatiotemporal Trajectory Inference We consider spatial transcriptomic snapshots collected at ordered developmental stages $t_1 < \cdots < t_T$. At stage t_m, the tissue is represented as

$$\mathcal{D}_m = \{(x_i^{(m)}, s_i^{(m)}, \theta_i^{(m)})\}_{i=1}^{n_m}, \tag{3}$$

where $x_i^{(m)} \in \mathbb{R}^d$ denotes the transcriptomic profile of spot i, $s_i^{(m)} \in \mathbb{R}^2$ its spatial coordinate, and $\theta_i^{(m)} \in \Delta^{K-1}$ its deconvolved cell-type composition over K cell types (Cable et al., 2022). Because spatial transcriptomic profiling is destructive, measurements from different stages are obtained from distinct cells or tissue sections, leaving cross-time correspondences unobserved (Schiebinger et al., 2019; Bryan et al., 2026). For consecutive stages t_m and t_{m+1}, we infer a probabilistic coupling

$$\Gamma^{(m,m+1)} \in \mathbb{R}_+^{n_m \times n_{m+1}}, \tag{4}$$

where $\Gamma_{ij}^{(m,m+1)}$ quantifies developmental correspondence from source spot i to target spot j.

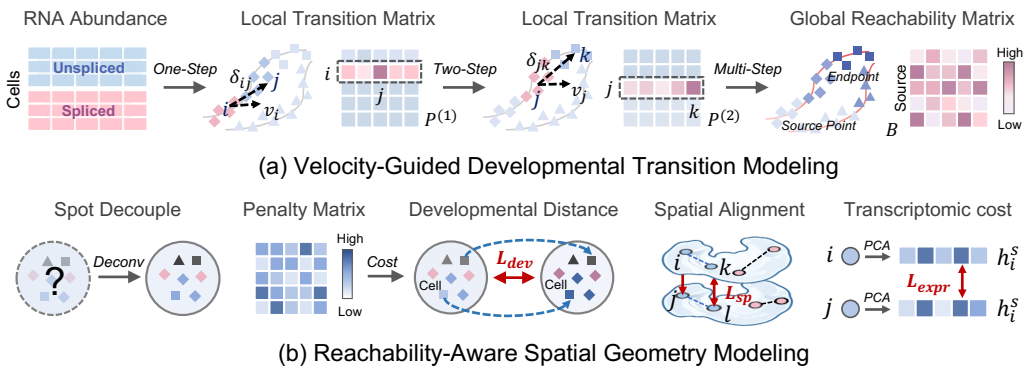


Figure 2: **Overview of the proposed DRIFT framework.** (a) Velocity-guided developmental transition modeling derives directed cell-type reachability from local RNA-velocity transitions. (b) Reachability-aware spatial geometry lifts cell-type reachability to spatial spots and integrates developmental, transcriptomic, and spatial information for cross-time alignment.

4 METHODS

Our framework transforms local cellular dynamics into developmental geometry between spatial spots as shown in Fig. 2. Directed cell-type reachability is inferred from a velocity-informed Markov process and lifted to the spot level via composition-aware optimal transport. The resulting developmental distance is integrated with transcriptomic and spatial information for cross-time alignment.

4.1 VELOCITY-GUIDED DEVELOPMENTAL TRANSITION MODELING

Given the cell-state and velocity representations $(\mathbf{z}_i, \mathbf{v}_i)$, we first construct a local Markov process describing one-step, velocity-consistent transitions between neighboring cells. We then propagate these local transitions to obtain directed, multi-step developmental reachability between cell types.

Local Transition Matrix As shown in Fig.2 (a), for each cell i, let $\mathcal{N}(i)$ denote its k nearest neighbors and $\boldsymbol{\delta}_{ij} = \mathbf{z}_j - \mathbf{z}_i$ the displacement to neighbor j. The directional agreement between the velocity of cell i and the displacement toward cell j is

$$c_{ij} = \frac{\mathbf{v}_i^\top \boldsymbol{\delta}_{ij}}{\|\mathbf{v}_i\|_2 \|\boldsymbol{\delta}_{ij}\|_2}.$$

A plausible transition should both follow the velocity direction and remain local in the cell-state space. We therefore define

$$w_{ij} = \exp[\kappa(c_{ij} - 1)] \exp\left(-\frac{\|\boldsymbol{\delta}_{ij}\|_2^2}{2\sigma_i^2}\right), \qquad j \in \mathcal{N}(i), \tag{5}$$

where $\kappa > 0$ controls directional selectivity and $\sigma_i > 0$ is the local neighborhood bandwidth. The corresponding one-step transition probability is

$$P_{ij} = \frac{w_{ij}}{\sum_{\ell \in \mathcal{N}(i)} w_{i\ell}}, \qquad j \in \mathcal{N}(i), \tag{6}$$

with $P_{ij} = 0$ otherwise. The resulting row-stochastic matrix P assigns higher probability to transitions that agree with both the inferred velocity field and local cell-state geometry.

Global Reachability Matrix The matrix P describes only one-step transitions. To capture multi-step developmental accessibility, representative late-state cells $\mathcal{A}_b$ are selected for each target-stage cell type $b \in \{1, \ldots, K\}$ and treated as absorbing endpoints. For each endpoint e, we set $P_{ee} = 1$ and $P_{ej} = 0$ for $j \neq e$. After reordering transient and absorbing states, the transition matrix becomes

$$\widetilde{P} = \begin{pmatrix} Q & R \\ 0 & I \end{pmatrix}, \tag{7}$$

where Q contains transient-to-transient transitions and R transient-to-endpoint transitions.

For a transient cell i, $(Q^n R)_{ie}$ is the probability of reaching endpoint e after n transient transitions. Summing over all path lengths gives the first-hitting probability matrix

$$\begin{aligned} B &= R + QR + Q^2 R + \cdots \\ &= \left(\sum_{n=0}^{\infty} Q^n \right) R = (I - Q)^{-1} R, \end{aligned} \tag{8}$$

where the last equality holds for an absorbing chain with $\rho(Q) < 1$. Thus, B_{ie} is the probability that a walk starting from transient cell i is eventually absorbed at endpoint e.

We aggregate these endpoint probabilities into a directed cell-type reachability matrix T:

$$T_{ab} = \frac{1}{|\mathcal{I}_a|} \sum_{i \in \mathcal{I}_a} \sum_{e \in \mathcal{A}_b} B_{ie}, \tag{9}$$

where $\mathcal{I}_a$ denotes source-stage cells of type a. Hence, T_{ab} measures the probability that a source cell of type a eventually reaches an endpoint of target type b.

4.2 Reachability-Aware Spot Geometry

The matrix T describes developmental accessibility between individual cell types, whereas spatial transcriptomic spots contain mixtures of cell types. We therefore lift cell-type reachability to the spot level through their full cell-type compositions as inllustrated in Fig.2 (b).

Composition-Aware Developmental Distance We first convert cell-type reachability into a normalized developmental cost that penalizes unlikely transitions:

$$C_{ab}^{\mathrm{dev}} = \frac{-\log(\max\{T_{ab}, \delta\})}{-\log \delta}, \qquad 0 < \delta < 1, \tag{10}$$

so that highly reachable transitions receive low cost. For source spot i and target spot j, let $\boldsymbol{\theta}_i^s, \boldsymbol{\theta}_j^t \in \Delta^{K-1}$ denote their cell-type compositions. For this spot pair, let γ_{ab}^{ij} denote the amount of mass transported from source cell type a to target cell type b, satisfying $\sum_b \gamma_{ab}^{ij} = \theta_{ia}^s$, $\sum_a \gamma_{ab}^{ij} = \theta_{jb}^t$, and $\gamma_{ab}^{ij} \geq 0$. We define their developmental distance as

$$D_{ij} = \min \sum_{a,b} \gamma_{ab}^{ij} C_{ab}^{\mathrm{dev}} + \tau \sum_{a,b} \gamma_{ab}^{ij} \log \frac{\gamma_{ab}^{ij}}{\theta_{ia}^s \theta_{jb}^t}, \tag{11}$$

where $\tau > 0$ controls entropic regularization. A small D_{ij} indicates that the composition of source spot i can be transported to target spot j through developmentally reachable cell-type transitions.

Theorem 1 (Developmental Plausibility Bound)**.** *For any reachability threshold $r \in [\delta, 1)$, let $c_r = \log r / \log \delta$. If $\gamma^{ij,\star}$ minimizes Eq. equation 11, then*

$$0 \leq D_{ij} \leq 1, \qquad \sum_{a,b:\, T_{ab} \leq r} \gamma_{ab}^{ij,\star} \leq \frac{D_{ij}}{c_r}.$$

A small D_{ij} therefore limits the transport mass assigned to low-reachability cell-type transitions. The proof and stability analysis are provided in Appendix C.1.

Transcriptomic and Spatial Geometry We additionally characterize transcriptomic similarity and spatial organization. Source and target expression profiles are projected into a shared PCA space, yielding embeddings $\mathbf{h}_i^s$ and $\mathbf{h}_j^t$. Their transcriptomic cost is

$$M_{ij} = \left\| \mathbf{h}_i^s - \mathbf{h}_j^t \right\|_2^2, \tag{12}$$

where smaller values indicate greater transcriptomic similarity. Spatial geometry is represented by pairwise distances within each snapshot:

$$S_{ik}^s = \left\| \mathbf{y}_i^s - \mathbf{y}_k^s \right\|_2, \qquad S_{jl}^t = \left\| \mathbf{y}_j^t - \mathbf{y}_l^t \right\|_2. \tag{13}$$

Thus, D, M, S^s, and S^t provide the developmental, transcriptomic, and spatial information used for cross-time alignment.

4.3 Overall Alignment Objective

We combine three sources of information above to infer the cross-time spot alignment. Let $\pi \in \mathbb{R}_+^{n_s \times n_t}$ denote the spot-level coupling, where π_{ij} is the matching mass between source and target spots. We use uniform marginals $\mathbf{a} = \frac{1}{n_s}\mathbf{1}$ and $\mathbf{b} = \frac{1}{n_t}\mathbf{1}$, and denote the feasible coupling set by

$$\Pi(\mathbf{a}, \mathbf{b}) = \left\{\pi \geq 0 : \pi\mathbf{1} = \mathbf{a},\ \pi^\top \mathbf{1} = \mathbf{b}\right\}.$$

Using the quantities defined in Sec. 4.2, we define the transcriptomic, developmental, and spatial alignment costs as follows:

$$\mathcal{L}_{\mathrm{expr}}(\pi) = \textstyle\sum_{i,j} M_{ij}\pi_{ij}, \quad \mathcal{L}_{\mathrm{dev}}(\pi) = \sum_{i,j} D_{ij}\pi_{ij}, \quad \mathcal{L}_{\mathrm{sp}}(\pi) = \sum_{i,k,j,l}(S^s_{ik} - S^t_{jl})^2\pi_{ij}\pi_{kl}.$$

To balance the numerical scales, we normalize the three costs using a fixed feasible reference coupling $\bar{\mathcal{L}}_q(\pi) = \mathcal{L}_q(\pi)/\mathcal{L}_q(\pi_0)$, where $q \in \{\mathrm{expr}, \mathrm{sp}, \mathrm{dev}\}$. The final alignment is obtained by

$$\pi^\star = \arg\min_{\pi \in \Pi(\mathbf{a},\mathbf{b})} \left\{(1-\alpha)\bar{\mathcal{L}}_{\mathrm{expr}}(\pi) + \alpha\bar{\mathcal{L}}_{\mathrm{sp}}(\pi) + \lambda\bar{\mathcal{L}}_{\mathrm{dev}}(\pi) - \varepsilon H(\pi)\right\}, \tag{14}$$

where $H(\pi) = -\sum_{i,j}\pi_{ij}\log\pi_{ij}$. Here, α balances transcriptomic and spatial information, λ controls developmental compatibility, and ε controls entropy regularization.

5 Experiments

5.1 Datasets and Evaluation Metrics

To evaluate DRIFT, we use four human spatiotemporal transcriptomics datasets covering *(i) fetal pancreas development* (Olaniru et al., 2023), *(ii) intestinal development* (Fawkner-Corbett et al., 2021a), *(iii) embryonic limb morphogenesis* (Zhang et al., 2024), and *(iv) acute skin wound healing* (Liu et al., 2025). Each dataset contains temporally ordered spatial snapshots and a matched single-cell reference. We consider two tasks, *(i) cross-time spatial alignment* and *(ii) held-out stage reconstruction and prediction*. The former evaluates whether spot couplings recover plausible developmental correspondences, while the latter reconstructs unseen intermediate stages by interpolation and predicts future stages by extrapolation. All alignment methods use the same linear transport and flow-matching predictors for reconstruction. Details are provided in Appendix E.

For cross-time alignment, we compare inferred couplings with independent directed lineage graphs of known developmental relationships. We report LCR, MLDD, CWLD, and LSC, where higher LCR and LSC and lower MLDD and CWLD indicate better consistency. For held-out reconstruction and prediction, we report MMD over joint expression and local context, cell-type Jensen–Shannon divergence, and Moran's I error, with lower values indicating better performance. The lineage graphs are used only for evaluation and unavailable to DRIFT during training. Throughout the paper, the best and second-best results are highlighted with blue and pink backgrounds, respectively.

5.2 Experimental Details

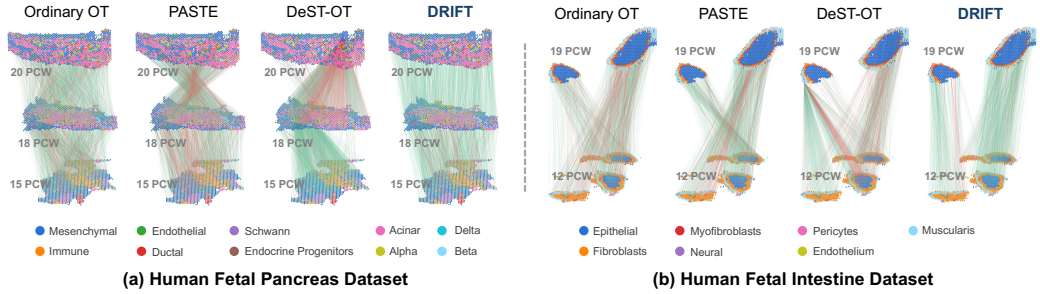
The alignment methods use their corresponding software implementations, including Python Optimal Transport for OT and FGW-OT, `paste3` for PASTE and PASTE2, the authors' implementation for DeST-OT, and our implementation for DRIFT. All methods use the same stage observations, training-data-fitted preprocessing, and downstream predictors to ensure a consistent comparison. Unless otherwise specified, we use 2,000 highly variable genes and 30 principal components. DRIFT uses $(\alpha, \omega_{\mathrm{sp}}, \lambda) = (0.5, 1.0, 1.0)$, a composition OT regularization coefficient of 0.05, and at most 100 outer iterations. The shared flow-matching predictor is trained with AdamW for 1,000 updates using a batch size of 2,048, a learning rate of 2×10^{-4}, and a weight decay of 10^{-5}. Efficiency experiments are conducted on one NVIDIA RTX PRO 6000 Blackwell Server Edition GPU with at most eight CPU threads per process. Further details are provided in Appendix B.4 and Appendix B.5.

5.3 Comparison Experiments

Cross-Time Spatial Alignment Table 1 evaluates whether inferred correspondences follow reference lineage graphs. At dataset level, DRIFT achieves the highest lineage consistency (LCR), the

Table 1: **Lineage alignment quality across developmental spatial datasets.** Numbers in parentheses indicate the number of spatial transcriptomics sections. Ordi. denotes Ordinary OT.

Method	Pancreas (2)				Intestine (2)			
	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑
Ordi. OT	0.539	0.449	0.557	0.270	0.354	0.593	0.611	0.267
FGW OT	0.432	0.555	0.657	0.217	0.358	0.583	0.642	0.263
PASTE	0.239	0.748	0.749	0.112	0.220	0.711	0.708	0.168
PASTE2	0.268	0.719	0.722	0.078	0.243	0.687	0.675	0.101
DeST-OT	0.586	0.406	0.458	0.300	0.386	0.560	0.575	0.295
DRIFT	0.845	0.152	0.284	0.485	0.573	0.384	0.599	0.433
Method	**Limb (3)**				**Skin Wound (5)**			
	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑
Ordi. OT	0.877	0.116	0.202	0.070	0.623	0.375	0.446	0.175
FGW OT	0.843	0.147	0.276	0.067	0.561	0.435	0.532	0.158
PASTE	0.425	0.572	0.573	0.015	0.332	0.666	0.671	0.089
PASTE2	0.393	0.603	0.593	0.009	0.268	0.729	0.727	0.047
DeST-OT	0.839	0.155	0.158	0.063	0.653	0.343	0.387	0.185
DRIFT	0.888	0.092	0.161	0.076	0.682	0.316	0.357	0.237



(a) Human Fetal Pancreas Dataset (b) Human Fetal Intestine Dataset

Figure 3: **Cross-time spot alignments on human fetal pancreas and intestine datasets.** We compare Ordinary OT, PASTE, DesT-OT and our proposed DRIFT using links between stages. Green lines denote lineage-consistent or unjudged correspondences, while red lines denote lineage-inconsistent correspondences.

lowest mean lineage deviation (MLDD), and the highest lineage-supported coverage (LSC) across all four datasets. The gains are particularly pronounced in pancreas and intestine, where similarity-based methods frequently connect transcriptionally close but lineage-distant states. For example, in pancreas, DRIFT increases LCR from 0.586 for strongest baseline to 0.845 while reducing MLDD from 0.406 to 0.152. A similar pattern is observed in intestine, with LCR increasing from 0.386 to 0.573 and MLDD decreasing from 0.560 to 0.384. Even in limb, where several methods already attain high consistency, DRIFT further reduces lineage deviation from 0.116 to 0.092.

At the coupling level, DRIFT achieves the lowest CWLD on pancreas and skin wound and ranks second on intestine and limb. Its CWLD is 0.599 on intestine and 0.161 on limb, compared with 0.575 and 0.158 for DeST-OT, respectively. LCR is conditional on evaluable strongest links. On limb, DRIFT achieves an LCR of 0.888, with top-link coverage of 0.084 and LSC of 0.076. Full alignment results are reported in Table S2.

Figure 3 visualizes cross-time spot correspondences for the pancreas and intestine datasets, where green links denote lineage-consistent or unjudged matches and red links indicate lineage-inconsistent matches. Compared with the baselines, DRIFT exhibits visibly fewer lineage-inconsistent links while preserving coherent correspondences across developmental stages, consistent with the quantitative results in Table 1. Stage-pair lineage scores and coverage are reported in Table S5, with corresponding visualizations for limb and skin wound provided in Figures S2 and S3.

Held-Out Stage Reconstruction and Prediction We next evaluate reconstruction of unseen tissue states from the inferred cross-time alignments. We consider two settings. In *interpolation*, an

Table 2: **Interpolation benchmark.** Reconstruction of withheld stages using MMD, Moran's I error, and cell-type JS divergence. Lower values are better.

