stVCR: Spatiotemporal dynamics of single cells

Qiangwei Peng[1], Peijie Zhou[2*], Tiejun Li[1,2*]

[1]LMAM and School of Mathematical Sciences, Peking University.
[2]Center for Machine Learning Research, Peking University.

*Corresponding author(s). E-mail(s): pjzhou@pku.edu.cn; tieli@pku.edu.cn;

Abstract

Time-series spatial transcriptome data with single-cell resolution provide an opportunity to study cell differentiation, proliferation and migration in physical space over time. Due to the destructive nature of sequencing, reconstruction of spatiotemporal dynamics from data remains challenging. Especially, the inference of migration in physical space remains a difficult task, as samples obtained at different temporal snapshots might not be in the same coordinate system due to the difference of biological replicates. Here we developed stVCR, a generative deep learning model, which integrates the dynamical optimal transport (OT) with the unbalanced setting, the density matching invariant to rigid body transformations as well as priors to model known biology and preserve spatial structure. stVCR achieves the end-to-end simultaneous reconstruction of continuous cell differentiation, proliferation, physical space migration, and spatial coordinates alignment from spatial transcriptome snapshots. In addition, stVCR allows the interpretable study of complex interactions between cell phenotype transition, spatial migration and proliferation. Through benchmarking on both simulation data and real datasets, we validated the effectiveness and robustness of stVCR and demonstrated its advantages over static OT or linear interpolation methods. We applied stVCR to dissect spatiotemporal dynamics underlying axolotl brain regeneration and 3D Drosophila embryo development.

Introduction

The development of a fertilized egg into a complete embryo is a highly complex and important process in biology [1–4]. This process involves intricate interactions between the dynamic regulation of gene expression, cell differentiation, cell division, apoptosis, as well as cell migration within physical space [5, 6].

The advent of spatial transcriptome (ST) technology has allowed obtaining both gene expression data and spatial coordinates [7–11]. As technology advances, spatial resolution has reached the single-cell or even subcellular level, exemplified by methods such as Stereo-seq [10] and 10x Visium HD [11]. However, due to the destructive nature of sequencing, ST data can only provide snapshots rather than a continuous trajectory. If ST sequencing technology is likened to an ultra-wide-angle camera, it can take pictures of living organisms but lacks video recording capability. Especially, when sequencing at multiple time points during embryonic development, the resulting time-series ST data often come from different biological replicates, therefore yielding multiple unpaired snapshots.

Recovering cells' dynamic trajectories from single-cell sequencing data or ST data is a challenging task. RNA velocity [12] utilizes unspliced/spliced RNA to infer the developmental direction of each cell. This inspired a series of subsequent works using unspliced/spliced RNA to more accurately infer RNA velocity [13–18]. These methods suffer from scale invariance due to the lack of temporal information [19]. Metabolic labeling scRNA-seq introduces temporal information into the data by distinguishing new/old RNA [20–27]. Dynamo [28] designed parameter inference methods for metabolic labeling scRNA-seq data based on steady-state assumptions and deterministic models, and Storm [29] extended it to be independent of steady-state assumptions and stochastic models. Time-series scRNA-seq data introduces temporal information into the data in another way.

Waddington-OT [30] pioneered the use of optimal transport(OT) for modeling time-series scRNA-seq data, finding the optimal mapping in cells at two adjacent time points. However, it approximates cell proliferation by using growth hallmark genes, which largely depends on the choice of database.

Targeted for time-series scRNA data, TrajectoryNet [31] combines dynamical OT [32] and normalizing flows to infer continuous trajectories of cell and takes cell proliferation into account through a separated static OT. MIOFlow [33] follows the geometry by operating in the latent space of an geodesic autoencoder. TIGON [34] uses dynamical unbalanced OT [35] to reconstructs dynamic trajectories and population growth simultaneously. Since the usual scRNA-seq data do not include the spatial coordinates of the cells, these methods have limitations to directly model cell migration in physical space. Some work considers stochastic cellular dynamics, such as FBSDE [36] and PI-SDE [37]. Additionally, DeepRUOT extends even further to the stochastic unbalanced case [38].

The availability of time-series ST data has made it possible to study how cells migrate in physical space. PASTE [39] uses fused Gromov-Wasserstein(GW) OT [40] to align 2D adjacent tissue ST slices to reconstruct the 3D structure of the tissue. Moscot [41] uses a similar fused GW-OT to find the optimal mapping of cells between slices at two adjacent time points, incorporating penalties for unbalanced and entropy regularization and employing a low-rank OT [42] to accommodate larger data sizes. Spateo [43] aligns the spatial coordinates of two adjacent time points by optimal mapping to obtain cell migration velocity and then learns a vector field of continuous spatial coordinates. However, it does not fully address the interplay between gene expression and spatial location, and processes such as cell division and apoptosis. DeST-OT [44] considers how to model cell proliferation in the static OT setting, in particular ST data. TopoVelo [45] uses spatial coordinates to model cellular neighborhoods when inferring usual RNA velocity based on unspliced/spliced RNA, and designs post-processing steps to infer cell migration velocity. STT [46] characterizes multistability in space by integrating unspliced/spliced RNA and ST through a multiscale dynamical model.

Reconstructing dynamical trajectories of cell differentiation, proliferation, and migration in physical space simultaneously for time-series ST data is a challenging task. Especially for quantifying the migration in physical space, improper treatment might introduce pseudo movements of cells as the cell coordinates obtained at different temporal snapshots are not in same coordinate system. Analogous to recovering a video from multiple photos, we aim to reconstruct the entire cellular developmental dynamics from multiple unpaired ST snapshots, thus obtaining a continuous spatiotemporal developmental trajectory. To achieve this goal, we developed an algorithm called spatio-temporal Video Cassette Recorder (stVCR), which is a dynamical optimal transport algorithm for resolving the issue of alignment of ST section data and unbalanced populations at different snapshots, and incorporation of biological structure priors in an integrative manner. As the result, stVCR reconstructs the spatiotemporal dynamical process for the considered system from multiple ST snapshots. Furthermore, stVCR also reveals the complex regulatory mechanisms behind the overall cellular dynamics, including how gene expression and spatial location affect each other, and how they affect cell proliferation.