Method	Pancreas			Intestine			Limb			Skin wound		
	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS
Linear Methods												
Ordi. OT	0.712	0.186	0.023	0.621	0.430	0.128	0.427	0.377	0.050	0.311	0.422	0.071
FGW OT	0.641	0.197	0.024	0.658	0.393	0.128	0.442	0.363	0.044	0.325	0.471	0.071
PASTE	0.750	0.143	0.021	0.484	0.222	0.129	0.387	0.191	0.033	0.329	0.200	0.071
PASTE2	0.646	0.282	0.022	0.484	0.359	0.137	0.428	0.238	0.039	0.314	0.332	0.060
DeST-OT	0.544	0.149	0.024	0.665	0.358	0.127	0.383	0.308	0.033	0.311	0.412	0.065
DRIFT	0.543	0.154	0.021	0.469	0.339	0.122	0.393	0.289	0.026	0.303	0.313	0.055
Flow matching Methods												
Ordi. OT	0.563	0.345	0.022	0.454	0.228	0.128	0.399	0.194	0.050	0.338	0.360	0.057
FGW OT	0.600	0.322	0.023	0.476	0.248	0.128	0.377	0.190	0.056	0.328	0.348	0.057
PASTE	0.667	0.212	0.022	0.458	0.239	0.128	0.414	0.190	0.050	0.336	0.273	0.065
PASTE2	0.566	0.460	0.025	0.506	0.346	0.136	0.421	0.188	0.039	0.326	0.303	0.063
DeST-OT	0.556	0.286	0.022	0.539	0.256	0.128	0.368	0.193	0.045	0.331	0.364	0.057
DRIFT	0.550	0.122	0.021	0.443	0.198	0.120	0.368	0.182	0.027	0.307	0.233	0.056

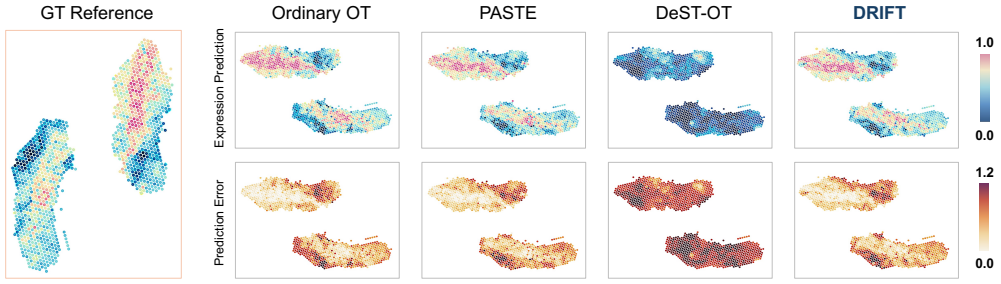


Figure 4: **Linear reconstruction of the developing limb.** PCW6.2 is reconstructed from PCW5.6 and PCW7.2 using Ordinary OT, PASTE, DeST-OT, and DRIFT, with observed slice as reference. The top row shows reconstructed expression, and bottom row shows absolute prediction errors.

intermediate spatial snapshot is held out and reconstructed from the surrounding observed stages. In *extrapolation*, the target stage lies beyond the observed temporal range. For each inferred coupling, we evaluate both direct linear transport and flow matching, allowing us to assess the alignment under both a simple transport rule and a learned continuous dynamics model. We report the joint expression–neighborhood MMD, cell-type JS divergence for cellular composition, and Moran's I error for spatial gene-expression organization. Results and the reporting protocol are in Appendix F.2.

Interpolation. Table 2 reports reconstruction of held-out intermediate stages. Under flow matching, DRIFT achieves the lowest Moran's I error on all four datasets and ranks first in joint MMD, with limb MMD rounding to 0.368 for both DRIFT and DeST-OT. Under linear interpolation, DRIFT achieves the lowest MMD on pancreas, intestine, and skin wound, while PASTE obtains the lowest Moran's I error on all four datasets. DRIFT also ranks first in cell-type JS across all eight conditions using unrounded values, with some ties after rounding. Figure 4 shows a PCW6.2 reconstruction from PCW5.6 and PCW7.2. DRIFT better preserves the spatial distribution of high- and low-expression regions with localized errors. Ordinary OT and PASTE recover the overall structure but show broader discrepancies, whereas DeST-OT underestimates expression across much of the tissue, producing larger and more widespread errors.

Extrapolation. Table 3 evaluates prediction beyond the observed temporal range. Across the eight settings, DRIFT achieves the lowest joint MMD in five and the lowest Moran's I error in five. Under flow matching, it attains the lowest Moran's I error on all four datasets and the lowest joint MMD on intestine, limb, and skin wound. Linear extrapolation yields metric-specific rankings across tissues, as detailed in Table 3. On pancreas, DRIFT has the lowest cell-type JS under both predictors.

Table 3: **Extrapolation benchmark.** Prediction beyond the observed time range measured by MMD, Moran's I error, and cell-type JS divergence. Lower is better.

Method	Pancreas			Intestine			Limb			Skin wound		
	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS
Linear Methods												
Ordi. OT	0.755	0.237	0.028	0.690	0.377	0.391	0.732	0.249	0.172	0.615	0.390	0.086
FGW OT	0.717	0.199	0.027	0.688	0.393	0.379	0.707	0.243	0.159	0.626	0.377	0.092
PASTE	0.693	0.178	0.022	0.708	0.414	0.374	0.710	0.398	0.142	0.668	0.186	0.086
PASTE2	0.695	0.281	0.037	0.711	0.524	0.388	0.715	0.335	0.162	0.670	0.223	0.127
DeST-OT	0.721	0.220	0.028	0.680	0.329	0.425	0.711	0.181	0.168	0.617	0.371	0.089
DRIFT	0.705	0.174	0.013	0.664	0.361	0.380	0.702	0.189	0.157	0.621	0.339	0.088
Flow Matching Methods (IVP)												
Ordi. OT	0.756	0.269	0.022	0.732	0.348	0.394	0.748	0.179	0.172	0.632	0.496	0.085
FGW OT	0.724	0.254	0.023	0.731	0.356	0.381	0.737	0.213	0.160	0.631	0.474	0.086
PASTE	0.698	0.266	0.024	0.731	0.472	0.372	0.740	0.382	0.141	0.653	0.355	0.086
PASTE2	0.724	0.301	0.032	0.732	0.537	0.391	0.744	0.610	0.157	0.674	0.436	0.127
DeST-OT	0.716	0.254	0.024	0.731	0.257	0.422	0.738	0.204	0.169	0.628	0.450	0.091
DRIFT	0.709	0.182	0.013	0.723	0.218	0.377	0.727	0.165	0.152	0.605	0.170	0.085

5.4 Ablation Experiments

We compare the full model with four variants in Fig. 5. *w/o development* removes the developmental cost, *One-step* retains only direct transitions with a Full-prior fallback when no direct jump mass is available, *Dominant* represents each spot by its dominant cell type, and *Independent* replaces the inner composition OT with $\theta_s^\top C^{\mathrm{dev}} \theta_t$. Bars show the mean change relative to the full model over ten paired seeds. Additional alignment and prediction ablations are provided in Appendix F.3.

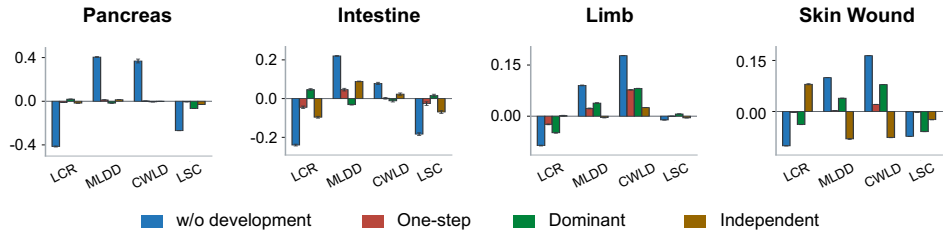


Figure 5: **Paired alignment ablations.** Bars show mean differences from the full model over ten paired seeds, with whiskers indicating sample standard deviation after stage-pair averaging. Higher LCR and LSC and lower MLDD and CWLD indicate better performance. The horizontal line denotes the full model baseline at zero.

Removing the developmental term degrades alignment across all four datasets, lowering LCR and LSC while increasing MLDD and CWLD, with the largest effects in pancreas and intestine. The One-step variant underperforms the full multi-step model on LCR, MLDD, and CWLD, supporting multi-step reachability. Composition simplifications are tissue dependent. Dominant improves several metrics in intestine but is less consistent elsewhere, while Independent improves LCR, MLDD, and CWLD in skin wound at the cost of lower LSC. Overall, developmental reachability provides more consistent gains than composition-aware transport.

6 Discussion

We presented DRIFT, a developmental geometry-aware framework for aligning unpaired spatial transcriptomic snapshots across time. By integrating velocity-derived directional reachability with composition-aware optimal transport, DRIFT captures developmental compatibility beyond transcriptomic similarity alone. Experiments across four human spatiotemporal transcriptomic datasets demonstrate that DRIFT improves lineage-consistent alignment and supports more accurate reconstruction of unseen tissue states. These results highlight the importance of incorporating directional developmental information into computational models of tissue dynamics.

AI USE STATEMENT

Generative AI tools were used as research and writing assistants during the development of this work. Their use included assisting with drafting and polishing parts of the manuscript, suggesting organizational and methodological structures, and supporting the implementation of research code, including components of the proposed method and experimental pipeline. AI tools were also used to help revise technical explanations and improve the clarity of the presentation.

All AI-assisted text, code, mathematical formulations, and methodological suggestions were reviewed, edited, and validated by the authors. Experimental results were obtained from the authors' implementations and analyses, and the authors take full responsibility for the methodology, results, claims, and final content of the paper.

REPRODUCIBILITY STATEMENT

All datasets used in this study with accession numbers provided in Appendix E. Complete preprocessing, model construction, optimization, prediction, and evaluation procedures are specified in Appendix B.4. Formal benchmark and paired comparisons share preprocessing, evaluation settings, and random seeds whenever applicable. Ablations modify only the designated component. Runtime and memory measurements follow the protocol in Appendix F.1. These specifications, together with the reported public data sources and implementation parameters, provide the information required to reproduce the experiments.

REFERENCES

Volker Bergen, Marius Lange, Stefan Peidli, F Alexander Wolf, and Fabian J Theis. Generalizing rna velocity to transient cell states through dynamical modeling. *Nature biotechnology*, 38(12): 1408–1414, 2020.

John P Bryan, Samouil L Farhi, and Brian Cleary. Accurate trajectory inference in time-series spatial transcriptomics with structurally-constrained optimal transport. *Nature Communications*, 2026.

Dylan M Cable, Evan Murray, Luli S Zou, Aleksandrina Goeva, Evan Z Macosko, Fei Chen, and Rafael A Irizarry. Robust decomposition of cell type mixtures in spatial transcriptomics. *Nature biotechnology*, 40(4):517–526, 2022.

Marco Cuturi. Lightspeed computation of optimal transportation distances. *Advances in Neural Information Processing Systems*, 26(2):2292–2300, 2013.

Matthew D Demmings, Luana da Silva Chagas, Marianela E Traetta, Rui S Rodrigues, Maria Florencia Acutain, Evgeny Barykin, Ashok Kumar Datusalia, Liliana German-Castelan, Vanesa S Mattera, Pedzisai Mazengenya, et al. (re) building the nervous system: A review of neuron–glia interactions from development to disease. *Journal of neurochemistry*, 169(1):e16258, 2025.

Rasa Elmentaite, Natsuhiko Kumasaka, Kenny Roberts, Aaron Fleming, Emma Dann, Hamish W King, Vitalii Kleshchevnikov, Monika Dabrowska, Sophie Pritchard, Liam Bolt, et al. Cells of the human intestinal tract mapped across space and time. *Nature*, 597(7875):250–255, 2021.

Silas Ruhrberg Estévez, Nicolas Huynh, Tennison Liu, Roderik M Kortlever, Gerard I Evan, David L Bentley, and Mihaela van der Schaar. Cellbridge: Learning cellular trajectories via interaction-aware alignment. *arXiv preprint arXiv:2605.30635*, 2026.

David Fawkner-Corbett, Agne Antanaviciute, Kaushal Parikh, Marta Jagielowicz, Ana Sousa Gerós, Tarun Gupta, Neil Ashley, Doran Khamis, Darren Fowler, Edward Morrissey, et al. Spatiotemporal analysis of human intestinal development at single-cell resolution. *Cell*, 184(3):810–826, 2021a.

David Fawkner-Corbett, Agne Antanaviciute, Kaushal Parikh, Marta Jagielowicz, Ana Sousa Geros, Tarun Gupta, Neil Ashley, Doran Khamis, Darren Fowler, Edward Morrissey, Chris Cunningham, Paul R. V. Johnson, Hashem Koohy, and Alison Simmons. Spatiotemporal analysis of human intestinal development at single cell resolution: Supplementary data, 2021b.

Rémi Flamary, Nicolas Courty, Alexandre Gramfort, Mokhtar Z Alaya, Aurélie Boisbunon, Stanislas Chambon, Laetitia Chapel, Adrien Corenflos, Kilian Fatras, Nemo Fournier, et al. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

Aden Forrow and Geoffrey Schiebinger. Lineageot is a unified framework for lineage tracing and trajectory inference. *Nature communications*, 12(1):4940, 2021.

Arthur Gretton, Karsten M Borgwardt, Malte J Rasch, Bernhard Schölkopf, and Alexander Smola. A kernel two-sample test. *The journal of machine learning research*, 13:723–773, 2012.

Laleh Haghverdi, Maren Büttner, F Alexander Wolf, Florian Buettner, and Fabian J Theis. Diffusion pseudotime robustly reconstructs lineage branching. *Nature methods*, 13(10):845–848, 2016.

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

Alice H Huang. Coordinated development of the limb musculoskeletal system: Tendon and muscle patterning and integration with the skeleton. *Developmental biology*, 429(2):420–428, 2017.

Geert-Jan Huizing, Jules Samaran, Daniele Capocefalo, Anna Audit, Gabriel Peyré, and Laura Cantini. Stories: learning cell fate landscapes from spatial transcriptomics using optimal transport. *Nature Methods*, 23(3):522–531, 2026.

Rachel E Jennings, Andrew A Berry, James P Strutt, David T Gerrard, and Neil A Hanley. Human pancreas development. *Development*, 142(18):3126–3137, 2015.

Andrew Jones, F William Townes, Didong Li, and Barbara E Engelhardt. Alignment of spatial genomics data using deep gaussian processes. *Nature methods*, 20(9):1379–1387, 2023.

Kacper Kapuśniak, Peter Potaptchik, Teodora Reu, Leo Zhang, Alexander Tong, Michael Bronstein, Avishek J Bose, and Francesco Di Giovanni. Metric flow matching for smooth interpolations on the data manifold. *Advances in Neural Information Processing Systems*, 37:135011–135042, 2024.

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.

Gioele La Manno, Ruslan Soldatov, Amit Zeisel, Emelie Braun, Hannah Hochgerner, Viktor Petukhov, Katja Lidschreiber, Maria E Kastriti, Peter Lönnerberg, Alessandro Furlan, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018.

Marius Lange, Volker Bergen, Michal Klein, Manu Setty, Bernhard Reuter, Mostafa Bakhti, Heiko Lickert, Meshal Ansari, Janine Schniering, Herbert B Schiller, et al. Cellrank for directed single-cell fate mapping. *Nature methods*, 19(2):159–170, 2022.

Prisca Liberali and Alexander F Schier. The evolution of developmental biology through conceptual and technological revolutions. *Cell*, 187(14):3461–3495, 2024.

Jianhua Lin. Divergence measures based on the shannon entropy. *IEEE Transactions on Information theory*, 37(1):145–151, 1991.

Longqi Liu, Ao Chen, Yuxiang Li, Jan Mulder, Holger Heyn, and Xun Xu. Spatiotemporal omics for biology and medicine. *Cell*, 187(17):4488–4519, 2024.

Xinhao Liu, Ron Zeira, and Benjamin J Raphael. Partial alignment of multislice spatially resolved transcriptomics data. *Genome research*, 33(7):1124–1132, 2023.

Zhuang Liu, Xiaowei Bian, Lihua Luo, Åsa K Björklund, Li Li, Letian Zhang, Yongjian Chen, Lei Guo, Juan Gao, Chunyan Cao, et al. Spatiotemporal single-cell roadmap of human skin wound healing. *Cell stem cell*, 32(3):479–498, 2025.

Oladapo Edward Olaniru, Ulrich Kadolsky, Shichina Kannambath, Heli Vaikkinen, Kathy Fung, Pawan Dhami, and Shanta J Persaud. Single-cell transcriptomic and spatial landscapes of the developing human pancreas. *Cell metabolism*, 35(1):184–199, 2023.

Giovanni Palla, Hannah Spitzer, Michal Klein, David Fischer, Anna Christina Schaar, Louis Benedikt Kuemmerle, Sergei Rybakov, Ignacio L Ibarra, Olle Holmberg, Isaac Virshup, et al. Squidpy: a scalable framework for spatial omics analysis. *Nature methods*, 19(2):171–178, 2022.

Oscar A Peña and Paul Martin. Cellular and molecular mechanisms of skin wound healing. *Nature Reviews Molecular Cell Biology*, 25(8):599–616, 2024.

Qiangwei Peng, Peijie Zhou, and Tiejun Li. stvcr: spatiotemporal dynamics of single cells. *Nature Methods*, 23(3):542–553, 2026.

Xiaojie Qiu, Qi Mao, Ying Tang, Li Wang, Raghav Chawla, Hannah A Pliner, and Cole Trapnell. Reversed graph embedding resolves complex single-cell trajectories. *Nature methods*, 14(10): 979–982, 2017.

Xiaojie Qiu, Daniel Y Zhu, Jiajun Yao, Zehua Jing, Lulu Zuo, Mingyue Wang, Kyung Hoi Min, Hailin Pan, Shuai Wang, Sha Liao, et al. Spateo: multidimensional spatiotemporal modeling of single-cell spatial transcriptomics. *BioRxiv*, pp. 2022–12, 2022.

Santanu Subhash Rathod, Francesco Ceccarelli, Sean B Holden, Pietro Liò, Xiao Zhang, and Jovan Tanevski. Contextflow: Context-aware flow matching for trajectory inference from spatial omics data. *arXiv preprint arXiv:2510.02952*, 2025.

Geoffrey Schiebinger, Jian Shu, Marcin Tabaka, Brian Cleary, Vidya Subramanian, Aryeh Solomon, Joshua Gould, Siyan Liu, Stacie Lin, Peter Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

Manu Setty, Vaidotas Kiseliovas, Jacob Levine, Adam Gayoso, Linas Mazutis, and Dana Pe'Er. Characterization of cell fate probabilities in single-cell data with palantir. *Nature biotechnology*, 37(4):451–460, 2019.

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.

Kelly Street, Davide Risso, Russell B Fletcher, Diya Das, John Ngai, Nir Yosef, Elizabeth Purdom, and Sandrine Dudoit. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC genomics*, 19(1):477, 2018.

Vayer Titouan, Nicolas Courty, Romain Tavenard, and Rémi Flamary. Optimal transport for structured data with application on graphs. In *International Conference on Machine Learning*, pp. 6275–6284. PMLR, 2019.

Alexander Tong, Jessie Huang, Guy Wolf, David Van Dijk, and Smita Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. In *International conference on machine learning*, pp. 9526–9536. PMLR, 2020.

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.

Shou-Wen Wang, Michael J Herriges, Kilian Hurley, Darrell N Kotton, and Allon M Klein. Cospar identifies early cell fate biases from single-cell transcriptomic and lineage information. *Nature Biotechnology*, 40(7):1066–1074, 2022.

F Alexander Wolf, Philipp Angerer, and Fabian J Theis. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19(1):15, 2018.

F Alexander Wolf, Fiona K Hamey, Mireya Plass, Jordi Solana, Joakim S Dahlin, Berthold Göttgens, Nikolaus Rajewsky, Lukas Simon, and Fabian J Theis. Paga: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome biology*, 20(1):59, 2019.

Chen-Rui Xia, Zhi-Jie Cao, Xin-Ming Tu, and Ge Gao. Spatial-linked alignment tool (slat) for aligning heterogenous slices. *Nature Communications*, 14(1):7236, 2023.

Grace Hui Ting Yeo, Sachit D Saksena, and David K Gifford. Generative modeling of single-cell time series with prescient enables prediction of cell trajectories with interventions. *Nature communications*, 12(1):3222, 2021.

Ron Zeira, Max Land, Alexander Strzalkowski, and Benjamin J Raphael. Alignment and integration of spatial transcriptomics data. *Nature Methods*, 19(5):567–575, 2022.

Bao Zhang, Peng He, John EG Lawrence, Shuaiyu Wang, Elizabeth Tuck, Brian A Williams, Kenny Roberts, Vitalii Kleshchevnikov, Lira Mamanova, Liam Bolt, et al. A human embryonic limb cell atlas resolved in space and time. *Nature*, 635(8039):668–678, 2024.

Stephen Zhang, Anton Afanassiev, Laura Greenstreet, Tetsuya Matsumoto, and Geoffrey Schiebinger. Optimal transport analysis reveals trajectories in steady-state systems. *PLoS computational biology*, 17(12):e1009466, 2021.

Xiang Zhou, Kangning Dong, and Shihua Zhang. Integrating spatial transcriptomics data across different conditions, technologies and developmental stages. *Nature Computational Science*, 3 (10):894–906, 2023.

APPENDIX

A COMPARISON METHODS **14**

B MODEL DETAILS **15**

B.1 Input Representation and Preprocessing 15

B.2 Developmental Reachability from Velocity 16

B.3 Developmental Distance from Composition 16

B.4 Reachability-Aware Spatial Alignment and Optimization 17

B.5 Downstream Interpolation and Extrapolation 17

C THEORETICAL ANALYSIS **18**

C.1 Proof of the Plausibility Certificate 18

C.2 Proof of Stability 20

C.3 Proof of Invalid Alignment Mass Control 23

D EVALUATION METRICS **25**

D.1 Transcriptomic Distribution Fidelity 25

D.2 Cellular Composition and Developmental Fidelity 26

D.3 Lineage-Graph Alignment Metrics 26

D.4 Spatial Organization Fidelity 28

E DATASETS **29**

F ADDITIONAL EXPERIMENTS **30**

F.1 Experimental Protocol 30

F.2 Benchmark Results 33

F.3 Component Ablations 38

F.4 Runtime, Memory, and Scaling 40

A COMPARISON METHODS

Ordinary optimal transport. Optimal transport (OT) provides a general framework for estimating probabilistic correspondences between two empirical distributions by minimizing the expected cost of transporting probability mass between them. In transcriptomic alignment, a natural formulation defines the transport cost directly from pairwise gene-expression dissimilarities, such that transcriptionally similar observations receive lower matching costs. Entropic regularization can further be introduced to obtain a smooth coupling that can be efficiently optimized using Sinkhorn iterations (Cuturi, 2013). We use this expression-only OT formulation as the simplest baseline in our experiments. It evaluates whether cross-time developmental correspondence can be recovered solely from transcriptomic proximity, without incorporating spatial organization or directional developmental information.

Fused Gromov–Wasserstein optimal transport. Fused Gromov–Wasserstein (FGW) optimal transport extends classical OT by jointly considering feature similarity between domains and relational structure within each domain (Titouan et al., 2019). Whereas ordinary Wasserstein transport compares observations according to their features, Gromov–Wasserstein transport compares the

pairwise relations among observations; FGW combines these two sources of information within a single objective. For spatial transcriptomics, the feature cost between domains can be defined from gene expression dissimilarity, while the relational matrices encode pairwise spatial distances within each tissue slice. We therefore use FGW as a generic baseline that accounts for structure and simultaneously exploits transcriptomic and spatial information. However, its relational term is symmetric and designed to preserve geometric structure, rather than to encode the direction or reachability of developmental transitions across time.

PASTE. PASTE is a method based on optimal transport that was developed specifically for pairwise alignment and integration of spatial transcriptomic slices (Zeira et al., 2022). It formulates slice alignment using a fused Gromov–Wasserstein objective that combines transcriptional similarity between spots with preservation of pairwise spatial relationships within the corresponding tissue sections. The resulting soft coupling can align adjacent slices and subsequently reconstruct tissue structure across several slices. PASTE assumes a global alignment in which the probability mass of both slices participates in the transport, making it particularly suitable when the two slices exhibit substantial spatial overlap and broadly comparable tissue composition. In our setting, PASTE serves as a specialized spatial transcriptomics baseline for evaluating whether expression and spatial geometry alone are sufficient to recover developmental correspondence across time.

PASTE2. PASTE2 extends PASTE to spatial transcriptomic slices that overlap only partially or contain tissue regions and cell populations unique to each slice (Liu et al., 2023). Rather than enforcing a full balanced coupling between the two slices, PASTE2 introduces a partial fused Gromov–Wasserstein formulation in which only a fraction s of the total probability mass is transported. This relaxation prevents observations in disjoint regions from being forced into potentially arbitrary correspondences and makes the method more robust to differences in tissue coverage and cellular composition. PASTE2 also provides a procedure for estimating the overlap fraction and can optionally incorporate histological information. We include PASTE2 to assess whether allowing partial spatial correspondence can account for the substantial morphological and compositional changes that may occur between developmental stages.

DeST-OT. Developmental SpatioTemporal Optimal Transport (DeST-OT) is designed specifically for aligning spatial transcriptomic measurements collected across developmental time points (Halmos et al., 2025). DeST-OT formulates alignment across time using semirelaxed optimal transport, relaxing the source marginal while constraining the target marginal so that variations in transported mass can capture heterogeneous cellular growth, death, and differentiation. The method further incorporates transcriptomic and spatial information to construct developmental alignments and derives quantities such as growth rates inferred for each spot to interpret tissue dynamics. Compared with OT, FGW, PASTE, and PASTE2, DeST-OT therefore provides the closest developmental baseline in our experiments. Its treatment of developmental dynamics primarily operates through variation in population mass and spatiotemporal compatibility, whereas our framework additionally constructs an explicit directed developmental geometry from single cell trajectories informed by RNA velocity and evaluates the reachability between heterogeneous spatial compositions.

B Model Details

This section describes the implementation of the alignment model and how its couplings are used for downstream prediction. The configurations below correspond to the saved priors and formal benchmark runs.

B.1 Input Representation and Preprocessing

The alignment model takes spatial expression profiles, two dimensional spot coordinates, and precomputed cell type compositions, together with a single cell reference containing spliced and unspliced abundances. Composition columns are reordered to match the cell types represented in the developmental prior, and each spot's composition is normalized to unit mass.

For spatial expression, we retain shared genes, normalize each spot to a total count of 10^4, and apply a $\log(1 + x)$ transformation. Highly variable genes are selected using Scanpy (Wolf et al.,

2018), with stage information used to account for batch. We standardize each selected gene and fit a joint PCA with up to 30 components on the observed stages. Pairwise alignment uses the two participating sections; prediction across several observed stages uses a common representation fitted to those stages. Withheld spatial observations are transformed using the fitted preprocessing and do not contribute to its estimation.

The single cell PCA used for velocity modeling is fitted separately from the spatial expression PCA. The two modalities are connected through the common cell type vocabulary and the resulting developmental cost between types.

B.2 Developmental Reachability from Velocity

B.2.1 Local transition dynamics

For each single cell reference pair, source and target cells are jointly processed by gene filtering, normalization, highly variable gene selection, and PCA. The primary velocity estimator is scVelo's stochastic model (Bergen et al., 2020); velocities in gene space are projected through the same PCA loadings as the cell states.

We construct a Euclidean graph from nearest neighbors in the joint single cell PCA space, excluding each cell itself. The local bandwidth is proportional to the median positive neighbor distance. The transition weights combine velocity–displacement cosine similarity with a Gaussian distance factor, as in Eq. (5). Velocity or displacement vectors with zero norm receive cosine similarity zero. Negative directional agreement reduces an edge's weight smoothly, without a hard directional cutoff.

For clarity, the main text presents the local transition kernel after direct normalization of the neighbor scores. In our implementation, we additionally allow a cell to remain in its current state when all candidate transitions are weak. Specifically, we define the total probability of leaving cell i as

$$m_i = m_{\max} \left(\max_{j \in \mathcal{N}(i)} w_{ij} \right)^{\eta},$$

where $m_{\max} \leq 1$ controls the maximum movement probability and $\eta > 0$ determines its sensitivity to the strongest local transition score. The transition kernel used in practice is therefore

$$P_{ij} = m_i \frac{w_{ij}}{\sum_{\ell \in \mathcal{N}(i)} w_{i\ell}}, \qquad j \in \mathcal{N}(i), \qquad P_{ii} = 1 - m_i.$$

This adaptive self-loop prevents the normalization step from forcing cells to move when the velocity field provides little support for any neighboring transition, while leaving the relative probabilities among candidate destinations unchanged.

B.2.2 Absorbing endpoints and global reachability

Within each supported target cell type, representative cells are selected in decreasing order of velocity pseudotime. Their transition rows are replaced by absorbing boundary conditions. Reverse graph traversal identifies states that can reach a selected endpoint, and the first passage problem is solved on this reachable subset using sparse linear algebra rather than explicitly constructing a matrix inverse.

Source cells are included in aggregation by type when their computed total hitting mass is at least $1 - 10^{-6}$. Their endpoint probabilities are normalized, summed by target type, and averaged within source type; unresolved source cells are excluded. The resulting nonzero rows for each type are normalized before constructing the developmental cost. This numerical treatment specifies how Eq. (9) is evaluated when a reference contains unresolved states.

B.3 Developmental Distance from Composition

Following Section 4.2, we convert reachability between types to a normalized negative logarithmic ground cost with $\delta = 10^{-6}$. For each pair of source and target spots, we solve the inner entropic transport problem using Sinkhorn scaling (Cuturi, 2013). A small positive floor is applied to composition entries before renormalization, and the regularization strength is $\tau = 0.05$.

The reported developmental distance includes both the transport cost and the KL term relative to the product of the two compositions. This comparison retains the full mixture of cell types in each spot. The implementation uses a finite iteration budget, so the computed distance approximates the exact optimum characterized by the theoretical results.

B.4 Reachability-Aware Spatial Alignment and Optimization

B.4.1 Alignment Setting

Expression dissimilarity is the squared Euclidean distance between spatial PCA embeddings. Within-section geometry uses Euclidean distances between spot coordinates. These costs and the composition-aware developmental distance enter the three-term objective in Section 4.3.

We use uniform spot marginals and initialize a positive random coupling, which is scaled to the prescribed marginals. The entropic coefficient is set to 0.05 times the median positive entry of the initial gradient of the normalized expression and spatial terms.

For $0 \leq \alpha < 1$, the developmental term is incorporated into the effective feature cost and the spatial distances are rescaled as specified below. We solve the resulting entropic FGW problem using projected gradient descent with Sinkhorn projections in POT (Flamary et al., 2021), allowing up to 100 outer iterations. In the formal benchmark, all datasets use up to 30 principal components and inner composition OT regularization $\tau = 0.05$. We select 2,000 highly variable genes for each dataset. The corresponding $(\alpha, \omega_{\mathrm{sp}}, \lambda)$ settings are $(0.5, 1.0, 1.00)$. The benchmark seed grid is $1, 11, \ldots, 91$.

B.4.2 Optimization

The common solver parameters above, also listed in Table S1, define

$$F_{\mathrm{bench}}(\pi) = (1-\alpha)\bar{\mathcal{L}}_{\mathrm{expr}}(\pi) + \alpha\omega_{\mathrm{sp}}\bar{\mathcal{L}}_{\mathrm{sp}}(\pi) + \lambda\bar{\mathcal{L}}_{\mathrm{dev}}(\pi) - \varepsilon H(\pi).$$

For $c = 1 - \alpha + \alpha\omega_{\mathrm{sp}} > 0$, the objective F_{bench}/c has the form of Eq. (14) with parameters

$$\alpha_{\mathrm{eff}} = \frac{\alpha\omega_{\mathrm{sp}}}{c}, \qquad \lambda_{\mathrm{eff}} = \frac{\lambda}{c}, \qquad \varepsilon_{\mathrm{eff}} = \frac{\varepsilon}{c}.$$

This positive rescaling preserves the minimizing couplings. With the common benchmark settings, $c = 1$ and all datasets use $(\alpha_{\mathrm{eff}}, \lambda_{\mathrm{eff}}) = (0.5, 1.0)$ in Eq. (14).

For efficient optimization, the developmental term is absorbed into the feature cost as $M_{\mathrm{eff}} = M/E_0 + \lambda D/[(1-\alpha)D_0]$, while the spatial matrices are rescaled by $\sqrt{\omega_{\mathrm{sp}}/G_0}$. The resulting objective is optimized with an entropic FGW solver from the Python Optimal Transport library (Flamary et al., 2021). Setting $\lambda = 0$ recovers spatial FGW, while $(\alpha, \lambda) = (0, 0)$ reduces to entropic OT.

B.5 Downstream Interpolation and Extrapolation

B.5.1 Trajectory sampling from couplings

For prediction, each spot is represented by a joint state $\mathbf{x} = [\mathbf{h}; \bar{\mathbf{y}}]$, concatenating its expression PCA embedding and spatial coordinates normalized by the bounding box of the observed stages. Endpoint pairs are sampled from the coupling after normalizing its total mass. For multiple observed stages, we sample the first pair jointly and extend each trajectory using successive couplings with normalized rows. This produces linked observations across the available stages.

Time is normalized to the span of the observed stages. Pancreas and limb use their numerical PCW values; intestine assigns Adult the computational time coordinate 30. Skin uses the ordered coordinates $0, 1, 2, 3$ for D0, D1, D7, and D30. These last two choices are modeling conventions, rather than literal biological ages or elapsed days.

B.5.2 Linear prediction

For a sampled endpoint pair $(\mathbf{x}_s, \mathbf{x}_t)$ and relative prediction time u, the linear predictor is

$$\widehat{\mathbf{x}}(u) = (1-u)\mathbf{x}_s + u\mathbf{x}_t.$$

Values $u \in [0,1]$ give interpolation, while $u > 1$ continues the endpoint secant. With several observed stages, linear extrapolation instead continues the most recent segment:

$$\widehat{\mathbf{x}}(t_\star) = \mathbf{x}_m + \frac{t_\star - t_m}{t_m - t_{m-1}}(\mathbf{x}_m - \mathbf{x}_{m-1}).$$

Both predictors propagate endpoint cell-type compositions linearly, followed by nonnegative clipping and row normalization. Multi-stage extrapolation uses the most recent segment. Cell-type JS uses the equally weighted mean composition of each predictor's saved prediction samples (Appendix D.2).

B.5.3 Conditional flow matching

We also train a time-dependent velocity field v_ψ using conditional flow matching (Tong et al., 2023). For endpoint pairs sampled from the inferred coupling, the training path and target velocity are

$$\mathbf{x}(t) = (1-t)\mathbf{x}_s + t\mathbf{x}_t, \qquad \mathbf{u} = \mathbf{x}_t - \mathbf{x}_s, \qquad t \sim \mathrm{Unif}[0,1].$$

The network minimizes the mean squared error between $v_\psi(\mathbf{x}(t), t)$ and $\mathbf{u}$. For multiple observed stages, a segment is sampled with probability proportional to its duration, the network receives the corresponding global normalized time, and the target velocity is the segment displacement divided by its duration.

The velocity field is an MLP with three hidden layers of width 192 and SiLU activations. Its input concatenates the joint state and scalar time, and its output has the same dimension as the joint state. We train for 1000 updates with batch size 2048 using AdamW, learning rate 2×10^{-4}, and weight decay 10^{-5}. The sampled training sets contain 20,000 trajectories for pancreas and limb, 40,000 for intestine, and 5,000 for skin. Each method uses its own coupling with the same sampling seed and network initialization.

Predictions are generated with 40 steps of fourth order Runge–Kutta integration. The default prediction solves an initial value problem from the earliest observed state and integrates to the requested time, including beyond the observed interval for extrapolation. The additional Next step variant starts from the last observed state at normalized time 1 and integrates forward using the same trained field. The field evolves expression and coordinates; compositions remain those obtained by the separate linear rule above.

C Theoretical Analysis

C.1 Proof of the Plausibility Certificate

To complement Theorem 1, define the unavoidable mass assigned to low reachability

$$m_r(i,j) = \min_{\gamma \in \Pi(\boldsymbol{\theta}_i^s, \boldsymbol{\theta}_j^t)} \sum_{(a,b)\in\mathcal{B}_r} \gamma_{ab}.$$

Since an optimal composition coupling is feasible for this minimum, the certificate also yields

$$m_r(i,j) \le \sum_{(a,b)\in\mathcal{B}_r} (\gamma^\star_{ij})_{ab} \le \frac{D_{ij}}{c_r}, \qquad D_{ij} \ge c_r\, m_r(i,j).$$

Thus, if the two compositions necessarily require transitions of reachability at most r, their developmental distance is bounded away from zero. This explains how full compositions can reveal incompatibilities missed by dominant cell type labels. The certificate bounds low-reachability mass relative to the inferred matrix T.

Proof of Theorem 1. We use the standard conventions for relative entropy: a term with zero numerator contributes zero, while a positive numerator over a zero reference mass contributes $+\infty$. Every feasible coupling vanishes wherever $\boldsymbol{\theta}_i^s \otimes \boldsymbol{\theta}_j^t$ vanishes. On its remaining support, the KL term is finite and continuous, including at zero coupling entries. Thus the objective in Eq. equation 11 is continuous on the nonempty compact coupling polytope and attains its minimum, even when some composition entries are zero.

By construction,

$$C_{ab}^{\mathrm{dev}} = \frac{-\log(\max\{T_{ab}, \delta\})}{-\log \delta},$$

with $T_{ab} \in [0, 1]$ and $0 < \delta < 1$. Hence

$$0 \le C_{ab}^{\mathrm{dev}} \le 1 \qquad \text{for all } a, b.$$

Both terms in Eq. equation 11 are therefore nonnegative, which immediately gives $D_{ij} \ge 0$. For the upper bound, consider the independent coupling

$$\gamma = \boldsymbol{\theta}_i^s \otimes \boldsymbol{\theta}_j^t.$$

It belongs to $\Pi(\boldsymbol{\theta}_i^s, \boldsymbol{\theta}_j^t)$ and satisfies

$$D_{\mathrm{KL}}\big(\gamma \big\| \boldsymbol{\theta}_i^s \otimes \boldsymbol{\theta}_j^t\big) = 0.$$

Since $C_{ab}^{\mathrm{dev}} \le 1$ and $\sum_{a,b} \gamma_{ab} = 1$,

$$D_{ij} \le \left\langle \boldsymbol{\theta}_i^s \otimes \boldsymbol{\theta}_j^t, C^{\mathrm{dev}} \right\rangle \le 1.$$

Now fix $r \in [\delta, 1)$ and consider any $(a, b) \in \mathcal{B}_r$. Since $T_{ab} \le r$ and $r \ge \delta$,

$$\max\{T_{ab}, \delta\} \le r,$$

and therefore

$$C_{ab}^{\mathrm{dev}} = \frac{-\log(\max\{T_{ab}, \delta\})}{-\log \delta} \ge \frac{-\log r}{-\log \delta} = c_r.$$

Let $\gamma_{ij}^\star$ be an optimizer of Eq. equation 11. Because the KL term is nonnegative,

$$D_{ij} \ge \left\langle \gamma_{ij}^\star, C^{\mathrm{dev}} \right\rangle.$$

Restricting the sum to $\mathcal{B}_r$ gives

$$\begin{aligned} D_{ij} &\ge \sum_{(a,b)\in\mathcal{B}_r} (\gamma_{ij}^\star)_{ab} C_{ab}^{\mathrm{dev}} \\ &\ge c_r \sum_{(a,b)\in\mathcal{B}_r} (\gamma_{ij}^\star)_{ab}. \end{aligned}$$

Since $c_r > 0$,

$$\sum_{(a,b)\in\mathcal{B}_r} (\gamma_{ij}^\star)_{ab} \le \frac{D_{ij}}{c_r}.$$

Finally, $m_r(i, j)$ is defined as the minimum mass with low reachability over all feasible couplings, while $\gamma_{ij}^\star$ is one such coupling. Hence

$$m_r(i, j) \le \sum_{(a,b)\in\mathcal{B}_r} (\gamma_{ij}^\star)_{ab}.$$

Combining the two inequalities yields

$$m_r(i, j) \le \sum_{(a,b)\in\mathcal{B}_r} (\gamma_{ij}^\star)_{ab} \le \frac{D_{ij}}{c_r},$$

and equivalently

$$D_{ij} \ge c_r \, m_r(i, j).$$

□

The result follows directly from the negative logarithmic reachability transformation: every unit of mass transported through a transition with $T_{ab} \le r$ incurs at least c_r developmental cost. Thus, a low value of D_{ij} necessarily limits the amount of mass that can use such transitions.

C.2 Proof of Stability

Theorem 2 (Stability of the developmental distance). *If $C^{\mathrm{dev}}, \widehat{C}^{\mathrm{dev}} \in [0,1]^{K\times K}$, then*

$$|D_{ij} - \widehat{D}_{ij}| \le \|C^{\mathrm{dev}} - \widehat{C}^{\mathrm{dev}}\|_{\max} + \mathrm{TV}\Big(\boldsymbol{\theta}_i^s, \widehat{\boldsymbol{\theta}}_i^s\Big) + \mathrm{TV}\Big(\boldsymbol{\theta}_j^t, \widehat{\boldsymbol{\theta}}_j^t\Big). \tag{S1}$$

Extension to perturbations of the Markov kernel. In addition to the bound for cost and composition in Theorem 2, we propagate errors through the first passage construction. Fix the finite single cell state space, cell type labels, nonempty source type groups $\mathcal{I}_a$, and absorbing endpoint groups $\mathcal{A}_b$. Let P and $\widehat{P}$ denote two transition kernels after imposing these same absorbing endpoints. Suppose that, under each kernel, every nonabsorbing state has a path with positive probability to an absorbing endpoint. Construct $T, \widehat{T}$ using their first passage probabilities and the same aggregation in Eq. equation 9, and construct $C^{\mathrm{dev}}, \widehat{C}^{\mathrm{dev}}$ using the same $0 < \delta < 1$. Define

$$\eta_P = \max_u \sum_v |P_{uv} - \widehat{P}_{uv}|,$$

and suppose the maximum expected absorption time under each kernel is at most $h_{\mathrm{abs}} < \infty$. For distances computed with the same $\tau > 0$, we then have

$$|D_{ij} - \widehat{D}_{ij}| \le \frac{h_{\mathrm{abs}}\eta_P}{\delta(-\log\delta)} + \mathrm{TV}\Big(\boldsymbol{\theta}_i^s, \widehat{\boldsymbol{\theta}}_i^s\Big) + \mathrm{TV}\Big(\boldsymbol{\theta}_j^t, \widehat{\boldsymbol{\theta}}_j^t\Big). \tag{S2}$$

The proof below establishes this extension together with the theorem. The assumption of fixed endpoints excludes perturbations caused by changing endpoint selection.