Results

Overview of stVCR

In stVCR, we adopt the dynamical OT formulation as a framework, yet with special treatments for different types of data (Fig. 1A). Specifically, for gene expression counts, we use the Wasserstein OT to model the temporal coupling of distributions (Fig. 1A Left and Methods). For spatial coordinates of cells, since rotations and translations may be involved to prevent a direct comparison of cell coordinates at different instants, we use the rigid-body transformation invariant OT to make the spatial alignment in time (Fig. 1A Middle and Methods). For the number of cells, due to the cell division and apoptosis, we use the unbalanced OT to model the unbalanced populations (Fig. 1A Right and Methods). Additionally, stVCR optionally takes known cell type transition prior (Fig. 1B Left) as well as the spatial structure preserving prior for specific cell types (or organs) (Fig. 1B) to produce more biologically meaningful results (Methods). We unified all three necessary modules and two optional modules into the form of dynamical OT, allowing us to study how a population of cells changes in gene expression, how they migrate in physical space, and how they divide and apoptose over time (Figure 1A,B,C and Methods). We take the spatial coordinates of cells at the first time point $t = t_0$ as the reference coordinates system, and the state of considered cell group at time t is described by a time-dependent distribution $\rho_t(\boldsymbol{x}, \boldsymbol{q})$, where ρ_t depends on the spatial coordinates $\boldsymbol{x} \in \mathbb{R}^{d_s}$ ($d_s = 2$ or 3) of the cells in the reference coordinates system and the gene expression variable $\boldsymbol{q} \in \mathbb{R}^{d_g}$ after dimensionality reduction. Generally, ρ_t is not a probability

distribution though we take normalization at $t = t_0$. stVCR finds the optimal rigid rotation matrix R_k and translation vector $\boldsymbol{r}_k$ for the coordinate system at time point $t = t_k$ except t_0, and uses the transport-with-growth PDE

$$\partial_t \rho_t + \nabla \cdot ((\boldsymbol{v}_t, \boldsymbol{p}_t)\rho_t) = g_t \rho_t \tag{1}$$

to interpolate the empirical density $\hat{\rho}^{(k)}$ and cell number n_k of the snapshot spatial transcriptomic data at $t = t_k$ after rigid body transformation (Fig. 1C), where $\boldsymbol{v}_t$, $\boldsymbol{p}_t$, and g_t are parameterized to be learned by neural networks. In physical meaning, $\boldsymbol{v}_t = \mathrm{d}\boldsymbol{x}/\mathrm{d}t$ describes the migration velocity of cells in physical space, $\boldsymbol{p}_t = \mathrm{d}\boldsymbol{q}/\mathrm{d}t$ describes the RNA velocity of cells in (reduced) gene expression space, and $g_t \in \mathbb{R}$ describes cell proliferation (Fig. 1C, D).

stVCR parameterizes the 2D rotation matrix by using rotation angle (or Euler angle for 3D case), and simultaneously finds the optimal rigid body transformations and parameterized dynamics by minimizing a total loss composed of dynamics loss, matching loss and optional spatial structure preserving loss

$$\mathcal{L} = \mathcal{L}_{\mathrm{Dyn}} + \lambda_{\mathrm{Mch}}\mathcal{L}_{\mathrm{Mch}} + \lambda_{\mathrm{SSP}}\mathcal{L}_{\mathrm{SSP}}^{(\mathrm{opt})}. \tag{2}$$

The dynamics loss $\mathcal{L}_{\mathrm{Dyn}}$ further contains three parts

$$\mathcal{L}_{\mathrm{Dyn}} = \mathcal{L}_{\mathrm{Spa}} + \alpha_{\mathrm{Exp}}\mathcal{L}_{\mathrm{Exp}} + \alpha_{\mathrm{Gro}}\mathcal{L}_{\mathrm{Gro}}, \tag{3}$$

promoting the least consumption of kinetic energy of spatial migration and gene expression change, and growth energy, respectively. The matching loss $\mathcal{L}_{\mathrm{Mch}}$ promotes the cell dynamics to match the aligned ST data as well as possible at different time points, and the *optional* spatial structure preserving loss $\mathcal{L}_{\mathrm{SSP}}^{(\mathrm{opt})}$ promotes a stable spatial structure for the the user-specified organ or cell type by encouraging adjacent cells to have similar spatial velocities, thereby preventing arbitrary deformations. (Fig. 1D and Methods). The training process involves OT optimization and integrating ODEs represented by neural networks, which we solve using the POT [47] and torchdiffeq packages [48], respectively.

Once we have completed the entire training process to obtain the optimal rigid-body transformation and parameterized dynamics, we can first apply the optimal rigid-body transformation to align the spatial coordinates of cells at different time points to the reference coordinate system, and then perform a series of downstream analyses (Fig. 1E and Methods): (1) Interpolation and prediction. We evolve forward or backward from the nearest observations to the interested time point (between observations or in the future) based on learned cellular dynamics (Fig. 1E Top left and Methods). (2) Gene-space interactions. We study cell-specific gene-gene, gene-space and space-space interactions by calculating the Jacobian matrices of learned spatial migration dynamics and gene expression dynamics and further calculating the directional derivatives along the cell migration direction of interest (Fig. 1E Bottom left and Methods). (3) Proliferation driver genes and migration direction. We study the effects of genes and migration on growth by calculating the gradient of the learned growth dynamics and further calculating the similar directional derivatives (Fig. 1E Top right and Methods). (4) Lineage inference/generation. For originally annotated data, we can infer temporal developmental lineages by learning a time-dependent classifier to annotate unobserved cells generated by interpolation or prediction (Fig. 1E Bottom right and Methods).