Proof of Theorem 2. We first establish stability with respect to the developmental ground cost and the two spot compositions, and then propagate perturbations of the Markov kernel through the first passage construction.

Step 1: perturbing the spot compositions. First keep the ground cost fixed at C^{dev}. Let $\gamma^\star$ be an optimizer of Eq. equation 11 for $\boldsymbol{\theta}_i^s$ and $\boldsymbol{\theta}_j^t$.

There exists a maximal coupling between $\boldsymbol{\theta}_i^s$ and $\widehat{\boldsymbol{\theta}}_i^s$ whose mismatch probability equals

$$\mathrm{TV}\Big(\boldsymbol{\theta}_i^s, \widehat{\boldsymbol{\theta}}_i^s\Big).$$

Equivalently, there exists a Markov kernel K^s mapping source type labels to perturbed source type labels such that, if $A \sim \boldsymbol{\theta}_i^s$ and $\widehat{A} \sim K^s(\cdot \mid A)$, then

$$\widehat{A} \sim \widehat{\boldsymbol{\theta}}_i^s, \qquad \Pr(A \ne \widehat{A}) = \mathrm{TV}\Big(\boldsymbol{\theta}_i^s, \widehat{\boldsymbol{\theta}}_i^s\Big).$$

Likewise, there exists a kernel K^t such that

$$\Pr(B \ne \widehat{B}) = \mathrm{TV}\Big(\boldsymbol{\theta}_j^t, \widehat{\boldsymbol{\theta}}_j^t\Big).$$

Let $(A, B) \sim \gamma^\star$, and conditionally on (A, B) generate $\widehat{A}$ using K^s and $\widehat{B}$ using K^t, independently. Denote the joint distribution of $(\widehat{A}, \widehat{B})$ by $\widehat{\gamma}$. By construction,

$$\widehat{\gamma} \in \Pi\Big(\widehat{\boldsymbol{\theta}}_i^s, \widehat{\boldsymbol{\theta}}_j^t\Big).$$

Applying the product channel $K^s \otimes K^t$ to both $\gamma^\star$ and $\boldsymbol{\theta}_i^s \otimes \boldsymbol{\theta}_j^t$, the data processing inequality for relative entropy gives

$$D_{\mathrm{KL}}\Big(\widehat{\gamma}\Big\|\widehat{\boldsymbol{\theta}}_i^s \otimes \widehat{\boldsymbol{\theta}}_j^t\Big) \le D_{\mathrm{KL}}\big(\gamma^\star\big\|\boldsymbol{\theta}_i^s \otimes \boldsymbol{\theta}_j^t\big).$$

Moreover, because all entries of C^{dev} lie in $[0,1]$,

$$C^{\mathrm{dev}}_{\widehat{A}\widehat{B}} \le C^{\mathrm{dev}}_{AB} + \mathbf{1}\{A \neq \widehat{A}\} + \mathbf{1}\{B \neq \widehat{B}\}.$$

Taking expectations yields

$$\left\langle \widehat{\gamma}, C^{\mathrm{dev}} \right\rangle \le \left\langle \gamma^\star, C^{\mathrm{dev}} \right\rangle + \mathrm{TV}\left(\boldsymbol{\theta}^s_i, \widehat{\boldsymbol{\theta}}^s_i\right) + \mathrm{TV}\left(\boldsymbol{\theta}^t_j, \widehat{\boldsymbol{\theta}}^t_j\right).$$

Since $\widehat{\gamma}$ is feasible for the perturbed compositions, optimality of the corresponding inner OT value implies that changing only the compositions increases the developmental distance by at most the two terms in total variation. Repeating the same argument with the original and perturbed compositions interchanged gives the corresponding lower bound. Therefore, with the ground cost fixed,

$$\left|D_{ij} - D^{\mathrm{comp}}_{ij}\right| \le \mathrm{TV}\left(\boldsymbol{\theta}^s_i, \widehat{\boldsymbol{\theta}}^s_i\right) + \mathrm{TV}\left(\boldsymbol{\theta}^t_j, \widehat{\boldsymbol{\theta}}^t_j\right),$$

where D^{comp}_{ij} denotes the inner OT value computed with the perturbed compositions but the original ground cost.

Step 2: perturbing the developmental ground cost. For fixed compositions, consider any feasible coupling γ. Since γ has unit total mass,

$$\left|\left\langle \gamma, C^{\mathrm{dev}} \right\rangle - \left\langle \gamma, \widehat{C}^{\mathrm{dev}} \right\rangle\right| \le \|C^{\mathrm{dev}} - \widehat{C}^{\mathrm{dev}}\|_{\max}.$$

The KL regularization term is unchanged when only the ground cost is perturbed. Evaluating the two objectives at each other's optimizers therefore gives

$$\left|D^{\mathrm{comp}}_{ij} - \widehat{D}_{ij}\right| \le \|C^{\mathrm{dev}} - \widehat{C}^{\mathrm{dev}}\|_{\max}.$$

Combining this inequality with Step 1 by the triangle inequality proves

$$\begin{aligned}|D_{ij} - \widehat{D}_{ij}| \le \|C^{\mathrm{dev}} - \widehat{C}^{\mathrm{dev}}\|_{\max} \\ + \mathrm{TV}\left(\boldsymbol{\theta}^s_i, \widehat{\boldsymbol{\theta}}^s_i\right) + \mathrm{TV}\left(\boldsymbol{\theta}^t_j, \widehat{\boldsymbol{\theta}}^t_j\right),\end{aligned}$$

which is Eq. equation S1.

Step 3: propagating perturbations of the Markov kernel. Fix the nonabsorbing states and absorbing endpoint set $\mathcal{A}$. Here P and $\widehat{P}$ denote the kernels after imposing the same absorbing endpoints, and Q, R and $\widehat{Q}, \widehat{R}$ denote their respective nonabsorbing and endpoint blocks. For each kernel, the assumed paths with positive probability and finiteness of the state space give an integer $L \ge 1$ and a number $\zeta \in (0,1)$ such that the probability of reaching $\mathcal{A}$ within L steps is at least ζ from every nonabsorbing state. For P, the Markov property then gives

$$\|Q^{kL}\|_\infty \le (1-\zeta)^k, \qquad k \ge 1.$$

Thus $\rho(Q) < 1$, absorption occurs almost surely, and $\sum_{n=0}^{\infty} Q^n = (I-Q)^{-1}$ exists. The same argument applies to $\widehat{Q}$, with possibly different L and ζ. Hence the nonabsorbing states are transient under both kernels.

For a target cell type b, define

$$h^{(b)}_u = \Pr_P\left(\text{the chain is eventually absorbed in } \mathcal{A}_b \mid X_0 = u\right)$$

for each transient state u, and define $\widehat{h}^{(b)}_u$ analogously under $\widehat{P}$. These quantities are exactly the endpoint probabilities in Eq. equation 8 after summing over endpoints in $\mathcal{A}_b$.

Let

$$N = (I-Q)^{-1}.$$

By the preceding argument, N exists and has nonnegative entries. The row sum

$$\sum_v N_{uv}$$

equals the expected number of visits to transient states before absorption when starting from u. By assumption,

$$\|N\|_\infty = \max_u \sum_v N_{uv} \le h_{\mathrm{abs}}.$$

For target type b, the hitting probability vectors satisfy their respective harmonic equations. Subtracting these equations gives

$$h^{(b)} - \widehat{h}^{(b)} = N\, e^{(b)},$$

where the uth component of $e^{(b)}$ can be written as

$$e_u^{(b)} = \sum_v (P_{uv} - \widehat{P}_{uv}) z_v^{(b)}$$

for values $z_v^{(b)} \in [0,1]$: for transient states these are perturbed hitting probabilities, while for absorbing states they are the indicator of membership in $\mathcal{A}_b$. Hence

$$|e_u^{(b)}| \le \sum_v |P_{uv} - \widehat{P}_{uv}| \le \eta_P.$$

It follows that

$$\|h^{(b)} - \widehat{h}^{(b)}\|_\infty \le \|N\|_\infty \eta_P \le h_{\mathrm{abs}} \eta_P.$$

The reachability T_{ab} between types in Eq. equation 9 is an average of $h_u^{(b)}$ over source cells of type a. Averaging cannot increase the maximum error, so

$$\|T - \widehat{T}\|_{\max} \le h_{\mathrm{abs}} \eta_P.$$

Finally, define

$$\phi(t) = \frac{-\log(\max\{t, \delta\})}{-\log \delta}, \qquad t \in [0,1].$$

Both cost matrices use this same function with $0 < \delta < 1$. For $t > \delta$,

$$|\phi'(t)| = \frac{1}{t(-\log \delta)} \le \frac{1}{\delta(-\log \delta)},$$

while ϕ is constant on $[0, \delta]$. Thus ϕ is globally Lipschitz on $[0,1]$ with constant

$$\frac{1}{\delta(-\log \delta)}.$$

Since $C_{ab}^{\mathrm{dev}} = \phi(T_{ab})$,

$$\|C^{\mathrm{dev}} - \widehat{C}^{\mathrm{dev}}\|_{\max} \le \frac{h_{\mathrm{abs}} \eta_P}{\delta(-\log \delta)}.$$

Substituting this bound into Eq. equation S1 proves

$$|D_{ij} - \widehat{D}_{ij}| \le \frac{h_{\mathrm{abs}} \eta_P}{\delta(-\log \delta)} + \mathrm{TV}\left(\boldsymbol{\theta}_i^s, \widehat{\boldsymbol{\theta}}_i^s\right) + \mathrm{TV}\left(\boldsymbol{\theta}_j^t, \widehat{\boldsymbol{\theta}}_j^t\right),$$

which is Eq. equation S2. □

The proof separates three sources of perturbation: cell type composition estimation, developmental geometry between types, and the upstream Markov transition kernel. In particular, the factor h_{abs} captures accumulation of local transition errors along paths with several steps, while δ controls amplification introduced by the negative logarithmic reachability transformation.

C.3 Proof of Invalid Alignment Mass Control

This subsection uses the solver parameters of Appendix B.4.

Theorem 3 (Extended reachability control)**.** *Suppose* $E_0, G_0, D_0 > 0$*,* $\lambda > 0$*, and* $\pi_\lambda \in \Pi(\mathbf{a}, \mathbf{b})$ *satisfies* $F_\lambda(\pi_\lambda) \le F_\lambda(\pi_D) + \xi$ *for* $\xi \ge 0$*. Then*

$$0 \le \langle D, \pi_\lambda \rangle - \langle D, \pi_D \rangle \le \frac{D_0(A_0 + \xi)}{\lambda}. \tag{S3}$$

If optimal OT dual potentials (f, g) *for cost* D *satisfy* $D_{ij} - f_i - g_j \ge \Delta > 0$ *on an invalid set defined by evaluation* $\mathcal{B}_{\text{spot}}$*, then*

$$\sum_{(i,j)\in\mathcal{B}_{\text{spot}}} (\pi_\lambda)_{ij} \le \min\left\{1, \frac{D_0(A_0 + \xi)}{\lambda\Delta}\right\}. \tag{S4}$$

Objective decomposition and comparison conditions. The nondevelopmental objective used in Theorem 3 is

$$F_0(\pi) = (1-\alpha)\frac{\langle M, \pi\rangle}{E_0} + \alpha\omega_{\text{sp}}\frac{\mathcal{G}(\pi)}{G_0} - \varepsilon H(\pi). \tag{S5}$$

Fix a minimizer π_D of the developmental cost over $\Pi(\mathbf{a}, \mathbf{b})$ and define

$$A_0 = F_0(\pi_D) - \min_{\pi\in\Pi(\mathbf{a},\mathbf{b})} F_0(\pi) \ge 0.$$

This quantity measures disagreement between developmental cost and the remaining objective. The comparison condition is

$$F_\lambda(\pi_\lambda) \le F_\lambda(\pi_D) + \xi, \qquad \xi \ge 0. \tag{S6}$$

This condition permits approximate solutions without asserting global convergence of the nonconvex FGW solver. For the ordinary linear OT problem with cost D, optimal dual potentials satisfy $f_i + g_j \le D_{ij}$ for all pairs. The additional margin condition is

$$D_{ij} - f_i - g_j \ge \Delta, \qquad (i,j) \in \mathcal{B}_{\text{spot}}, \qquad \Delta > 0. \tag{S7}$$

It requires the inferred developmental geometry to separate the invalid pairs defined by evaluation after accounting for the marginal constraints. The resulting guarantee controls their total soft coupling mass, rather than hard assignment error computed by row.

Proof of Theorem 3. Since $E_0, G_0, D_0 > 0$, the normalized objectives are defined. With the convention $0 \log 0 = 0$, F_0 is finite and continuous on the nonempty compact polytope $\Pi(\mathbf{a}, \mathbf{b})$. Consequently its minimum is attained, as is the minimum of the linear developmental cost, and A_0 is finite and nonnegative.

Recall that

$$F_\lambda(\pi) = F_0(\pi) + \frac{\lambda}{D_0}\langle D, \pi\rangle,$$

and that

$$\pi_D \in \arg\min_{\pi\in\Pi(\mathbf{a},\mathbf{b})} \langle D, \pi\rangle.$$

Because π_D minimizes developmental cost,

$$\langle D, \pi_\lambda\rangle - \langle D, \pi_D\rangle \ge 0.$$

By the comparison condition in Eq. equation S6,

$$F_\lambda(\pi_\lambda) \le F_\lambda(\pi_D) + \xi.$$

Expanding F_λ gives

$$F_0(\pi_\lambda) + \frac{\lambda}{D_0}\langle D, \pi_\lambda\rangle \le F_0(\pi_D) + \frac{\lambda}{D_0}\langle D, \pi_D\rangle + \xi.$$

Rearranging,

$$\frac{\lambda}{D_0}\left(\langle D, \pi_\lambda\rangle - \langle D, \pi_D\rangle\right) \le F_0(\pi_D) - F_0(\pi_\lambda) + \xi.$$

Since

$$F_0(\pi_\lambda) \ge \min_{\pi \in \Pi(\mathbf{a},\mathbf{b})} F_0(\pi),$$

the definition of A_0 yields

$$F_0(\pi_D) - F_0(\pi_\lambda) \le A_0.$$

Since $\lambda/D_0 > 0$, dividing by this coefficient gives

$$0 \le \langle D, \pi_\lambda\rangle - \langle D, \pi_D\rangle \le \frac{D_0(A_0 + \xi)}{\lambda},$$

which proves Eq. equation S3.

We next translate excess developmental cost into mass assigned to invalid correspondences. Consider the linear optimal transport problem

$$\min_{\pi \in \Pi(\mathbf{a},\mathbf{b})} \langle D, \pi\rangle.$$

Its dual is

$$\max_{f,g} \left\{ \sum_i a_i f_i + \sum_j b_j g_j : f_i + g_j \le D_{ij} \right\}.$$

Let (f, g) be an optimal dual solution. By strong duality,

$$\langle D, \pi_D\rangle = \sum_i a_i f_i + \sum_j b_j g_j.$$

For any feasible π, the marginal constraints imply

$$\begin{aligned}\langle D, \pi\rangle - \langle D, \pi_D\rangle &= \sum_{i,j} D_{ij}\pi_{ij} - \sum_i a_i f_i - \sum_j b_j g_j \\ &= \sum_{i,j} (D_{ij} - f_i - g_j)\pi_{ij}.\end{aligned}$$

Applying this identity to π_λ gives

$$\langle D, \pi_\lambda\rangle - \langle D, \pi_D\rangle = \sum_{i,j} (D_{ij} - f_i - g_j)(\pi_\lambda)_{ij}.$$

Dual feasibility implies

$$D_{ij} - f_i - g_j \ge 0 \qquad \text{for all } i, j.$$

For every $(i, j) \in \mathcal{B}_{\text{spot}}$, the margin assumption in Eq. equation S7 strengthens this to

$$D_{ij} - f_i - g_j \ge \Delta.$$

Consequently,

$$\begin{aligned}\langle D, \pi_\lambda\rangle - \langle D, \pi_D\rangle &\ge \sum_{(i,j)\in\mathcal{B}_{\text{spot}}} (D_{ij} - f_i - g_j)(\pi_\lambda)_{ij} \\ &\ge \Delta \sum_{(i,j)\in\mathcal{B}_{\text{spot}}} (\pi_\lambda)_{ij}.\end{aligned}$$

Combining this lower bound with Eq. equation S3 yields

$$\sum_{(i,j)\in\mathcal{B}_{\text{spot}}} (\pi_\lambda)_{ij} \le \frac{D_0(A_0 + \xi)}{\lambda\Delta}.$$

Finally, because π_λ is a coupling between probability marginals,

$$\sum_{i,j} (\pi_\lambda)_{ij} = 1,$$

so the invalid mass is also trivially bounded above by 1. Hence

$$\sum_{(i,j)\in\mathcal{B}_{\text{spot}}} (\pi_\lambda)_{ij} \le \min\left\{1, \frac{D_0(A_0+\xi)}{\lambda\Delta}\right\},$$

which proves Eq. equation S4. For the asymptotic statements, fix the costs, marginals, reference coupling, choice of π_D, and the remaining objective parameters. Then D_0 and A_0 are independent of λ. If ξ is uniformly bounded, Eq. equation S3 gives an $O(\lambda^{-1})$ upper bound; with a fixed $\Delta > 0$, Eq. equation S4 gives an $O((\lambda\Delta)^{-1})$ upper bound. Neither argument requires the actual errors of approximate solutions to decrease monotonically. □

The exact optimizer is obtained by setting $\pi_\lambda = \pi^\star$ and $\xi = 0$ in Theorem 3.

The proof separates three factors governing the final guarantee. The quantity A_0 measures disagreement between developmental cost and the remaining alignment objective, ξ captures optimization error relative to π_D, and Δ measures how strongly invalid correspondences are separated by the developmental geometry after accounting for the marginal constraints. With the costs, marginals, reference coupling, choice of π_D, and other objective parameters fixed, a uniformly bounded ξ and a fixed $\Delta > 0$ yield an $O((\lambda\Delta)^{-1})$ upper bound on invalid coupling mass.

D Evaluation Metrics

We evaluate each interpolated or extrapolated spatial transcriptomic slice against the withheld observation at the same stage. Three primary prediction metrics measure expression distribution, cellular composition, and spatial organization; four lineage metrics evaluate strongest links and the full soft coupling between stages. The PCA transformation, evaluation sampling rules, and neighborhood construction rules are shared across methods; each predicted slice has its own spatial graph. Unless explicitly stated otherwise, lower is better.

Evaluation versions. The main prediction tables and their aggregate and target-specific appendix counterparts use the reevaluated joint MMD, corrected Moran's I, and deterministic cell-type JS described below. Full measured library totals are used before selecting the training-fitted genes when transforming withheld counts. These evaluations reuse the saved predictions without retraining. The component ablations use the same four lineage definitions as Table 1 and the same three primary prediction definitions. Lineage scores are evaluated on the saved couplings; prediction scores are read from the existing evaluation results.

D.1 Transcriptomic Distribution Fidelity

Scaled FGW calibrated against random couplings. Let $\widehat{X}$ and X be the predicted and withheld PCA coordinates and let $\widehat{S}$ and S be their spatial coordinates. We form the squared cost between expression profiles $M_{ij} = \|\widehat{X}_i - X_j\|_2^2$ and the Euclidean distance matrices $\widehat{C}$ and C within each slice. Uniform marginals define the feasible set $\Pi(a,b)$. A seeded positive random matrix is projected onto $\Pi(a,b)$ by iterative proportional fitting, producing π_0. To make all methods in a task directly comparable, the calibration denominators are computed once from the FGW OT prediction:

$$E_0 = \left\langle M^{\text{ref}}, \pi_0 \right\rangle, \qquad G_0 = \mathcal{L}_{\text{GW}}\left(\widehat{C}^{\text{ref}}, C, \pi_0\right). \tag{S8}$$

Every method is then evaluated with the same E_0 and G_0 by solving

$$\pi^\star = \arg\min_{\pi\in\Pi(a,b)} (1-\alpha_e)\frac{\langle M,\pi\rangle}{E_0} + \alpha_e\omega_e\frac{\mathcal{L}_{\text{GW}}(\widehat{C},C,\pi)}{G_0}, \tag{S9}$$

and reporting the square root of the optimized objective,

$$d_{\text{sFGW}} = \left[(1-\alpha_e)\frac{\langle M,\pi^\star\rangle}{E_0} + \alpha_e\omega_e\frac{\mathcal{L}_{\text{GW}}(\widehat{C},C,\pi^\star)}{G_0}\right]^{1/2}. \tag{S10}$$

The same withheld subsample and random coupling seed are used for every method in a task. The evaluation weights α_e and ω_e for each dataset are listed in Table S1.

Joint expression–neighborhood MMD. Our primary Maximum Mean Discrepancy compares the joint distribution of a spot's PCA expression x_i and its local expression context h_i (Gretton et al., 2012). Within each slice separately, let $h_i = |\mathcal{N}_6(i)|^{-1} \sum_{j \in \mathcal{N}_6(i)} x_j$. The neighborhood excludes the spot itself and includes all observations tied at the sixth-nearest distance in native spatial coordinates. Saved per-axis coordinate normalization is undone before constructing this graph; duplicate observations retain their multiplicity. For $z = (x, h)$, the kernel is

$$k_\lambda(z, z') = \exp\left(-\frac{\|x - x'\|_2^2}{2\sigma_x^2}\right) \exp\left(-\lambda_{\mathrm{MMD}} \frac{\|h - h'\|_2^2}{2\sigma_h^2}\right), \qquad \lambda_{\mathrm{MMD}} = 1. \tag{S11}$$

This evaluation weight is distinct from the developmental weight in the alignment objective. Given the joint distributions P and $\widehat{P}$, the squared MMD is

$$\mathrm{MMD}^2(P, \widehat{P}) = \mathbb{E}_{x,x' \sim P} k(x, x') + \mathbb{E}_{y,y' \sim \widehat{P}} k(y, y') - 2\mathbb{E}_{x \sim P, y \sim \widehat{P}} k(x, y). \tag{S12}$$

Here the arguments in Eq. equation S12 are joint states z. We remove within-sample diagonal kernel entries and report $\sqrt{\max(\widehat{\mathrm{MMD}}_U^2, 0)}$, retaining the unclamped squared estimates in the result files. The usual unbiasedness of the U estimator requires independent samples and is not asserted for coupled predictions or neighborhood-derived features.

The six-neighbor representation is invariant to rigid coordinate changes and does not require anatomical registration, but it cannot identify globally relocated patches with identical local context. Sampling density and repeated predicted states can affect the neighborhoods.

D.2 Cellular Composition and Developmental Fidelity

Cell Type Composition JS Divergence. We compare observed and predicted aggregate cell-type proportions using Jensen–Shannon (JS) divergence (Lin, 1991). Let $\theta_i \in \Delta^{K-1}$ be the composition of observed spot i. For predictor $q \in \{\mathrm{Linear}, \mathrm{Flow}\}$, let N_q be the number of saved prediction samples and $\widehat{\theta}_j^{(q)}$ the composition propagated from sample j's endpoints using the rule in Appendix B.5. Each sample has equal weight $1/N_q$, giving

$$p = \frac{1}{n} \sum_{i=1}^{n} \theta_i, \qquad \widehat{p} = \frac{1}{N_q} \sum_{j=1}^{N_q} \widehat{\theta}_j^{(q)}. \tag{S13}$$

This empirical aggregate is deterministic for a fixed saved sample set. Linear and Flow use the same linear propagation, nonnegative clipping, and row normalization rule; their saved prediction samples can differ, yielding different aggregate compositions and cell-type JS values. Writing $m = (p + \widehat{p})/2$, we compute

$$\mathrm{JS}_{\mathrm{cell}} = \frac{\frac{1}{2} D_{\mathrm{KL}}(p \| m) + \frac{1}{2} D_{\mathrm{KL}}(\widehat{p} \| m)}{\log 2}. \tag{S14}$$

D.3 Lineage-Graph Alignment Metrics

For each dataset d, we encode the literature-supported biological relationships in Fig. S1 as a directed reference lineage graph $\mathcal{G}_d = (V_d, E_d)$. The graph topology is compiled from the corresponding human atlas and supporting developmental literature (Olaniru et al., 2023; Jennings et al., 2015; Fawkner-Corbett et al., 2021a; Elmentaite et al., 2021; Zhang et al., 2024; Huang, 2017; Liu et al., 2025; Peña & Martin, 2024). Its nodes contain all cell types used in the alignment evaluation together with biologically necessary branching states. Contextual states that lie only along an unbranched path and are not evaluated are contracted, whereas genuine branch points and informative terminal branches are retained. Distinct major lineage compartments remain separated. For skin wound, $\mathcal{G}_d$ represents repair-lineage and repair-state relationships rather than embryonic ancestry. The graph and the mapping from analysis labels to graph nodes are fixed for each dataset and shared across all methods and ablations. Model fitting follows the benchmark configuration in Appendix F.1; the graph is subsequently applied to the inferred couplings to compute the lineage-based evaluation scores.

For source type a, let $\mathcal{R}(a)$ contain its admissible downstream states, including itself. The fixed raw deviation ℓ_{ab} is zero for $b \in \mathcal{R}(a)$, the minimum undirected gragh distance from b to $\mathcal{R}(a)$

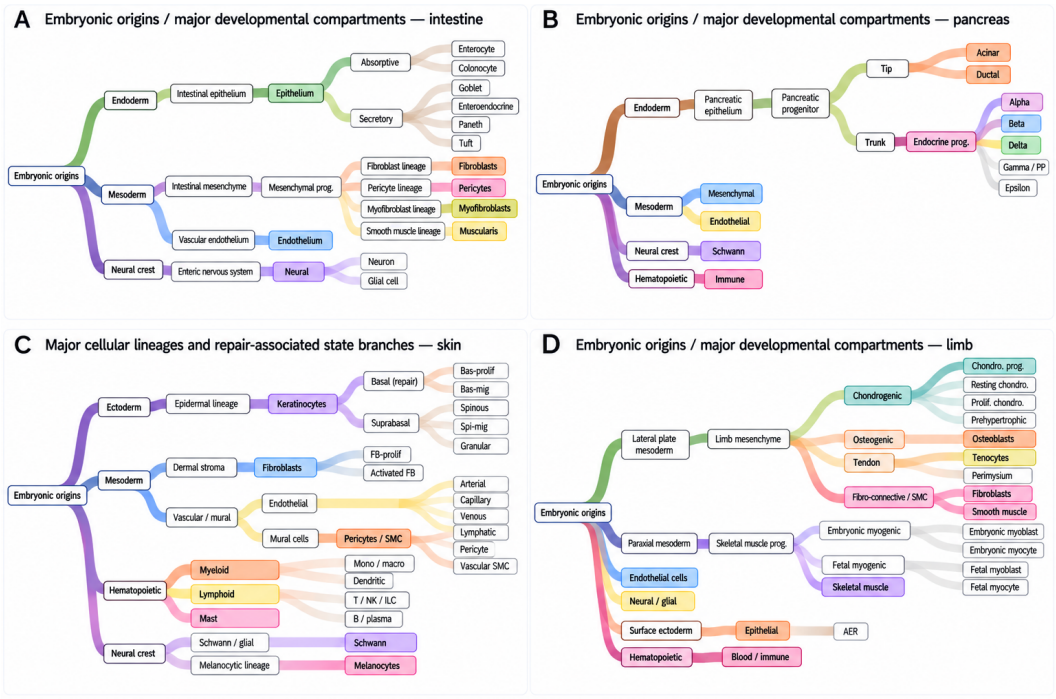


Figure S1: **Literature-supported reference lineage graphs used for biological evaluation.** Reference graphs are shown for **(A)** developing human intestine, **(B)** fetal human pancreas, **(C)** human acute skin wound repair, and **(D)** human embryonic limb. Nodes represent major cell types or lineage states, and directed edges denote literature-supported developmental or repair-associated relationships. Colored nodes indicate cell types used in the alignment evaluation. These reference graphs are given by the corresponding atlases and biological literature.

within a lineage compartment, and a fixed maximal penalty $D_\perp$ for cross-compartment pairs. We use $C^{\mathrm{lin}}_{ab} = \ell_{ab}/D_\perp \in [0,1]$, with $D_\perp = 5, 3, 5, 3$ for pancreas, intestine, limb, and skin wound, respectively. This matrix is generally asymmetric because the admissible region depends on the source type.

Let a_i, b_j be the mapped dominant source and target types and q^s_i, q^t_j the corresponding dominant fractions. Define $E_{ij} = 1$ when both types are mapped and both fractions are at least the dataset threshold, and zero otherwise. The thresholds are $\tau = 0.5$ for pancreas and skin wound and $\tau = 0.4$ for intestine and limb. Unmapped types remain unevaluable; in particular, the broad limb mesenchymal label is not assigned to a narrower lineage state, and skin hair follicle labels are excluded. For each source with positive row mass, retain its saved strongest target

$$j^\star(i) = \arg\max_j \pi_{ij}, \qquad \mathcal{J} = \{i : \textstyle\sum_j \pi_{ij} > 0,\ E_{i,j^\star(i)} = 1\}. \tag{S15}$$

Saved top-link indices preserve the original tie decision when archived coupling precision rounds two maxima to the same value.

Lineage Consistent Rate (LCR) and Mean Lineage Deviation Distance (MLDD). Let $\mathcal{S} = \{i \in \mathcal{J} : C^{\mathrm{lin}}_{a_i b_{j^\star(i)}} = 0\}$ be the supported evaluable links. We report

$$\mathrm{LCR} = \frac{|\mathcal{S}|}{|\mathcal{J}|}, \qquad \mathrm{MLDD} = \frac{1}{|\mathcal{J}|}\sum_{i\in\mathcal{J}} C^{\mathrm{lin}}_{a_i b_{j^\star(i)}}. \tag{S16}$$

Higher LCR and lower MLDD are better; both condition on evaluable links.

Coupling-weighted Lineage Deviation (CWLD). The soft-coupling score uses the same matrix and endpoint eligibility:

$$\mathrm{CWLD} = \frac{\sum_{i,j:E_{ij}=1} \pi_{ij} C^{\mathrm{lin}}_{a_i b_j}}{\sum_{i,j} \pi_{ij} E_{ij}}. \tag{S17}$$

Lower CWLD is better. It measures deviation conditional on evaluable mass.

Lineage-Supported Coverage (LSC) and evaluability. With n_s the total number of source spots, including zero-mass rows,

$$\mathrm{LSC} = \frac{|\mathcal{S}|}{n_s}, \qquad \mathrm{Coverage}_{\mathrm{top}} = \frac{|\mathcal{J}|}{n_s}, \qquad \mathrm{Coverage}_{\mathrm{mass}} = \frac{\sum_{i,j}\pi_{ij}E_{ij}}{\sum_{i,j}\pi_{ij}}. \tag{S18}$$

Higher LSC is better. For each run and stage pair, $\mathrm{LSC} = \mathrm{LCR}\ \mathrm{Coverage}_{\mathrm{top}}$; we aggregate the per-pair LSC values rather than multiplying aggregate means. All four lineage metrics lie in $[0,1]$. LCR and MLDD are undefined if $|\mathcal{J}| = 0$, and CWLD is undefined when evaluable coupling mass is zero; coverage is reported alongside these conditional metrics.

D.4 Spatial Organization Fidelity

Cell Type Neighborhood JS Divergence. Matching global cell type proportions does not guarantee that the populations are arranged correctly within the tissue. We therefore evaluate local spatial organization by comparing cell type neighborhood distributions. Spatial neighborhood analysis and joint occurrence of cell types are widely used for characterizing local organization in spatial omics data (Palla et al., 2022). For each slice, we construct a spatial graph $\mathcal{E}$ that connects the k nearest neighbors, using the same neighborhood size across methods. Rather than assigning each mixed spot to a single dominant type, we retain its complete cell type composition. Each undirected edge (i,j) contributes

$$\theta_i\theta_j^\top + \theta_j\theta_i^\top \tag{S19}$$

to a $K \times K$ interaction matrix between cell types, giving

$$N = \sum_{(i,j)\in\mathcal{E}} \left(\theta_i\theta_j^\top + \theta_j\theta_i^\top\right). \tag{S20}$$

After normalizing N to sum to one, we flatten the interaction matrix into a probability distribution over neighboring pairs of cell types and compute the normalized JS divergence between the withheld and interpolated slices (Lin, 1991). This metric is insensitive to the absolute orientation of the tissue and specifically measures whether the local spatial relationships between cellular populations are reproduced.

Moran's I Error. We next evaluate spatial organization at the gene-expression level using Moran's I, a classical measure of spatial autocorrelation. For gene g, Moran's I is defined as

$$I_g = \frac{n}{W} \frac{\sum_{i,j} w_{ij}(x_{ig} - \bar{x}_g)(x_{jg} - \bar{x}_g)}{\sum_i (x_{ig} - \bar{x}_g)^2}, \qquad W = \sum_{i,j} w_{ij}, \tag{S21}$$

where w_{ij} is the binary symmetric adjacency constructed separately for each predicted or withheld slice in the saved per-axis-normalized coordinates. The implementation queries seven nearest entries ($k + 1$, with $k = 6$), removes self-indices, and symmetrizes the remaining edges. With duplicate coordinates, a query can omit its own index and retain seven nonself entries; this differs from the distance-tie-inclusive native-coordinate graph used for MMD. The construction rule is shared across methods, and weights are not standardized by row. Both predicted and withheld expression are reconstructed in the same training-fitted gene space by inverse PCA transform. Repeated predicted states retain multiplicity; duplicate-collapse sensitivity is retained in the accompanying result files. Genes with nearly zero variance in either the withheld or predicted slice are discarded, and the remaining shared genes define

$$\mathrm{MoranError} = \sqrt{\frac{1}{G} \sum_{g=1}^{G} \left(I_g^{\mathrm{pred}} - I_g^{\mathrm{true}} \right)^2}. \tag{S22}$$

Thus the reported value is the root mean squared difference between the two Moran vectors, not an error after fitting a gene subset chosen separately for each method.

Gromov–Wasserstein Distance (GW). Finally, we assess global spatial organization using the Gromov–Wasserstein discrepancy, which compares relational geometry without requiring observations from the two domains to share identical coordinates or explicit pointwise correspondence. Let C^{true} and C^{pred} denote the pairwise Euclidean spatial-distance matrices of the held-out and interpolated slices, respectively. Each matrix is divided by the median of its positive entries. Both slices are deterministically subsampled to at most 2,000 spots, and uniform empirical marginals a and b define

$$\mathrm{GW}^2 = \min_{\gamma \in \Pi(a,b)} \sum_{i,k,j,l} \left(C_{ik}^{\mathrm{true}} - C_{jl}^{\mathrm{pred}} \right)^2 \gamma_{ij} \gamma_{kl}. \tag{S23}$$

We report $\sqrt{\mathrm{GW}^2}$, clipped at zero before taking the square root. The resulting metric is invariant to translation, rotation, and reflection and is calibrated for coordinate scale by the median within each slice normalization.

Together, archived scaled FGW and the primary joint MMD describe reconstruction fidelity under their stated evaluation versions; cell type JS measures population abundance; neighborhood JS and Moran's I error measure local cellular and expression organization; GW measures global tissue geometry; and the statistics from judged strongest links evaluate developmental plausibility directly.

E DATASETS

Developing human pancreas. We use the human fetal pancreas dataset from Olaniru et al. (Olaniru et al., 2023), with scRNA-seq and 10x Visium data available under GEO accessions `GSE197064` and `GSE197317`, respectively. We focus on the common developmental stages at 15, 18, and 20 post-conception weeks (PCW), with two spatial sections available at each stage. The scRNA-seq data provide the single-cell developmental reference, including spliced and unspliced counts used to infer developmental dynamics, whereas the Visium sections provide the spatial observations to be aligned across time. Although both modalities originate from the same study, tissue, and developmental stages, the public metadata do not establish donor-level or cell-level correspondence. We therefore treat them as stage-matched but cross-modally unpaired observations.

Developing human intestine. The fetal intestine dataset is derived from Fawkner-Corbett et al. (Fawkner-Corbett et al., 2021a). The single cell RNA sequencing data (GSE158702) cover approximately 8–22 PCW and characterize diverse epithelial, mesenchymal, immune, neural, and vascular populations during intestinal development. The spatial transcriptomic data (GSE158328) include fetal colon and terminal ileum sections at 12 and 19 PCW. Additional processed annotations are available through the authors' Mendeley repository (Fawkner-Corbett et al., 2021b) and CELLxGENE. The relatively dense temporal coverage of the single cell atlas provides intermediate developmental states between the more sparsely sampled spatial measurements, allowing us to evaluate whether developmental dynamics from single cells can support spatial alignment and reconstruction across time. The two modalities are treated as population measurements rather than pointwise matched observations.

Developing human limb. We use the human embryonic limb atlas of Zhang et al. (Zhang et al., 2024). Human scRNA-seq data are available under ArrayExpress accession E-MTAB-8813, and 10x Visium data under E-MTAB-10367, with processed data and annotations additionally provided through the accompanying embryonic-limb data portal. The single-cell atlas spans approximately PCW5–PCW9, while the spatial measurements cover PCW5–PCW8. During this interval, the limb undergoes rapid lineage diversification and large-scale morphogenesis, including the emergence and spatial organization of skeletal, muscular, neural, vascular, and connective-tissue compartments. This pronounced change in both cellular composition and tissue geometry makes the dataset a challenging test of cross-time developmental alignment.

Human acute skin wound healing. We additionally evaluate our framework on the human acute skin wound atlas of Liu et al. (Liu et al., 2025). The dataset profiles intact skin and wounds collected at days 1, 7, and 30 after injury, representing distinct stages of the wound-healing process. The scRNA-seq data are deposited under GSE241132, while the spatial transcriptomic data are available under GSE241124; raw sequencing files used in our analysis were obtained through controlled access from the European Genome-phenome Archive (EGA), study EGAS50000000571, dataset EGAD50000000813, under the corresponding data-access agreement. The scRNA-seq series contains samples from three donors and the spatial series from four donors across the four time points. We therefore use their shared temporal ordering without assuming complete donor-level cross-modality matching. Unlike the developmental datasets above, this dataset captures injury-induced tissue regeneration and provides an additional setting for evaluating spatiotemporal alignment beyond embryonic development.

F Additional Experiments

F.1 Experimental Protocol

All methods use the same observations from each stage, PCA representations fitted on training data, and downstream predictors. Endpoint sampling uses a shared random seed and each method's own coupling. The cell-type JS aggregate is the equally weighted mean of propagated compositions over each predictor's saved samples (Appendix D.2). The benchmark order is Ordinary OT, FGW OT, PASTE, PASTE2, DeST-OT, and DRIFT. The reporting protocol for the three primary tables is specified in Appendix F.2. Flow extrapolation uses IVP integration. For alignment, pancreas, intestine, limb, and skin contain two, two, three, and five evaluated stage pairs, respectively; the alignment table reports sample variation across those stage-pair values. Ablation prediction summaries average targets within each seed, regime, and predictor. All result tables below preserve unrounded values in CSV exports for machine reading. Most show three decimal places in print. A nonzero standard deviation below the rounding threshold is shown as < 0.001. Rankings in the primary tables use unrounded values. For reevaluated cell-type JS, differences within 10^{-8} are treated as ties.

We select highly variable genes on the ordered training stages, normalize full measured library sizes to 10^4 before restricting to those genes, apply $\log(1 + x)$, standardize features, and fit a shared PCA with 30 dimensions. The entropy fraction is 0.05 and the outer FGW solver uses at most 100 iterations in every dataset. Flow matching is trained for 1,000 optimization steps and integrated with 40 fourth order Runge–Kutta steps. The inner cell type OT regularization is 0.05. The common preprocessing and solver settings, along with dataset-specific prediction and evaluation settings, are listed in Table S1.