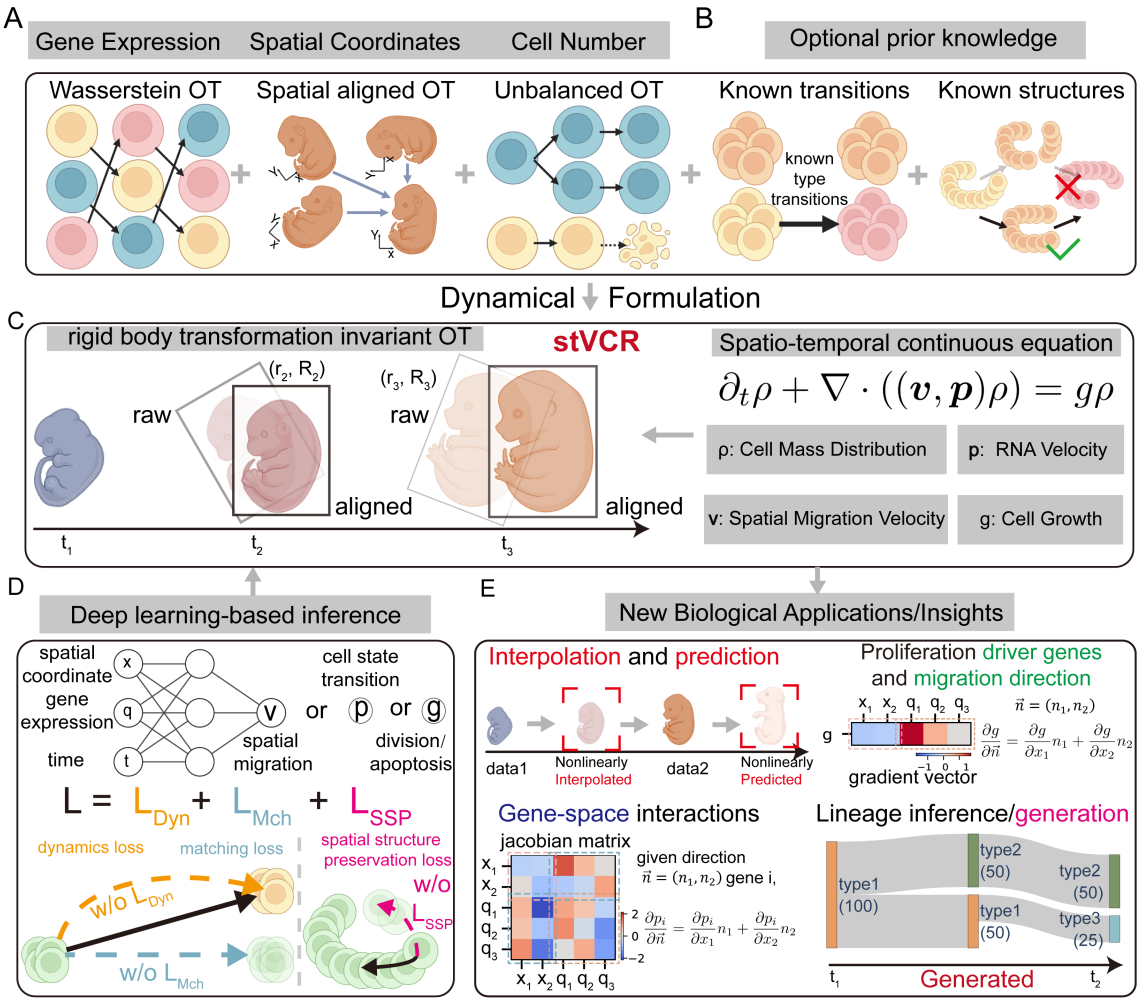

Fig. 1: **Overview of stVCR. A.** stVCR adopts dynamical OT framework yet with special treatments for different types of data in the spatial transcriptome. Specifically we use Wasserstein OT for gene expression data (Left), rigid body transformation invariant OT for spatial coordinates (Middle), and unbalanced OT for cell number change due to cell division and apoptosis (Right). **B.** stVCR can optionally model prior knowledge, including biological priors for known type transitions (Left) and spatial structure preserving priors (Right). **C.** stVCR unifies the three necessary modules and two optional modules into a dynamical OT. The input spatial transcriptome snapshots are described as distributions $\rho^{(k)}$, and the permissible rotations and translations are characterized by $(R_k, \boldsymbol{r}_k)$ at $t = t_k$. The modeling density ρ_t is governed by a partial differential equation involving spatial velocity $\boldsymbol{v}$, RNA velocity $\boldsymbol{p}$, and growth rate g. **D.** stVCR solves the problem in **B** based on deep learning. $\boldsymbol{v}_t(\boldsymbol{x}, \boldsymbol{q})$, $\boldsymbol{p}_t(\boldsymbol{x}, \boldsymbol{q})$ and $g_t(\boldsymbol{x}, \boldsymbol{q})$ are modeled by three neural networks. The loss function includes three parts: dynamics loss, matching loss and spatial structure preserving loss. **E.** stVCR can perform downstream analyses, including interpolation and prediction (Top left), studying cell-specific gene-gene, gene-space, space-space interactions (Bottom left), exploring the cell-specific effects of gene expression and spatial variations of growth rates (Top right) and inferring temporal cell-type developmental lineages (Bottom right).

Benchmark on the simulated time-series ST data for accuracy, scalability and robustness

To demonstrate the necessity of aligning the spatial coordinates of different temporal snapshots into the same coordinate system, and benchmark the ability of the stVCR to recover spatiotemporal dynamics and reveal key regulatory mechanisms, we generated the simulated dataset of gene circuits and two spatial dimensions (Fig. 2A, B and Methods). The three genes are named *Red*, *Green* and *Blue* genes. There are regulatory relationships between different genes and different spatial coordinates, in addition, gene expression and cell migration also affect cell proliferation (Fig. 2A). *Red* and *Green* genes form a toggle switch circuit and they have opposite effects on growth. In addition, the difference in spatial location makes them unequal in status (Fig. 2A, B and Methods).

In the simulated dataset, there are three groups of cells at the initial moment (Fig. 2 B). The first group highly expressed gene *Blue* and was in steady state (Fig. 2 B). The second and third groups had similar low gene expression at the initial moment, but their different spatial locations determined their different fates of transition and growth (Fig. 2B). Without spatial information, it is not possible to distinguish between the second and third group at the initial moment, which would lead to erroneous trajectories (Fig. 2B Left and Fig. S4 last row). If cell proliferation is ignored, it will lead to incorrect cell trajectories of the small number group to the large number group (Fig. 2B Right and Fig. S5 last row). The input data totaled 6 time points, and we rotated the spatial coordinates by different angles to simulate the possible rotation of tissues by spatial transcriptome sequencing (Fig. 2C and Fig. S1A).

To illustrate the ability of stVCR to align the spatial coordinates of different temporal snapshots and reconstruct the entire spatiotemporal dynamics, we took the data from the first time point and evolved them according to the learned dynamics, demonstrating consistency with real dynamics (Supplementary Video 1; Fig. 2D and Fig. S1B). Specifically, the first group of cells remained virtually unchanged. The second group of cells gradually overexpressed the *Red* gene, moved outwards in the horizontal direction, and continuously proliferated. The third group of cells gradually overexpressed *Green* gene and continued apoptosis. In addition, we observed that in the aligned space by stVCR, cells only moved horizontally and did not rotate, indicating that we found the optimal rigid body transformation to align the data at different time points while finding the optimal dynamics of cell evolution (Supplementary Video 1, Fig. 2D and Fig. S1B). In addition, stVCR interpolated the unobserved intermediate moments $t = 0.25$, 0.75, 1.25, 1.75 and 2.25 and predicted the future moments $t = 2.75$ based on the learned dynamics, and the results are close to ground truth (Fig. 2E,F and Fig. S1C).