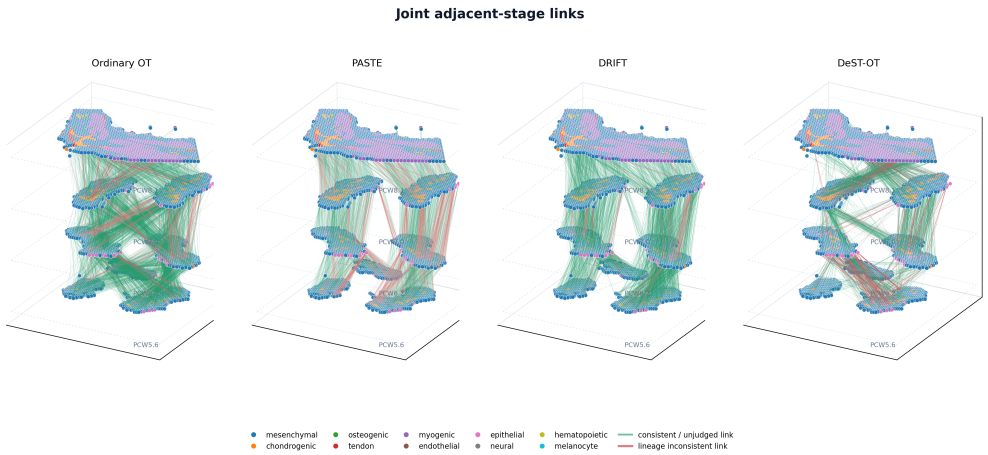


Figure S2: **Joint cross-time spot alignments on the developing limb dataset.** We visualize the spot-to-spot correspondences inferred by Ordinary OT, PASTE, DeST-OT, and DRIFT across four consecutive developmental stages, PCW5.6 → PCW6.2 → PCW7.0 → PCW8.1. Each line represents an inferred OT correspondence between spots in adjacent stages. Green lines denote lineage-consistent or unjudged links, whereas red lines denote lineage-inconsistent links. Evaluable-link coverage is reported in Table S5; green links include correspondences outside this evaluation subset. Spot colors indicate dominant cell types.

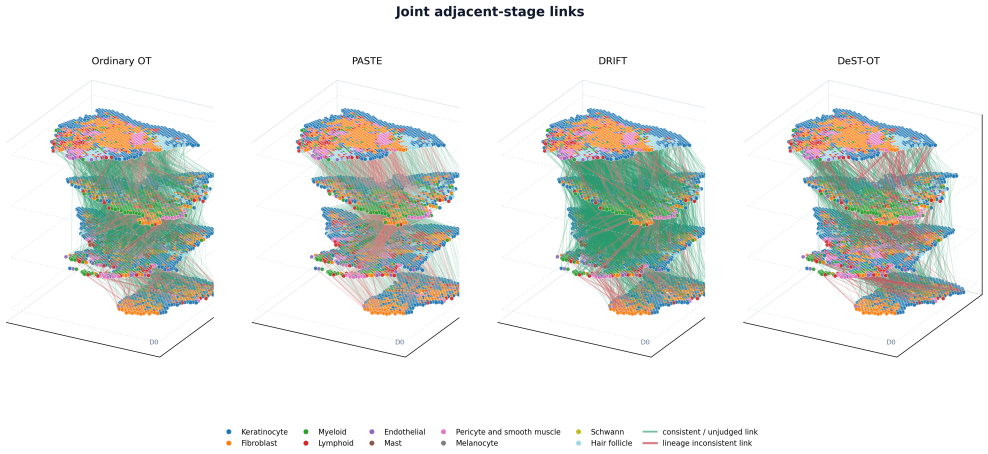


Figure S3: **Joint cross-time spot alignments on the human acute skin wound dataset.** We visualize the spot-to-spot correspondences inferred by Ordinary OT, PASTE, DeST-OT, and DRIFT across four consecutive stages, D0 → D1 → D7 → D30. Each line represents an inferred OT correspondence between spots in adjacent stages. Green lines denote lineage-consistent or unjudged links, whereas red lines denote lineage-inconsistent links. Evaluable-link coverage is reported in Table S5; green links include correspondences outside this evaluation subset. Spot colors indicate dominant cell types.

Table S1: Experimental settings by dataset. "Pred." is the number of predicted points, "Flow pairs" is the number of coupling samples used for flow training, $(\alpha, \omega_{\mathrm{sp}}, \lambda)$ are the DRIFT solver parameters defined in Appendix B.4, "Inner" is the number of inner composition OT iterations, (α_e, ω_e) are the scaled FGW evaluation weights, and τ is the dominant type threshold.

Dataset	HVGs	Pred.	Flow pairs	α	ω_{sp}	λ	Inner	α_e	ω_e	τ
Pancreas	2,000	2,000	20,000	0.5	1	1.00	100	0.5	1	0.5
Intestine	2,000	3,000	40,000	0.5	1	1.00	80	0.5	20	0.4
Limb	2,000	3,000	20,000	0.5	1	1.00	100	0.6	8	0.4
Skin wound	2,000	3,000	5,000	0.5	1	1.00	200	0.6	20	0.5

F.2 BENCHMARK RESULTS

Tables S2–S4 show the dataset-level results behind the main-text tables. Table S5 shows alignment scores for each stage pair. Alignment uses the four lineage metrics and reports top-link and coupling-mass coverage. Stage pairs receive equal weight, with sample SD across stage-pair means in the dataset-level alignment table. Stage-pair scores show mean and sample SD across ten runs. Within each comparable dataset and condition, bold and underlined entries denote the best and second-best unrounded metric means, respectively; tied ranks are unmarked. Prediction tables show mean and sample SD after averaging targets within each run and displayed condition. Interpolation combines the 6.2 and 7.0 PCW targets for limb and the D1 and D7 targets for skin wound; all other prediction groups contain one target.

Stage-level rankings vary within these dataset averages. For limb PCW6.2 → PCW7.0, Ordinary OT has higher LCR and lower MLDD and CWLD than DRIFT. For skin D1 → D30, Ordinary OT has higher LCR and lower MLDD, and DRIFT has the highest LSC among the evaluated methods (Table S5).

Limb reconstruction illustration. Figure 4 shows a separate linear reconstruction of PCW6.2 from PCW5.6 and PCW7.2. The limb benchmark uses the stage sequence PCW5.6, PCW6.2, PCW7.0, and PCW8.1. The figure's nearest-neighbour expression similarity colours provide a qualitative overlay; quantitative comparisons use the MMD, cell-type JS, and Moran's I errors reported in the prediction tables.

Table S2: **Alignment results.** Cells report mean and sample SD across stage-pair means. Top coverage divides evaluable strongest links by all source spots; mass coverage divides evaluable coupling mass by total coupling mass. Unmapped labels are excluded from evaluation, including broad limb mesenchymal labels and skin hair follicle labels.

Pancreas Method	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	Top coverage	Mass coverage
Ordinary OT	0.539±0.066	0.449±0.084	0.557±0.209	0.270±0.031	0.502±0.004	0.485±0.002
FGW OT	0.432±0.115	0.555±0.132	0.657±0.151	0.217±0.063	0.501±0.012	0.478±0.003
PASTE	0.239±0.093	0.748±0.111	0.749±0.106	0.112±0.046	0.466±0.011	0.466±0.010
PASTE2	0.268±0.116	0.719±0.134	0.722±0.130	0.078±0.041	0.283±0.030	0.493±0.034
DeST-OT	0.586±0.173	0.406±0.183	0.458±0.245	0.300±0.094	0.511±0.009	0.501±0.008
DRIFT	**0.845±0.160**	**0.152±0.164**	**0.284±0.305**	**0.485±0.117**	**0.571±0.031**	**0.548±0.029**
Intestine Method	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	Top coverage	Mass coverage
Ordinary OT	0.354±0.321	0.593±0.283	0.611±0.267	0.267±0.222	0.797±0.095	0.805±0.099
FGW OT	0.358±0.330	0.583±0.300	0.642±0.229	0.263±0.221	0.785±0.108	0.805±0.098
PASTE	0.220±0.157	0.711±0.141	0.708±0.142	0.168±0.104	0.800±0.099	0.801±0.099
PASTE2	0.243±0.203	0.687±0.170	0.675±0.150	0.101±0.066	0.464±0.119	0.771±0.087
DeST-OT	0.386±0.367	0.560±0.330	**0.575±0.333**	0.295±0.261	**0.810±0.092**	0.810±0.094
DRIFT	**0.573±0.272**	**0.384±0.247**	0.599±0.270	**0.433±0.157**	0.778±0.097	**0.811±0.093**
Limb Method	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	Top coverage	Mass coverage
Ordinary OT	0.877±0.198	0.116±0.185	0.202±0.333	0.070±0.030	0.077±0.020	0.075±0.009
FGW OT	0.843±0.251	0.147±0.234	0.276±0.429	0.067±0.022	0.079±0.008	0.070±0.015
PASTE	0.425±0.244	0.572±0.240	0.573±0.230	0.015±0.001	0.051±0.043	0.052±0.045
PASTE2	0.393±0.253	0.603±0.249	0.593±0.244	0.009±0.003	0.037±0.038	0.052±0.039
DeST-OT	0.839±0.225	0.155±0.214	**0.158±0.229**	0.063±0.029	0.072±0.019	0.072±0.025
DRIFT	**0.888±0.171**	**0.092±0.137**	0.161±0.245	**0.076±0.027**	**0.084±0.016**	**0.080±0.005**
Skin wound Method	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	Top coverage	Mass coverage
Ordinary OT	0.623±0.104	0.375±0.104	0.446±0.135	0.175±0.034	0.288±0.079	0.285±0.092
FGW OT	0.561±0.095	0.435±0.097	0.532±0.130	0.158±0.045	0.287±0.092	0.279±0.092
PASTE	0.332±0.229	0.666±0.229	0.671±0.231	0.089±0.068	0.264±0.099	0.264±0.095
PASTE2	0.268±0.163	0.729±0.163	0.727±0.163	0.047±0.031	0.172±0.085	0.257±0.099
DeST-OT	0.653±0.123	0.343±0.124	0.387±0.142	0.185±0.040	0.289±0.076	0.290±0.080
DRIFT	**0.682±0.178**	**0.316±0.179**	**0.357±0.173**	**0.237±0.045**	**0.366±0.107**	**0.339±0.112**

Table S3: **Interpolation results.** Cells show mean and sample SD over ten runs after averaging targets within each run and displayed condition. The three metrics use the primary evaluator; lower is better. Each dataset is shown beneath both predictor headings.

	Linear			**Flow matching**		
		Pancreas			Pancreas	
Method	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS
Ordinary OT	0.712 ± 0.115	0.186 ± 0.003	0.023 ± 0.003	0.563 ± 0.005	0.345 ± 0.159	0.022 ± 0.002
FGW OT	0.641 ± 0.130	0.197 ± 0.006	0.024 ± 0.003	0.600 ± 0.088	0.322 ± 0.179	0.023 ± 0.003
PASTE	0.750 ± 0.066	$\mathbf{0.143 \pm 0.005}$	0.021 ± 0.000	0.667 ± 0.100	0.212 ± 0.146	0.022 ± 0.002
PASTE2	0.646 ± 0.092	0.282 ± 0.172	0.022 ± 0.001	0.566 ± 0.018	0.460 ± 0.261	0.025 ± 0.003
DeST-OT	0.544 ± 0.002	0.149 ± 0.010	0.024 ± 0.003	0.556 ± 0.003	0.286 ± 0.165	0.022 ± 0.002
DRIFT	$\mathbf{0.543 \pm 0.005}$	0.154 ± 0.015	$\mathbf{0.021 \pm 0.001}$	$\mathbf{0.550 \pm 0.002}$	$\mathbf{0.122 \pm 0.002}$	$\mathbf{0.021 \pm 0.001}$

	Linear			**Flow matching**		
		Intestine			Intestine	
Method	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS
Ordinary OT	0.621 ± 0.078	0.430 ± 0.005	$0.128 \pm <0.001$	0.454 ± 0.021	0.228 ± 0.094	0.128 ± 0.002
FGW OT	0.658 ± 0.036	0.393 ± 0.003	$0.128 \pm <0.001$	0.476 ± 0.060	0.248 ± 0.086	0.128 ± 0.002
PASTE	0.484 ± 0.061	$\mathbf{0.222 \pm 0.046}$	0.129 ± 0.002	0.458 ± 0.001	0.239 ± 0.091	0.128 ± 0.002
PASTE2	0.484 ± 0.008	0.359 ± 0.142	0.137 ± 0.009	0.506 ± 0.060	0.346 ± 0.132	0.136 ± 0.009
DeST-OT	0.665 ± 0.003	0.358 ± 0.006	$0.127 \pm <0.001$	0.539 ± 0.087	0.256 ± 0.110	0.128 ± 0.002
DRIFT	$\mathbf{0.469 \pm 0.005}$	0.339 ± 0.032	$\mathbf{0.122 \pm 0.008}$	$\mathbf{0.443 \pm 0.006}$	$\mathbf{0.198 \pm 0.006}$	$\mathbf{0.120 \pm 0.010}$

	Linear			**Flow matching**		
		Limb			Limb	
Method	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS
Ordinary OT	0.427 ± 0.076	0.377 ± 0.010	0.050 ± 0.022	0.399 ± 0.074	0.194 ± 0.003	0.050 ± 0.022
FGW OT	0.442 ± 0.076	0.363 ± 0.007	0.044 ± 0.022	0.377 ± 0.002	0.190 ± 0.001	0.056 ± 0.020
PASTE	0.387 ± 0.001	$\mathbf{0.191 \pm 0.052}$	0.033 ± 0.015	0.414 ± 0.057	0.190 ± 0.004	0.050 ± 0.022
PASTE2	0.428 ± 0.067	0.238 ± 0.105	0.039 ± 0.014	0.421 ± 0.034	0.188 ± 0.067	0.039 ± 0.014
DeST-OT	$\mathbf{0.383 \pm 0.002}$	0.308 ± 0.007	0.033 ± 0.015	0.368 ± 0.002	0.193 ± 0.003	0.045 ± 0.021
DRIFT	0.393 ± 0.005	0.289 ± 0.032	$\mathbf{0.026 \pm <0.001}$	$\mathbf{0.368 \pm 0.004}$	$\mathbf{0.182 \pm 0.007}$	$\mathbf{0.027 \pm <0.001}$

	Linear			**Flow matching**		
		Skin wound			Skin wound	
Method	MMD	Moran I	Cell JS	MMD	Moran I	Cell JS
Ordinary OT	0.311 ± 0.001	0.422 ± 0.005	0.071 ± 0.011	0.338 ± 0.003	0.360 ± 0.061	0.057 ± 0.001
FGW OT	0.325 ± 0.002	0.471 ± 0.006	0.071 ± 0.011	0.328 ± 0.004	0.348 ± 0.058	0.057 ± 0.001
PASTE	0.329 ± 0.003	$\mathbf{0.200 \pm 0.002}$	0.071 ± 0.011	0.336 ± 0.004	0.273 ± 0.033	0.065 ± 0.012
PASTE2	0.314 ± 0.009	0.332 ± 0.124	0.060 ± 0.007	0.326 ± 0.005	0.303 ± 0.088	0.063 ± 0.010
DeST-OT	0.311 ± 0.002	0.412 ± 0.004	0.065 ± 0.012	0.331 ± 0.002	0.364 ± 0.058	0.057 ± 0.001
DRIFT	$\mathbf{0.303 \pm 0.005}$	0.313 ± 0.055	$\mathbf{0.055 \pm 0.001}$	$\mathbf{0.307 \pm 0.013}$	$\mathbf{0.233 \pm 0.005}$	$\mathbf{0.056 \pm <0.001}$

Table S4: **Extrapolation results.** Cells show mean and sample SD over ten runs after averaging targets within each run and displayed condition. The three metrics use the primary evaluator; lower is better. Each dataset is shown beneath both predictor headings. Flow matching uses IVP integration.

	Linear			Flow matching		
Method	MMD	Pancreas Moran I	Cell JS	MMD	Pancreas Moran I	Cell JS
Ordinary OT	0.755±0.003	0.237±0.008	0.028±0.009	0.756±0.006	0.269±0.114	0.022±0.008
FGW OT	0.717±0.004	0.199±0.006	0.027±0.009	0.724±0.003	0.254±0.117	0.023±0.008
PASTE	**0.693±0.001**	0.178±0.020	0.022±0.008	**0.698±0.002**	0.266±0.134	0.024±0.008
PASTE2	0.695±0.002	0.281±0.019	0.037±0.009	0.724±0.022	0.301±0.103	0.032±0.007
DeST-OT	0.721±0.001	0.220±0.005	0.028±0.009	0.716±0.006	0.254±0.112	0.024±0.009
DRIFT	0.705±0.003	**0.174±0.011**	**0.013±<0.001**	0.709±0.007	**0.182±0.003**	**0.013±<0.001**
	Linear			**Flow matching**		
Method	MMD	Intestine Moran I	Cell JS	MMD	Intestine Moran I	Cell JS
Ordinary OT	0.690±0.002	0.377±0.005	0.391±0.000	0.732±0.002	0.348±0.125	0.394±0.003
FGW OT	0.688±0.001	0.393±0.007	0.379±<0.001	0.731±0.001	0.356±0.123	0.381±0.003
PASTE	0.708±0.002	0.414±0.086	**0.374±0.009**	0.731±0.001	0.472±0.114	**0.372±0.008**
PASTE2	0.711±0.003	0.524±0.028	0.388±0.010	0.732±0.001	0.537±0.061	0.391±0.009
DeST-OT	0.680±0.003	**0.329±0.017**	0.425±0.005	0.731±0.004	0.257±0.038	0.422±<0.001
DRIFT	**0.664±0.027**	0.361±0.036	0.380±0.003	**0.723±0.001**	**0.218±0.003**	0.377±0.002
	Linear			**Flow matching**		
Method	MMD	Limb Moran I	Cell JS	MMD	Limb Moran I	Cell JS
Ordinary OT	0.732±0.002	0.249±0.007	0.172±<0.001	0.748±0.006	0.179±0.010	0.172±<0.001
FGW OT	0.707±0.001	0.243±0.006	0.159±<0.001	0.737±0.002	0.213±0.131	0.160±0.001
PASTE	0.710±0.002	0.398±0.093	**0.142±0.002**	0.740±0.001	0.382±0.180	**0.141±0.002**
PASTE2	0.715±0.003	0.335±0.049	0.162±0.009	0.744±0.002	0.610±0.017	0.157±0.012
DeST-OT	0.711±0.002	**0.181±0.007**	0.168±<0.001	0.738±0.002	0.204±0.063	0.169±0.001
DRIFT	**0.702±0.003**	0.189±0.021	0.157±0.005	**0.727±0.011**	**0.165±0.004**	0.152±0.010
	Linear			**Flow matching**		
Method	MMD	Skin wound Moran I	Cell JS	MMD	Skin wound Moran I	Cell JS
Ordinary OT	**0.615±0.003**	0.390±0.004	0.086±0.009	0.632±0.004	0.496±0.107	**0.085±0.001**
FGW OT	0.626±<0.001	0.377±0.003	0.092±0.012	0.631±0.006	0.474±0.102	0.086±0.007
PASTE	0.668±0.008	**0.186±0.092**	**0.086±0.008**	0.653±0.010	0.355±0.158	0.086±0.008
PASTE2	0.670±0.005	0.223±0.089	0.127±0.005	0.674±0.007	0.436±0.075	0.127±0.006
DeST-OT	0.617±<0.001	0.371±0.006	0.089±0.009	0.628±0.004	0.450±0.135	0.091±0.008
DRIFT	0.621±<0.001	0.339±0.012	0.088±0.001	**0.605±0.003**	**0.170±0.006**	0.085±0.002

Table S5: **Alignment by stage pair.** Cells show mean and sample SD over ten runs. Coverage uses the denominators in Table S2.

Method	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	Top coverage	Mass coverage
Pancreas: 15 PCW → 18 PCW						
Ordinary OT	0.492±0.000	0.508±0.000	0.705±0.000	0.248±0.000	0.504±0.000	0.483±0.000
FGW OT	0.351±0.015	0.648±0.015	0.764±0.036	0.173±0.006	0.493±0.008	0.480±0.006
PASTE	0.173±0.002	0.827±0.003	0.823±0.001	0.079±0.002	0.458±0.002	0.459±0.003
PASTE2	0.186±0.020	0.814±0.020	0.814±0.019	0.049±0.007	0.262±0.016	0.516±0.031
DeST-OT	0.464±0.005	0.535±0.005	0.631±0.010	0.234±0.003	0.504±0.002	0.496±0.001
DRIFT	**0.732±0.021**	**0.267±0.021**	**0.499±0.024**	**0.402±0.013**	**0.549±0.012**	**0.528±0.004**
Pancreas: 18 PCW → 20 PCW						
Ordinary OT	0.585±0.000	0.390±0.000	0.409±0.000	0.292±0.000	0.499±0.000	0.486±0.000
FGW OT	0.513±0.006	0.461±0.007	0.550±<0.001	0.262±0.004	0.510±0.003	0.476±<0.001
PASTE	0.305±0.001	0.669±0.002	0.674±0.003	0.145±0.001	0.474±0.005	0.473±0.002
PASTE2	0.350±0.048	0.624±0.044	0.630±0.035	0.107±0.018	0.304±0.015	0.469±0.023
DeST-OT	0.709±0.003	0.276±0.004	0.285±0.010	0.367±0.005	0.517±0.006	0.507±0.003
DRIFT	**0.958±0.020**	**0.036±0.019**	**0.068±0.022**	**0.568±0.018**	**0.592±0.007**	**0.569±0.007**

Continued on next page

Table S5: Alignment by stage pair (continued).