To investigate the ability of stVCR to restore the effects of gene interactions, we compared the partial derivatives of *Green* gene velocity with respect to *Red* gene expression with ground truth, and visualized them in spatial coordinates (Fig. 2G Left). Qualitatively, they were consistent, and *Green* gene inhibited *Red* gene expression mainly in the second and third group of cells.

Next, to investigate the ability of stVCR to restore the cell migration effects on gene expression, we calculated the directional derivative of *Red* gene expression for the given direction $\boldsymbol{n} = (1, 0)$ (i.e., cells moving horizontally to the right) for both learned and true dynamics (Fig. 2G Right). Cells at the right end of the second group moving to the right will promote gene *Red* expression, and cells at the left end moving to the right will inhibit gene *Red* expression, which overall suggests that moving horizontally outward in the second group of cells will promote *Red* expression.

To evaluate the spatial variability of cell proliferation and the effect of cell migration on growth, we compared the true and learned cell proliferation rates (Fig. 2H Left). The results show that the first group of cells has a growth rate close to 0, the second group has a large positive growth rate, and the third group has a large negative growth rate. Additionally we calculated the directional derivative of the growth rate g with respect to a given direction $\boldsymbol{n} = (1, 0)$, similarly showing that cells moving outward in the horizontal direction will promote cell proliferation (Fig. 2H Right).

Finally, we checked the scalability and the robustness with respect to important hyperparameters of the stVCR. We first performed a scalability analysis, which shows that stVCR is scalable for dataset size, model size, and number of observation times when the proper sample batch size is chosen (Fig. S2). Next, we tested the robustness of stVCR with respect to the important hyperparameters λ_{Mch} (Fig. S3 and Supplementary Video 2), κ_{Exp} (Fig. S4 and Supplementary Video 3), and α_{Gro} (Fig. S5 and Supplementary Video 4), where λ_{Mch} measures the importance of the matching loss, κ_{Exp} weighs the importance of gene expression and spatial coordinates in the matching term and α_{Gro} measures the flexibility of cell proliferation. The results show that stVCR is robust over a wide range of these parameters.

In summary, our benchmark tests on this simulated data show that (1) it is necessary to align the spatial coordinates of different time snapshots to the same coordinate system; (2) stVCR simultaneously reconstructs cell transition, migration and growth are the keys to reconstructing correct spatiotemporal dynamics; (3) and stVCR can accurately reconstruct key regulatory mechanisms. In addition, stVCR is a scalable and robust algorithm under key hyperparameters tuning.

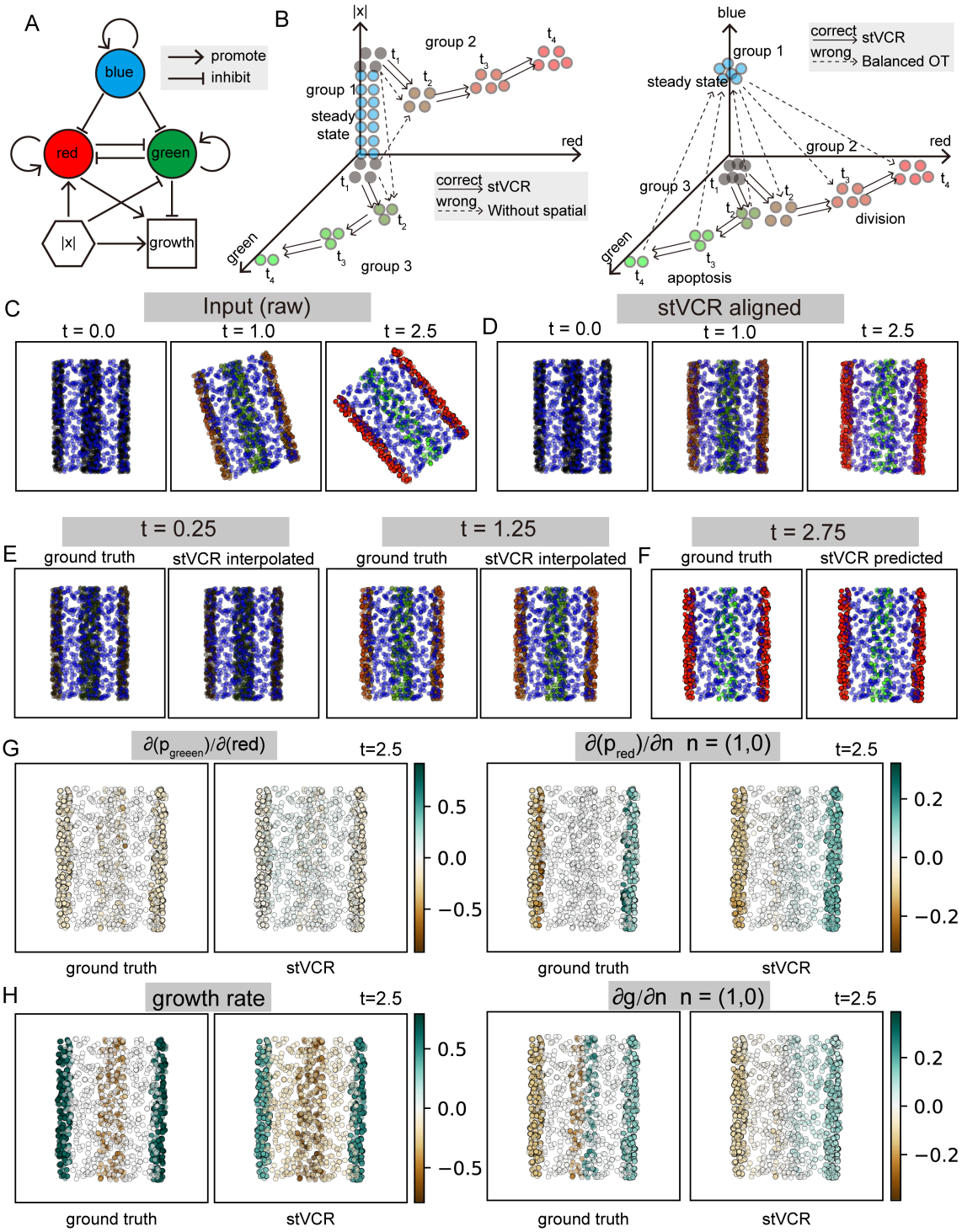

Fig. 2: **Benchmark of stVCR on the simulated time-series ST data. A.** Regulation relationship diagram to generate simulation data. **B.** Dynamic diagram of cell evolution over time. Left: Dynamics of the second group and third group of cells over time in r (red gene expression), g (green gene expression) and $|x|$ (absolute values of spatial coordinates) dimension. Right: Dynamics of all cells over time in r, g and b (blue gene expression) dimension. **C.** Input data to stVCR at $t = 0.0$, 1.0 and 2.5. The color was determined by the expression of three genes *Red*, *Green* and *Blue*. **D.** The reconstructed results at $t = 0.0$, 1.0 and 2.5 of cells at $t = 0.0$ according to learned dynamics using stVCR. **E.** Results of stVCR interpolation at $t = 0.25$ and $t = 1.25$ and comparison with ground truth. Left: ground truth; Right: stVCR. **F.** Similar to E, but for the results of stVCR prediction at $t = 2.75$. **G.** Left: Derivative of *Green* gene velocity with respect to *Red* gene on cells at $t = 2.5$ of true dynamics and learned dynamics. Right: similar to Left, but for derivative of *Red* with respect to given direction $\boldsymbol{n}$, where $\boldsymbol{n} = (1, 0)$. **H.** Left: Growth rates of cells at $t = 2.5$ of true dynamics and learned dynamics. Right: similar to Left, but for derivative of cell proliferation with respect to given direction $\boldsymbol{n}$, where $\boldsymbol{n} = (1, 0)$.

stVCR reconstructs cell transition and growth dynamics of axolotl brain regeneration

To validate stVCR's capability to learn complex continuous dynamics from spatial snapshots, we next applied stVCR for the axolotl brain regeneration dataset using single-cell Stereo-seq technology [49]. The dataset includes brain samples at 2, 5, 10, 15, 20, 30, and 60 days post-injury (DPI) to dissect both immediate wound responses and later regeneration processes. According to the original study [49], the regeneration process is mainly concentrated in 2 DPI to 20 DPI, so we took the data of 5 temporal points at 2, 5, 10, 15 and 20 and reconstructed the dynamic regeneration process using stVCR (Fig. 3, Fig. S6 and Fig. S7).

To inspect stVCR's effect in aligning different samples, we demonstrated the aligned coordinate at different time points Fig. 3A and Fig. S6A. stVCR aligns the spatial coordinates of the data collected at different time points to the same coordinate system making them blend well (Fig. 3B). We further compared the spatial coordinates of each time point before and after the stVCR alignment (Fig. 3B and Fig. S6B), and observed that at each time point the data were adjusted to varying degrees based on the inferred rigid-body transformation, especially for the 20 DPI data (Fig. 3C). This suggests the necessity of sample alignment to infer dynamics correctly.

To further illustrate the continuous dynamics reconstructed, we trained a classifier based on existing cell annotations using a neural network, which allowed us to annotate cells at unobserved time points (Methods). stVCR recovered the gene expression, physical location, possible division and apoptosis, and possible transformation of the cell type of each cell at each moment (Supplementary Video S5). We visualized the calculated spatial velocity in coordinate space and observed cells in the vicinity of the wound migrating toward the wound when the wound was not yet fully healed, showing a response to injury, especially reactive ependymoglial cell (reaEGC) and microglial cell (MCG) (Fig. 3D).

Consistent with cell transition dynamics in response to the injury, we studied the spatial distribution of cell proliferation rates (Fig. 3E and Fig. S6D), which showed that cell proliferation rates in the injured hemispheres were significantly higher than those in the uninjured hemispheres especially near the wound site (Fig. 3E), implying that cell division was more active in the injured hemispheres. This phenomenon may be due to the need to compensate for cells lost due to injury. In addition, we show the interpolation results at 3.5, 7.5, 12.5 and 17.5 DPI (Fig. 3F and Fig. S6C).

To highlight stVCR's function to generate unobserved lineage dynamics, we calculated the number of cells of each type over time based on reconstructed continuous trajectories (Fig. S7A). Interestingly, the inferred number of many types of cells does not simply vary monotonically and linearly outside the observed time point. Among the three ependymoglial cell (EGC) types, the number of reaEGC are increasing first and then decreasing, while the population of Wnt^+ EGC (wntEGC) and $Sfrp^+$ (sfrpEGC) are decreasing first and followed by increasing trend. Such a trend coincides with the original study [49] which revealed that sfrpEGC and wntEGC are transitioned into reaEGC in the earlier stage of immediate wound responses. In contrast, later reaEGC are transitioned into mature neurons (Fig. S7B). In particular, the number of wntEGC decreased while reaEGC population expanded synchronously from 5DPI to 10DPI (Fig. 3G Top).

To better visualize the lineage dynamics inferred by stVCR, we constructed the temporal developmental lineage of wntEGC from 5DPI to 10DPI, which allowed us to study the transformation of cell types at time periods other than the observed time points (Fig. 3H Top). The results showed that wntEGC were indeed partially transformed into reaEGC. Next we constructed the temporal developmental profile of reaEGC from 15DPI to 20DPI, and the results showed that it transformed into wntEGC and some neurons in intermediate and mature states (Fig. 3H Bottom), which also coincided with the trend of their cell number (Fig. 3G Bottom). In addition, we noticed a rapid increase in the number of immature neuron (IMN) and dorsal palliumexcitatory neuron (dpEX) when the number of regeneration intermediate progenitor cell (rIPC)1 and rIPC2 was sharply decreasing (Fig. S7C). Therefore, we constructed temporal developmental lineages of rIPC1 and rIPC2 from 15DPI to 20DPI, and the results showed that rIPC1 were mainly transitioned into IMN and rIPC2 were mainly transitioned into dpEX (Fig. S7D), which is also consistent with the experimental observations [49].

In summary, stVCR describes the complex dynamics of axolotl brain regeneration. In the early wound response phase of an injury, stVCR revealed that sfrpEGC and wntEGC transitioned into reaEGC, and the proliferation of cells in the injured hemisphere became active, especially EGC types in the vicinity of the wound. In addition, reaEGC moved toward the wound. As the wound gradually healed, reaEGC transformed back to wntEGC or differentiated into certain intermediate-state neurons. Eventually, neurons in the intermediate state are then transitioned into mature neurons to compensate for the loss of mature neurons due to injury.