Method	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	Top coverage	Mass coverage
Intestine: 12 PCW → 19 PCW						
Ordinary OT	0.581±0.000	0.393±0.000	0.422±0.000	0.424±0.000	0.730±0.000	0.736±0.000
FGW OT	0.591±0.008	0.371±0.007	0.480±<0.001	0.419±0.003	0.709±0.007	0.736±<0.001
PASTE	0.331±0.015	0.611±0.015	0.608±0.017	0.242±0.015	0.730±0.018	0.731±0.014
PASTE2	0.387±0.031	0.567±0.032	0.569±0.026	0.147±0.015	0.380±0.011	0.710±0.008
DeST-OT	0.645±0.002	0.327±0.002	**0.339±<0.001**	0.480±0.002	**0.744±0.001**	0.744±<0.001
DRIFT	**0.765±0.017**	**0.210±0.016**	0.408±0.004	**0.543±0.010**	0.710±0.010	**0.745±0.001**
Intestine: 19 PCW → Adult						
Ordinary OT	0.127±0.000	0.793±0.000	0.800±0.000	0.110±0.000	0.864±0.000	0.875±0.000
FGW OT	0.124±0.002	0.795±0.002	0.804±<0.001	0.107±0.002	0.862±0.001	0.875±<0.001
PASTE	0.109±0.007	0.811±0.006	0.808±0.007	0.095±0.006	0.870±0.010	0.871±0.009
PASTE2	0.100±0.005	0.807±0.006	**0.782±0.010**	0.054±0.005	0.548±0.057	0.833±0.022
DeST-OT	0.126±0.003	0.794±0.002	0.810±<0.001	0.110±0.002	**0.875±0.002**	**0.877±<0.001**
DRIFT	**0.380±0.038**	**0.559±0.036**	0.790±0.001	**0.322±0.031**	0.847±0.004	0.877±0.001
Limb: 5.6 PCW → 6.2 PCW						
Ordinary OT	0.992±0.000	0.008±0.000	0.008±<0.001	**0.091±0.000**	**0.092±0.000**	0.083±<0.001
FGW OT	1.000±0.000	0.000±0.000	0.011±<0.001	0.073±0.001	0.073±0.001	0.055±<0.001
PASTE	0.597±0.010	0.403±0.010	0.420±0.028	0.014±0.002	0.024±0.004	0.023±0.004
PASTE2	0.477±0.250	0.523±0.250	0.525±0.246	0.006±0.004	0.012±0.004	0.023±0.008
DeST-OT	0.971±0.003	0.029±0.003	0.027±<0.001	0.081±0.001	0.084±0.001	**0.092±<0.001**
DRIFT	1.000±0.000	0.000±0.000	**0.002±0.001**	0.091±0.007	0.091±0.007	0.082±0.012
Limb: 6.2 PCW → 7.0 PCW						
Ordinary OT	**0.991±0.000**	**0.009±0.000**	**0.011±<0.001**	0.083±0.000	0.084±0.000	0.078±<0.001
FGW OT	0.975±0.004	0.025±0.004	0.046±<0.001	0.086±0.001	0.088±0.001	0.070±<0.001
PASTE	0.534±0.027	0.466±0.027	0.463±0.017	0.015±<0.001	0.029±0.001	0.030±0.001
PASTE2	0.593±0.149	0.405±0.147	0.391±0.135	0.011±0.005	0.019±0.008	0.036±0.014
DeST-OT	0.966±0.004	0.034±0.004	0.025±0.001	0.079±0.001	0.081±0.001	0.081±<0.001
DRIFT	0.973±0.029	0.027±0.029	0.038±0.053	**0.092±0.009**	**0.095±0.012**	**0.084±0.006**
Limb: 7.0 PCW → 8.1 PCW						
Ordinary OT	0.648±0.000	0.330±0.000	0.587±<0.001	0.035±0.000	0.054±0.000	0.065±<0.001
FGW OT	0.554±0.019	0.417±0.017	0.771±<0.001	0.042±0.002	0.076±0.003	0.085±<0.001
PASTE	0.146±0.014	0.848±0.015	0.837±0.014	0.015±0.002	**0.100±0.003**	**0.104±0.003**
PASTE2	0.109±0.047	0.883±0.046	0.864±0.032	0.009±0.005	0.081±0.019	0.096±0.020
DeST-OT	0.579±0.018	0.401±0.020	**0.422±0.006**	0.029±0.002	0.050±0.002	0.044±<0.001
DRIFT	**0.691±0.149**	**0.249±0.141**	0.442±0.157	**0.044±0.008**	0.065±0.009	0.075±0.023

Continued on next page

Table S5: Alignment by stage pair (continued).

Method	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑	Top coverage	Mass coverage
Skin wound: D0 → D1						
Ordinary OT	0.510±0.000	0.489±0.000	0.549±0.000	0.207±0.000	0.405±0.000	0.425±0.000
FGW OT	0.519±0.004	0.481±0.003	0.562±<0.001	0.223±0.002	0.429±0.002	0.416±<0.001
PASTE	0.409±0.000	0.591±0.000	0.590±0.000	0.173±0.000	0.423±0.000	0.414±0.000
PASTE2	0.281±0.061	0.718±0.060	0.698±0.057	0.088±0.020	0.312±0.025	0.398±0.027
DeST-OT	0.572±0.011	0.425±0.012	0.452±0.004	0.236±0.004	0.412±0.003	0.414±0.004
DRIFT	**0.589±0.016**	**0.409±0.015**	**0.441±0.012**	**0.296±0.010**	**0.502±0.011**	**0.473±0.006**
Skin wound: D0 → D7						
Ordinary OT	0.668±0.000	0.332±0.000	0.379±0.000	0.196±0.000	0.293±0.000	0.246±0.000
FGW OT	0.614±0.004	0.380±0.005	0.480±0.001	0.168±0.003	0.273±0.005	0.242±<0.001
PASTE	0.111±0.000	0.889±0.000	0.904±0.000	0.022±0.000	0.195±0.000	0.197±0.000
PASTE2	0.328±0.172	0.671±0.170	0.682±0.173	0.061±0.040	0.161±0.052	0.209±0.042
DeST-OT	0.720±0.010	0.275±0.010	0.291±0.005	0.202±0.003	0.280±0.002	0.252±<0.001
DRIFT	**0.762±0.045**	**0.235±0.045**	**0.246±0.024**	**0.236±0.009**	**0.311±0.019**	**0.257±0.013**
Skin wound: D1 → D30						
Ordinary OT	**0.515±0.000**	**0.483±0.000**	0.624±0.000	0.151±0.000	0.294±0.000	0.328±0.000
FGW OT	0.438±0.005	0.560±0.005	0.695±<0.001	0.136±0.003	0.311±0.004	0.329±<0.001
PASTE	0.109±0.000	0.889±0.000	0.889±0.000	0.032±0.000	0.296±0.000	0.301±0.000
PASTE2	0.164±0.064	0.833±0.063	0.832±0.063	0.030±0.012	0.181±0.007	0.322±0.010
DeST-OT	0.493±0.013	0.506±0.012	**0.590±0.002**	0.145±0.006	0.294±0.004	0.325±0.005
DRIFT	0.413±0.025	0.586±0.025	0.615±0.014	**0.187±0.009**	**0.453±0.007**	**0.447±0.007**
Skin wound: D1 → D7						
Ordinary OT	0.734±0.000	0.262±0.000	0.300±0.000	0.193±0.000	0.263±0.000	0.232±0.000
FGW OT	0.688±0.014	0.306±0.014	0.343±0.001	0.165±0.001	0.239±0.004	0.220±<0.001
PASTE	0.652±0.000	0.347±0.000	0.352±0.000	0.147±0.000	0.226±0.000	0.223±0.000
PASTE2	0.499±0.105	0.498±0.102	0.496±0.096	0.052±0.007	0.107±0.013	0.191±0.020
DeST-OT	0.805±0.005	0.191±0.005	0.226±0.005	0.196±0.002	0.244±0.002	0.243±0.002
DRIFT	**0.821±0.055**	**0.178±0.054**	**0.185±0.044**	**0.263±0.014**	**0.321±0.005**	**0.277±0.003**
Skin wound: D7 → D30						
Ordinary OT	0.686±0.000	0.309±0.000	0.381±0.000	0.128±0.000	0.186±0.000	0.196±0.000
FGW OT	0.548±0.007	0.449±0.007	0.577±0.001	0.101±0.001	0.185±0.003	0.189±<0.001
PASTE	0.379±0.000	0.617±0.000	0.620±0.000	0.069±0.000	0.182±0.000	0.184±0.000
PASTE2	0.070±0.030	0.925±0.032	0.925±0.032	0.007±0.003	0.101±0.017	0.166±0.025
DeST-OT	0.674±0.008	0.320±0.010	0.375±0.002	0.144±0.004	0.214±0.004	0.217±0.003
DRIFT	**0.823±0.041**	**0.172±0.038**	**0.298±0.018**	**0.201±0.010**	**0.244±0.007**	**0.239±0.005**

F.3 COMPONENT ABLATIONS

We compare four variants with Full: w/o development removes the developmental cost; One-step uses direct transitions with a fallback to the Full prior; Dominant replaces each spot composition by a hard indicator for its dominant cell type; and Independent uses $\theta_s^\top C_{\mathrm{dev}} \theta_t$ without solving the inner composition OT problem. For One-step, we remove transient self loops, normalize the jump kernel from source to target by row, aggregate it by cell type, and fall back to the corresponding row of the full prior when no direct jump mass is available. All other weights, initialization, entropy regularization, endpoints, and solver budgets remain fixed.

Each variant is compared with Full at the same dataset and seed across ten paired seeds. Alignment scores first average stage pairs equally within each seed. Prediction scores first average targets within each seed, regime, and predictor. The three tables report the mean and sample SD of variant-minus-Full seed-level differences; they do not display absolute metric values. Bold and underlined entries rank the four displayed variants only. Full is the unshown zero reference.

Table S6: **Paired alignment ablation effects.** Each entry is the mean and sample SD across ten seed-level differences (variant − paired Full). Stage pairs are equally averaged within each seed. Positive differences favor the variant for LCR and LSC; negative differences favor it for MLDD and CWLD. Bold and underlined entries rank only the four displayed variants; Full is the unshown within-seed zero reference.

Pancreas Variant	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑
w/o development	-0.4146 ± 0.0048	$+0.4047 \pm 0.0047$	$+0.3689 \pm 0.0178$	-0.2696 ± 0.0007
One-step	-0.0125 ± 0.0000	$+0.0123 \pm 0.0000$	$+0.0021 \pm < 0.0001$	$\mathbf{-0.0051 \pm 0.0000}$
Dominant	$\mathbf{+0.0193 \pm 0.0000}$	$\mathbf{-0.0172 \pm 0.0000}$	$\mathbf{-0.0057 \pm < 0.0001}$	-0.0669 ± 0.0000
Independent	-0.0172 ± 0.0001	$+0.0141 \pm 0.0001$	$\underline{-0.0006 \pm < 0.0001}$	$\underline{-0.0301 \pm 0.0001}$
Intestine Variant	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑
w/o development	-0.2400 ± 0.0048	$+0.2205 \pm 0.0019$	$+0.0784 \pm 0.0059$	-0.1836 ± 0.0069
One-step	-0.0450 ± 0.0066	$+0.0463 \pm 0.0069$	$+0.0019 \pm 0.0044$	-0.0259 ± 0.0093
Dominant	$\mathbf{+0.0464 \pm 0.0048}$	$\mathbf{-0.0313 \pm 0.0019}$	$\mathbf{-0.0116 \pm 0.0059}$	$\mathbf{+0.0158 \pm 0.0069}$
Independent	-0.0962 ± 0.0048	$+0.0888 \pm 0.0019$	$+0.0233 \pm 0.0059$	-0.0693 ± 0.0069
Limb Variant	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑
w/o development	-0.0862 ± 0.0012	$+0.0902 \pm 0.0012$	$+0.1774 \pm 0.0004$	-0.0111 ± 0.0003
One-step	$\underline{-0.0240 \pm 0.0013}$	$\underline{+0.0228 \pm 0.0013}$	$\underline{+0.0773 \pm 0.0009}$	0.0000 ± 0.0003
Dominant	-0.0487 ± 0.0012	$+0.0384 \pm 0.0012$	$+0.0815 \pm 0.0004$	$\mathbf{+0.0065 \pm 0.0003}$
Independent	$\mathbf{+0.0014 \pm 0.0012}$	$\mathbf{-0.0039 \pm 0.0012}$	$\mathbf{+0.0251 \pm 0.0004}$	-0.0050 ± 0.0003
Skin wound Variant	LCR ↑	MLDD ↓	CWLD ↓	LSC ↑
w/o development	-0.1007 ± 0.0004	$+0.0995 \pm 0.0003$	$+0.1640 \pm 0.0004$	-0.0728 ± 0.0002
One-step	$\underline{-0.0026 \pm 0.0004}$	$\underline{+0.0027 \pm 0.0004}$	$\underline{+0.0208 \pm 0.0004}$	$\mathbf{-0.0019 \pm 0.0002}$
Dominant	-0.0382 ± 0.0004	$+0.0391 \pm 0.0003$	$+0.0791 \pm 0.0004$	-0.0591 ± 0.0002
Independent	$\mathbf{+0.0800 \pm 0.0016}$	$\mathbf{-0.0804 \pm 0.0015}$	$\mathbf{-0.0763 \pm 0.0006}$	$\underline{-0.0238 \pm 0.0006}$

Table S7: **Paired interpolation ablation effects.** Each entry is the mean and sample SD across ten seed-level differences (variant − paired Full). Targets are averaged within each seed and predictor before taking the difference. Lower is better for all three metrics, so negative differences favor the variant. Bold and underlined entries rank only the four displayed variants; Full is the unshown within-seed zero reference. Dataset names appear beneath each predictor heading.

Variant	MMD	Moran I	Cell JS
	Linear Pancreas		
w/o development	$+0.0073 \pm 0.0035$	$+0.0158 \pm 0.0071$	$0.0000 \pm < 0.0001$
One-step	$+0.0016 \pm 0.0039$	$+0.0076 \pm 0.0044$	$0.0000 \pm < 0.0001$
Dominant	$\mathbf{-0.0017 \pm 0.0034}$	$\mathbf{+0.0012 \pm 0.0032}$	$0.0000 \pm < 0.0001$
Independent	$\underline{+0.0011 \pm 0.0018}$	$\underline{+0.0047 \pm 0.0040}$	$0.0000 \pm < 0.0001$
	Linear Intestine		
Variant	MMD	Moran I	Cell JS
w/o development	$+0.0081 \pm 0.0033$	$+0.0241 \pm 0.0137$	$0.0000 \pm < 0.0001$
One-step	$\mathbf{-0.0004 \pm 0.0015}$	$\mathbf{-0.0046 \pm 0.0048}$	$0.0000 \pm < 0.0001$
Dominant	$\underline{+0.0014 \pm 0.0026}$	$\underline{-0.0044 \pm 0.0058}$	$0.0000 \pm < 0.0001$
Independent	$\underline{+0.0020 \pm 0.0029}$	$\underline{+0.0061 \pm 0.0110}$	$0.0000 \pm < 0.0001$
	Linear Limb		
Variant	MMD	Moran I	Cell JS
w/o development	$\underline{+0.0013 \pm 0.0033}$	$\underline{-0.0133 \pm 0.0169}$	$0.0000 \pm < 0.0001$
One-step	$\underline{+0.0021 \pm 0.0031}$	$\underline{+0.0081 \pm 0.0137}$	$0.0000 \pm < 0.0001$
Dominant	$\mathbf{+0.0008 \pm 0.0041}$	$\mathbf{-0.0143 \pm 0.0175}$	$0.0000 \pm < 0.0001$
Independent	$+0.0033 \pm 0.0014$	0.0000 ± 0.0060	$0.0000 \pm < 0.0001$
	Linear Skin wound		
Variant	MMD	Moran I	Cell JS
w/o development	$+0.0079 \pm 0.0022$	$+0.0614 \pm 0.0053$	$0.0000 \pm < 0.0001$
One-step	$\mathbf{-0.0005 \pm 0.0019}$	$\mathbf{+0.0037 \pm 0.0062}$	$0.0000 \pm < 0.0001$
Dominant	$\underline{+0.0026 \pm 0.0022}$	$\underline{+0.0121 \pm 0.0059}$	$0.0000 \pm < 0.0001$
Independent	$\underline{+0.0041 \pm 0.0016}$	$\underline{+0.0155 \pm 0.0073}$	$0.0000 \pm < 0.0001$
	Flow matching Pancreas		
Variant	MMD	Moran I	Cell JS
w/o development	$\underline{+0.0005 \pm 0.0030}$	$+0.0111 \pm 0.0029$	$0.0000 \pm < 0.0001$
One-step	$\underline{+0.0018 \pm 0.0014}$	$+0.0005 \pm 0.0022$	$0.0000 \pm < 0.0001$
Dominant	$\mathbf{-0.0015 \pm 0.0014}$	$\mathbf{-0.0015 \pm 0.0022}$	$0.0000 \pm < 0.0001$
Independent	$+0.0007 \pm 0.0019$	$\underline{+0.0002 \pm 0.0026}$	$0.0000 \pm < 0.0001$
	Flow matching Intestine		
Variant	MMD	Moran I	Cell JS
w/o development	$\mathbf{-0.0002 \pm 0.0042}$	$\mathbf{-0.0111 \pm 0.0076}$	$0.0000 \pm < 0.0001$
One-step	$+0.0005 \pm 0.0025$	-0.0006 ± 0.0102	$0.0000 \pm < 0.0001$
Dominant	$+0.0024 \pm 0.0028$	-0.0017 ± 0.0071	$0.0000 \pm < 0.0001$
Independent	$\underline{+0.0003 \pm 0.0043}$	$\underline{-0.0052 \pm 0.0126}$	$0.0000 \pm < 0.0001$
	Flow matching Limb		
Variant	MMD	Moran I	Cell JS
w/o development	$+0.0020 \pm 0.0020$	$\mathbf{-0.0092 \pm 0.0085}$	$0.0000 \pm < 0.0001$
One-step	$\underline{+0.0019 \pm 0.0017}$	$+0.0048 \pm 0.0086$	$0.0000 \pm < 0.0001$
Dominant	$\mathbf{+0.0005 \pm 0.0030}$	$\underline{-0.0085 \pm 0.0067}$	$0.0000 \pm < 0.0001$
Independent	$+0.0021 \pm 0.0023$	$\underline{-0.0001 \pm 0.0058}$	$0.0000 \pm < 0.0001$
	Flow matching Skin wound		
Variant	MMD	Moran I	Cell JS
w/o development	$\underline{-0.0023 \pm 0.0015}$	$\mathbf{-0.0030 \pm 0.0069}$	$0.0000 \pm < 0.0001$
One-step	$\underline{-0.0009 \pm 0.0029}$	$+0.0009 \pm 0.0052$	$0.0000 \pm < 0.0001$
Dominant	$\mathbf{-0.0034 \pm 0.0029}$	-0.0006 ± 0.0061	$0.0000 \pm < 0.0001$
Independent	-0.0017 ± 0.0028	$\underline{-0.0017 \pm 0.0054}$	$0.0000 \pm < 0.0001$

Table S8: **Paired extrapolation ablation effects.** Each entry is the mean and sample SD across ten seed-level differences (variant − paired Full). Targets are averaged within each seed and predictor before taking the difference. Lower is better for all three metrics, so negative differences favor the variant. Bold and underlined entries rank only the four displayed variants; Full is the unshown within-seed zero reference. Dataset names appear beneath each predictor heading.