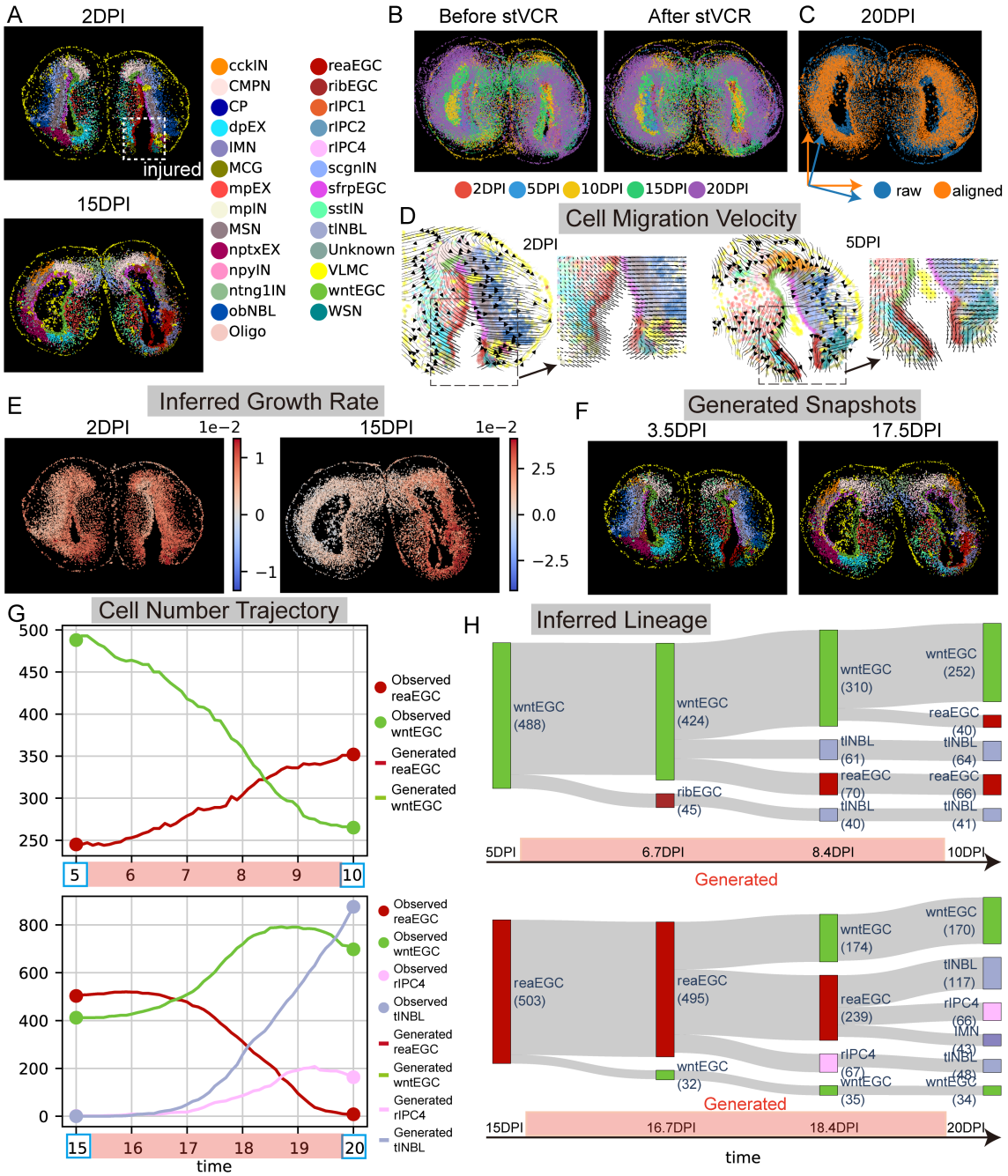

Fig. 3: **stVCR reconstructs spatiotemporal dynamics of axolotl brain regeneration. A.** stVCR aligns the spatial coordinates of data at different time points to the same coordinate system. Top 2DPI; Bottom: 15DPI. Cell type annotations come from the original study. dpEX,dorsal palliumexcitatory neuron; IMN, immature neuron; MCG, microglial cell; MSN, medium spiny neuron; nptxEX, Nptx$^+$ lateral pallium excitatory neuron; EGC, ependymoglial cell; reaEGC, reactive EGC; ribEGC, ribosomal EGC; rIPC, regeneration intermediate progenitor cell; sfrpEGC, Sfrp$^+$ EGC; tlNBL, telencephalon neuroblast; VLMC, vascular leptomeningeal cell; wntEGC, Wnt$^+$ EGC. **B.** Comparison of spatial coordinates of the data for all time points before and after stVCR alignment. Left: Before stVCR. Right: After stVCR. **C.** Comparison of spatial coordinates of 20 DPI data before and after alignment. **D.** stVCR inferred spatial cell migration velocity at 2DPI and 5DPI injured hemispheres. Left: Streamline plot; Right: Locally amplified grid velocity. **E.** stVCR inferred cell proliferation rate at 2DPI and 15DPI data. **F.** stVCR interpolated snapshots at 3.5DPI and 17.5DPI. Cell type annotations come from the stVCR's time-dependent classifier based on the generated continuous gene expression values. **G.** Changes in cell number over time. Top: reaEGC and wntEGC from 5DPI to 10DPI. Bottom: reaEGC, wntEGC, rIPC4 and tlNBL from 15DPI to 20DPI. **H.** stVCR reconstructs the time-varying developmental lineages. The number in parentheses is the number of cells. Top: wntEGC from 5DPI to 10DPI; Bottom: reaEGC from 15DPI to 20DPI.

Gene-level mechanisms of axolotl brain regeneration revealed by stVCR

stVCR reconstructed the dynamics of axolotl brain regeneration process, in which cell migration and reaEGC proliferation play important roles. To further reveal the mechanism on the gene level, we performed stVCR analysis of these two biological processes based on the learned kinetics, and in addition, reconstructed the regulatory network between key genes (Fig. 4, Fig. S8 and Fig. S9).

To infer gene-spatial interactions, we used stVCR to identify the top 100 migration driver genes (Methods). Through Gene Ontology (GO) biological process analysis, we identified several processes associated with cell migration (Fig. 4A), including neuron migration and negative regulation of homotypic cell-cell adhesion. Neuron migration is a crucial process for the proper positioning of neurons, while the negative regulation of homotypic cell-cell adhesion facilitates cell movement by reducing cell interactions. Interestingly, several marker genes of EGC-type cells identified in the original study, were included in stVCR's migration driver genes, such as *GFAP*, *TNC*, *PTN*, *SLC1A3*, *GLUD1* and *ECM1*. We further visualized the migration driver gene score, i.e. $\partial\|\boldsymbol{v}\|/\partial q^j$ of four example genes *GFAP*, *TNC*, *PTN* and *SLC1A3* (Fig. 4B and Fig. S8A), and showed that they have a promoting effect on cell migration in EGC-type cells. In addition, these genes are indeed highly expressed in EGC-type cells (Fig. 4C and Fig. S8B), especially the *GFAP*, *TNC* genes in reaEGC cells (Fig. 4C).

Next, we used stVCR to infer cell proliferation driver genes in reaEGC and ranked the results to obtain the top 100 growth-promoting genes (Methods). The GO analysis identified several processes closely related to cell proliferation (Fig. S8C) essential for ribosome biogenesis, protein synthesis, and the proper targeting of proteins. Similar to the migration driver gene results, the growth driver genes inferred by stVCR overlapped with several marker genes for EGCs, such as *FABP7* and *SFRP1* (Fig. S8E). We visualized the growth driver gene score $\partial g/\partial q^j$ of these two example genes (Fig. S8D), and showed that they are significantly promoting cell proliferation and division at EGCs in the injured hemisphere.

Finally, to utilize stVCR's function to infer dynamic gene interactions(Methods), we selected some genes (*KRT18*, *ECM1*, *GFAP*, *VIM*, *TNC*, *S100A10* and *HMOX1*) that were highly expressed in reaEGC (Fig. S9A), and investigated the regulatory relationship between these genes 1) at different time points and 2) in different cell types, and visualized the gene regulatory network (GRN) (Fig. 4D,E and Fig. S9B,C). We observed that in reaEGC these genes inhibit each other at an early stage (2DPI), followed by a gradual weakening of the inhibition (5DPI; 10DPI), and at a later stage (15DPI) they turn to promote each other (Fig. 4D and Fig. S9B). Thus, stVCR analysis suggests that gene regulatory relationships may be changing over time, even in the same cell type, which may be related to the discovery that reaEGC play different roles in early and late stages of injury. In addition, to investigate the gene regulatory relationships as affected by spatial distribution, we selected wntEGC and sfrpEGC, which are closer to reaEGC, and vascular leptomeningeal cell (VLMC), whereas more apart from reaEGC (Fig. 4E). We recovered the regulatory relationships between these cells at 2DPI at the previously mentioned genes (Fig. 4F and Fig. S9C). The results showed that the regulatory relationships of these genes were close in wntEGC and sfrpEGC, and closer to reaEGC, although there were some minor differences (Fig. 4F Left and Middle). In contrast, the regulatory relationship in VLMC was distinct from reaEGC (Fig. 4F Right). Thus, our results suggest that gene regulatory relationships might also be influenced by cell type and spatial location.

In summary, our gene-level mechanism of axolotl brain regeneration datasets demonstrates the ability of the stVCR to (1) find migration-driven and growth-driven genes; and (2) infer time-dependent and cell type- dependent GRN, suggesting its advantages over static OT-based methods.