Variant	MMD	Moran I	Cell JS
	Linear Pancreas		
w/o development	$+0.0032 \pm 0.0056$	$+0.0006 \pm 0.0099$	$+0.0026 \pm 0.0003$
One-step	-0.0001 ± 0.0017	-0.0011 ± 0.0033	$\mathbf{-0.0005 \pm < 0.0001}$
Dominant	$+0.0001 \pm 0.0020$	$\mathbf{-0.0016 \pm 0.0055}$	$+0.0003 \pm < 0.0001$
Independent	$\mathbf{-0.0022 \pm 0.0013}$	$+0.0006 \pm 0.0047$	$+0.0006 \pm < 0.0001$
	Linear Intestine		
w/o development	$+0.0094 \pm 0.0035$	-0.0613 ± 0.0499	-0.0086 ± 0.0011
One-step	$\mathbf{-0.0014 \pm 0.0025}$	-0.0111 ± 0.0333	-0.0025 ± 0.0008
Dominant	$+0.0044 \pm 0.0036$	-0.0647 ± 0.0497	$\mathbf{-0.0093 \pm 0.0011}$
Independent	$+0.0041 \pm 0.0024$	$\mathbf{-0.0650 \pm 0.0480}$	-0.0042 ± 0.0011
	Linear Limb		
w/o development	-0.0037 ± 0.0027	$+0.0017 \pm 0.0211$	$\mathbf{-0.0075 \pm 0.0007}$
One-step	$+0.0011 \pm 0.0030$	$+0.0132 \pm 0.0265$	$+0.0020 \pm 0.0007$
Dominant	$\mathbf{-0.0043 \pm 0.0023}$	$+0.0001 \pm 0.0213$	-0.0074 ± 0.0007
Independent	-0.0017 ± 0.0022	$\mathbf{-0.0073 \pm 0.0202}$	-0.0022 ± 0.0007
	Linear Skin wound		
w/o development	-0.0005 ± 0.0022	$+0.0281 \pm 0.0050$	$\mathbf{-0.0070 \pm < 0.0001}$
One-step	$\mathbf{-0.0007 \pm 0.0024}$	$\mathbf{+0.0060 \pm 0.0051}$	$+0.0012 \pm < 0.0001$
Dominant	-0.0005 ± 0.0017	$+0.0204 \pm 0.0070$	$-0.0062 \pm < 0.0001$
Independent	$+0.0002 \pm 0.0026$	$+0.0163 \pm 0.0052$	$-0.0013 \pm < 0.0001$
	Flow matching Pancreas		
w/o development	$+0.0013 \pm 0.0059$	$\mathbf{-0.0032 \pm 0.0046}$	$+0.0026 \pm 0.0003$
One-step	$\mathbf{-0.0011 \pm 0.0032}$	-0.0015 ± 0.0042	$\mathbf{-0.0005 \pm < 0.0001}$
Dominant	$+0.0013 \pm 0.0028$	$+0.0010 \pm 0.0052$	$+0.0003 \pm < 0.0001$
Independent	$+0.0016 \pm 0.0041$	-0.0016 ± 0.0034	$+0.0006 \pm < 0.0001$
	Flow matching Intestine		
w/o development	$+0.0049 \pm 0.0033$	$+0.0151 \pm 0.0230$	-0.0086 ± 0.0011
One-step	$\mathbf{-0.0006 \pm 0.0027}$	$\mathbf{+0.0039 \pm 0.0215}$	-0.0025 ± 0.0008
Dominant	$+0.0025 \pm 0.0032$	$+0.0281 \pm 0.0294$	$\mathbf{-0.0093 \pm 0.0011}$
Independent	$+0.0026 \pm 0.0029$	$+0.0278 \pm 0.0210$	-0.0042 ± 0.0011
	Flow matching Limb		
w/o development	$+0.0022 \pm 0.0038$	$\mathbf{-0.0058 \pm 0.0066}$	$\mathbf{-0.0075 \pm 0.0007}$
One-step	-0.0007 ± 0.0045	$+0.0014 \pm 0.0089$	$+0.0020 \pm 0.0007$
Dominant	$\mathbf{-0.0010 \pm 0.0049}$	-0.0010 ± 0.0076	-0.0074 ± 0.0007
Independent	$+0.0006 \pm 0.0047$	$+0.0023 \pm 0.0095$	-0.0022 ± 0.0007
	Flow matching Skin wound		
w/o development	$+0.0057 \pm 0.0048$	$+0.0029 \pm 0.0110$	$\mathbf{-0.0070 \pm < 0.0001}$
One-step	$\mathbf{-0.0024 \pm 0.0048}$	-0.0056 ± 0.0076	$+0.0012 \pm < 0.0001$
Dominant	-0.0006 ± 0.0050	$\mathbf{-0.0072 \pm 0.0109}$	$-0.0062 \pm < 0.0001$
Independent	-0.0005 ± 0.0056	$+0.0001 \pm 0.0097$	$-0.0013 \pm < 0.0001$

F.4 Runtime, Memory, and Scaling

We profile all six methods with spatial caps of 250, 500, 1,000, and the maximum available value for each stage (2,097 for pancreas, 2,316 for intestine, 2,616 for limb, and 2,019 for skin wound), skipping and deduplicating unsupported levels. The corresponding totals at maximum scale are 5,054, 4,635, 6,272, and 6,123 spots because stage sizes differ. DRIFT is additionally profiled

with 25%, 50%, and 100% of each single cell reference while spatial size remains maximal. Every configuration is repeated five times.

All timing jobs run sequentially in a dedicated window after the quality queue has completed. Each process is pinned to at most eight CPU threads and one NVIDIA RTX PRO 6000 Blackwell Server Edition GPU. "Cold" model time is the sum of data reading, model preprocessing, prior construction, construction of composition costs, coupling, flow training and sampling, and sampling of coupling pairs. It starts from standardized spatial inputs and saved single cell PCA and velocity summaries; raw sequencing processing and deconvolution are upstream and excluded. The time "with saved prior" subtracts prior construction from the same run. The "OT solve" is the recorded internal OT solve for DRIFT and the complete coupling stage for the baselines. GPU operations are synchronized at phase boundaries. Evaluation, plotting, and result writing are excluded from these timing totals, but the directly sampled memory peaks cover the complete process, including evaluation and writing. Consequently, stage times can be added, whereas memory peaks from different stages must not be added.

Across the four maximum-scale benchmark configurations, DRIFT requires 48.86–216.44 seconds of median cold model time, with median process-tree peak RSS of 2.81–3.75 GiB and median GPU-memory peaks below 0.79 GiB. Reusing the skin-wound prior reduces median runtime from 216.44 to 166.33 seconds.

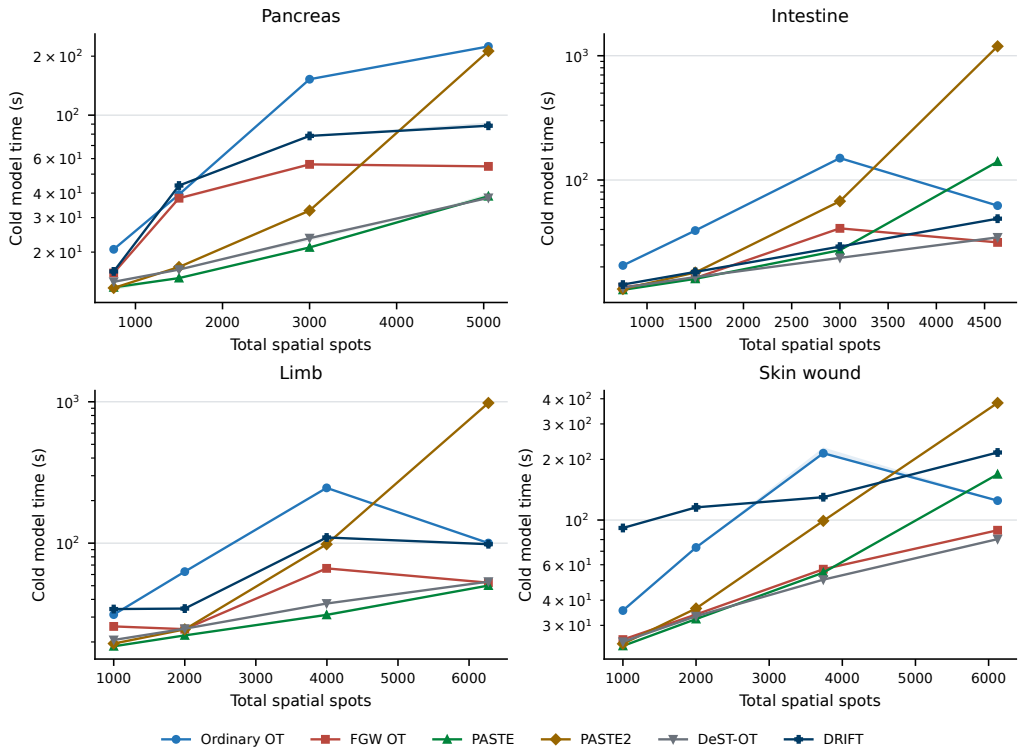


Figure S4: **Spatial size scaling.** Cold model time as the total number of retained spatial spots increases. Lines show medians and bands show interquartile ranges over five isolated runs. The time axis is logarithmic. Each nominal size is a cap for every stage, so the horizontal coordinate reports the actual total across stages.

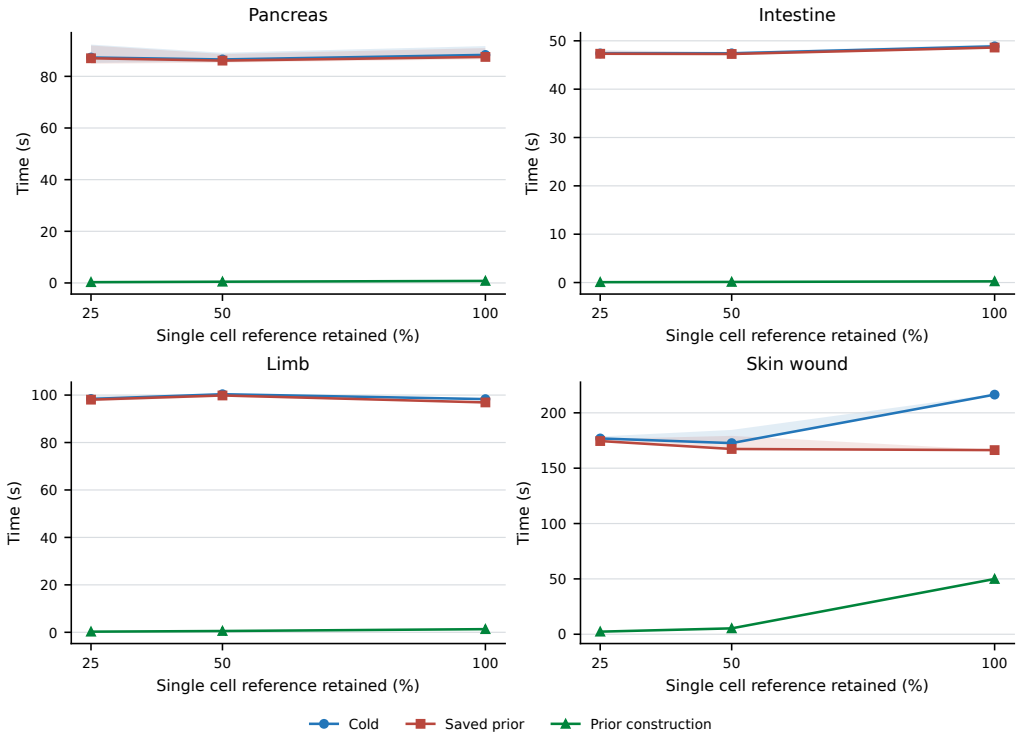


Figure S5: **Single cell reference scaling for DRIFT.** Cold time, time with a saved prior, and prior construction time at maximum spatial size as the retained stratified fraction of each single cell reference varies. Lines show medians and bands show interquartile ranges over five isolated runs.

Table S9: **Maximum-scale model time.** Median seconds $[Q_1, Q_3]$ over five isolated runs.

Dataset	Method	Cold	Saved prior	OT solve
Pancreas	Ordinary OT	224.24 [223.59, 224.79]	224.24 [223.59, 224.79]	204.81 [203.66, 205.83]
	FGW OT	54.77 [54.47, 55.30]	54.77 [54.47, 55.30]	35.70 [35.45, 35.75]
	PASTE	38.56 [38.52, 39.81]	38.56 [38.52, 39.81]	20.10 [20.00, 20.44]
	PASTE2	212.36 [211.84, 214.25]	212.36 [211.84, 214.25]	194.37 [192.72, 195.18]
	DeST-OT	**37.82 [37.58, 38.15]**	**37.82 [37.58, 38.15]**	**18.49 [18.48, 18.51]**
	DRIFT	88.30 [87.49, 91.73]	87.52 [86.61, 90.94]	45.72 [44.43, 47.39]
Intestine	Ordinary OT	62.23 [62.08, 63.36]	62.23 [62.08, 63.36]	43.34 [43.27, 43.45]
	FGW OT	**31.49 [31.06, 31.54]**	**31.49 [31.06, 31.54]**	**12.03 [11.59, 12.11]**
	PASTE	140.59 [140.57, 140.64]	140.59 [140.57, 140.64]	121.42 [121.39, 121.69]
	PASTE2	1191.11 [1189.48, 1192.59]	1191.11 [1189.48, 1192.59]	1172.12 [1169.62, 1173.25]
	DeST-OT	34.56 [34.50, 35.27]	34.56 [34.50, 35.27]	15.81 [15.77, 15.83]
	DRIFT	48.86 [48.66, 49.23]	48.61 [48.42, 48.99]	13.64 [13.59, 14.11]
Limb	Ordinary OT	100.15 [99.77, 100.31]	100.15 [99.77, 100.31]	70.86 [70.00, 70.98]
	FGW OT	52.52 [52.11, 52.59]	52.52 [52.11, 52.59]	22.85 [22.66, 22.88]
	PASTE	**50.12 [49.82, 50.18]**	**50.12 [49.82, 50.18]**	**20.77 [20.75, 20.91]**
	PASTE2	981.23 [981.18, 983.82]	981.23 [981.18, 983.82]	951.90 [951.85, 954.11]
	DeST-OT	53.41 [53.14, 53.78]	53.41 [53.14, 53.78]	23.80 [23.71, 23.85]
	DRIFT	98.31 [98.16, 98.41]	96.90 [96.85, 97.10]	25.70 [25.64, 26.10]
Skin wound	Ordinary OT	124.86 [123.78, 125.54]	124.86 [123.78, 125.54]	67.93 [67.69, 68.29]
	FGW OT	88.98 [88.91, 89.70]	88.98 [88.91, 89.70]	31.56 [31.52, 31.71]
	PASTE	168.71 [167.47, 169.39]	168.71 [167.47, 169.39]	111.16 [110.95, 111.65]
	PASTE2	381.45 [379.60, 382.00]	381.45 [379.60, 382.00]	324.23 [322.75, 324.94]
	DeST-OT	**80.40 [79.77, 81.34]**	**80.40 [79.77, 81.34]**	**22.79 [22.77, 22.96]**
	DRIFT	216.44 [216.25, 216.98]	166.33 [165.99, 166.50]	53.60 [52.62, 54.53]

Table S10: **Maximum-scale phase times.** Median seconds over five isolated runs; phases correspond to Table S9.

Dataset	Method	Read	Prep.	Prior	Comp.	Coupling	Flow	Sample
Pancreas	Ordinary OT	1.02	15.97	0.00	0.00	204.81	2.84	0.04
	FGW OT	**0.99**	15.46	**0.00**	0.00	35.70	2.83	**0.03**
	PASTE	1.00	**14.99**	0.00	0.00	20.10	**2.56**	0.03
	PASTE2	1.00	15.47	0.00	0.00	194.37	2.57	0.04
	DeST-OT	1.01	15.75	0.00	0.00	**18.49**	2.59	0.03
	DRIFT	1.00	15.75	0.79	20.60	45.72	2.80	0.03
Intestine	Ordinary OT	0.98	15.33	0.00	0.00	43.34	2.81	0.03
	FGW OT	0.99	15.54	0.00	0.00	**12.03**	2.82	0.03
	PASTE	0.98	15.75	0.00	0.00	121.42	**2.54**	0.03
	PASTE2	0.98	16.02	**0.00**	0.00	1172.12	2.61	0.03
	DeST-OT	**0.98**	**15.26**	0.00	0.00	15.81	2.55	**0.02**
	DRIFT	0.98	15.68	0.25	15.30	13.64	2.82	0.02
Limb	Ordinary OT	1.19	**24.18**	0.00	0.00	70.86	3.89	0.07
	FGW OT	**1.17**	24.59	0.00	0.00	22.85	3.82	0.07
	PASTE	1.18	24.50	0.00	0.00	**20.77**	3.54	0.07
	PASTE2	1.18	24.60	0.00	0.00	951.90	**3.54**	**0.06**
	DeST-OT	1.18	24.62	**0.00**	0.00	23.80	3.57	0.07
	DRIFT	1.18	24.66	1.33	41.41	25.70	3.83	0.07
Skin wound	Ordinary OT	0.93	52.65	0.00	0.00	67.93	3.61	0.08
	FGW OT	0.93	52.83	0.00	0.00	31.56	3.60	0.08
	PASTE	**0.92**	52.50	**0.00**	0.00	111.16	3.62	0.08
	PASTE2	0.93	**52.28**	0.00	0.00	324.23	**3.55**	**0.07**
	DeST-OT	0.93	53.00	0.00	0.00	**22.79**	3.59	0.08
	DRIFT	0.93	53.00	49.92	54.15	53.60	3.83	0.09

Table S11: **Memory at maximum scale.** Values are medians over five runs. RSS is the peak resident set size over the process tree; NVML is the peak memory of the study process tree on its assigned GPU. PyTorch columns are allocator peaks. These peaks cover the full run, including evaluation and result writing.

Dataset	Method	RSS (GiB)	NVML (GiB)	Alloc. (GiB)	Reserved (GiB)
Pancreas	Ordinary OT	2.80	0.76	0.08	0.09
	FGW OT	**2.78**	0.76	0.08	0.09
	PASTE	3.23	7.04	4.40	6.36
	PASTE2	3.14	7.04	4.40	6.36
	DeST-OT	3.19	1.36	0.42	0.68
	DRIFT	2.81	0.76	0.08	0.09
Intestine	Ordinary OT	3.22	0.78	0.09	0.11
	FGW OT	3.23	0.78	0.09	0.11
	PASTE	3.46	5.94	3.95	5.27
	PASTE2	3.56	5.95	3.95	5.27
	DeST-OT	3.65	1.29	0.42	0.62
	DRIFT	**3.19**	0.78	0.09	0.11
Limb	Ordinary OT	3.66	0.78	0.09	0.11
	FGW OT	3.66	0.78	0.09	0.11
	PASTE	3.88	6.41	4.25	5.74
	PASTE2	3.98	6.41	4.25	5.74
	DeST-OT	4.08	1.58	0.52	0.90
	DRIFT	**3.62**	0.78	0.09	0.11
Skin wound	Ordinary OT	3.74	0.76	0.08	0.09
	FGW OT	**3.71**	0.76	0.08	0.09
	PASTE	3.95	8.19	4.69	7.51
	PASTE2	4.06	8.19	4.69	7.51
	DeST-OT	4.18	1.65	0.47	0.97
	DRIFT	3.75	0.76	0.08	0.09

Table S12: **DRIFT single cell reference scaling at maximum spatial size.** Cell counts are summed over rebuilt adjacent priors; timings are medians with $[Q_1, Q_3]$ over five runs.

Dataset	Fraction	Source cells	Target cells	Cold (s)	Prior (s)
Pancreas	0.25	995	1,233	87.29 [85.18, 92.31]	0.30 [0.30, 0.31]
	0.5	1,988	2,286	86.54 [85.78, 89.14]	0.48 [0.47, 0.49]
	1	3,974	3,974	88.30 [87.49, 91.73]	0.79 [0.79, 0.80]
Intestine	0.25	1,326	921	47.40 [47.38, 48.05]	0.08 [0.08, 0.08]
	0.5	2,652	1,746	47.41 [47.19, 47.73]	0.13 [0.13, 0.14]
	1	5,301	3,243	48.86 [48.66, 49.23]	0.25 [0.24, 0.25]
Limb	0.25	5,125	5,985	98.38 [97.61, 99.61]	0.27 [0.27, 0.31]
	0.5	10,250	11,445	100.37 [99.81, 100.44]	0.54 [0.53, 0.54]
	1	20,500	20,500	98.31 [98.16, 98.41]	1.33 [1.32, 1.33]
Skin wound	0.25	10,934	15,026	176.86 [176.47, 178.65]	2.35 [2.33, 2.35]
	0.5	21,864	27,898	172.66 [172.50, 184.58]	5.32 [5.31, 5.36]
	1	43,727	47,233	216.44 [216.25, 216.98]	49.92 [49.59, 50.46]