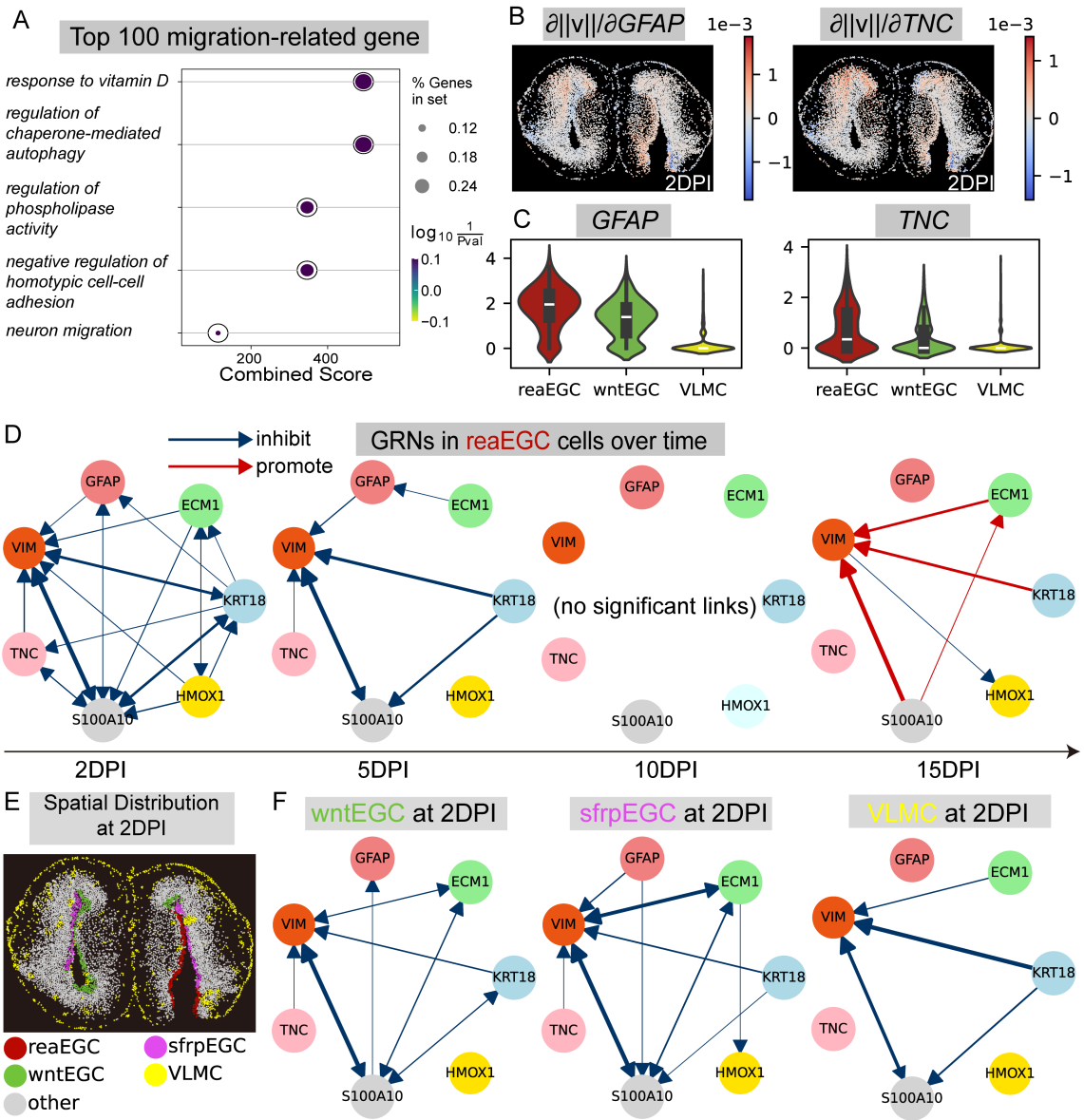

Fig. 4: stVCR gene-level analysis of axolotl brain regeneration. A. GO biological process enrichment analysis of the top 100 migration-promoting genes in all cells. **B.** The partial derivative of the norm of spatial velocity $\|v_z\|$ with respect to gene expression. Two example genes in **A**, *GFAP* (Left) and *TNC* (Right). **C.** Violin plots of gene *GFAP* (Left) and *TNC* (Right) expression in reaEGC, wntEGC and VLMC cells. **D.** Gene regulatory networks in reaEGC cells for genes highly expressed in reaEGC. From left to right: 2DPI, 5DPI, 10DPI and 15DPI. **E.** Spatial distribution of reaEGC, wntEGC, sfrpEGC and VLMC cells at 2 DPI. **F.** Similar to **D**, but for sfrpEGC (Left), wntEGC (Middle) and VLMC (Right) cells at 2 DPI.

stVCR analysis of 3D Drosophila embryos and organs with optional prior

To illustrate the necessity of incorporating known biological priors for ST data with sparse temporal observations, we begin with a specially designed simulation dataset for benchmarking (Methods and Supplementary Note 3). The data consists of three types of cells named type 1, type 2, and type 3 (Fig. 5A), where type 3 cells express the *Red* and *Green* genes moderately and are at steady state. The type 1 cells will first highly express *Red* and *Green* genes, gradually decrease the expression of *green* genes, and migrate over time, transitioning to the type 2 cells that highly express only the *Red* gene. When there are sufficient observations and the time intervals are small enough, the correct result can be inferred by stVCR without any prior (Fig. S11 and Supplementary Video S8). Indeed, we can theoretically prove that the stVCR reconstructed dynamics will converge to the true dynamics when the sampling time intervals between consecutive observations converge to zero, which provides a rigorous guarantee for the algorithm (Supplementary Note 4). Meanwhile, due to the high cost of ST sequencing, the number of measurement time points is usually fewer

and at longer intervals in real experiments. Thus we develop the strategy to allow users to assign known biology priors such as cell state-transition relations and structural continuity of tissues into stVCR (Methods).

To show the effects of adding biological priors, we only input two observations at $t = 0$ and $t = 1$ (Fig. 5A,C) to stVCR. Without prior knowledge, stVCR would infer the wrong type 1 to type 3 and type 3 to type 2 transitions rather than the correct type 1 to type 2 transition (Fig. 5, Supplementary video S6-S7). In comparison, the interpolation results show that stVCR with a type 1 to type 2 state-transition prior is closer to the ground truth at unobserved time points (Fig. 5D and Fig. S10). Overall, the above experimental results illustrate the benefit of adding the correct biological prior into datasets with fewer observations and longer intervals for more accurate results.

In order to validate the effectiveness of the strategy of combining biological prior and spatial structure preserving prior on real datasets, we next applied stVCR priors for the 3D Drosophila embryos and organs dataset using single-cell Stereo-seq technology [43]. This datasets include only two time points E7-9h and E9-10h. We set the former moment (E7-9h) in the 3D Drosophila embryo data to $t = 8$h and the latter moment (E9-10h) to $t = 9.5$h. The data contains 9 tissues (Fig 5E), and we added biological priors central nervous system (CNS) transition to CNS, midgut transition to midgut, and amnioserosa transition to amnioserosa. Additionally, we added the spatial structure preserving priors for CNS and midgut. We aligned the two observations of data and reconstructed the dynamics between the two moments using the stVCR with the above priors (Supplementary video S9). Fig. 5E shows the aligned 3D Drosophila embryo. We focused on the CNS and midgut (Fig. 5F). We observed that the anterior of the CNS of Drosophila at the latter moment overlapped with the posterior of the CNS at the former moment (Fig. 5F Left), and the midgut consisted of two parts at the former moment and only one part at the latter moment (Fig. 5F Right).

To benchmark stVCR with other static OT-based methods and highlight its unique function to model continuous dynamics, we compared the spatial migration dynamic trajectories of CNS reconstructed by stVCR with Spateo [43] and Moscot [41] (Supplementary video S10-S12 and Fig. 5G). Since both Spateo and Moscot are based on static OT and do not directly reconstruct the intermediate process, we obtained the intermediate process by linear interpolation based on the inferred static optimal map. In the spatial trajectory reconstructed by stVCR, the cells in the posterior of CNS gradually migrated to the anterior along the internal structure of the CNS (Supplementary video S10 and Fig. 5G Top row). In Spateo, the cells at the posterior of the CNS were disconnected from the main body and then migrated to the anterior to merge into one part (Supplementary video S11 and Fig. 5G Middle row). One possible explanation is that Spateo is based on static OT and does not constrain the consistency of the intermediate trajectory. In Moscot, cells migrate and aggregate to a few locations (Supplementary video S12 and Fig. 5G Bottom row). We speculate the possible reason is that Moscot tackles the unbalanced OT problem by adding the KL divergence penalty, so that the cells at the first moment correspond to a few cells at the second moment. In addition, we reconstructed the spatial trajectory of midgut using stVCR and compared it with Spateo and Moscot (Supplementary video S13-S15 and Fig. S12). The results showed that both stVCR and Spateo observed that two parts of the midgut at the first moment merged into one part at the later moment (Supplementary video 13-14 and Fig. S11C First two row). In Moscot, the spatial trajectories of midgut migrated and aggregated to a small number of locations similarly to the CNS results.

In summary, we demonstrate the theoretical convergence of stVCR in large sample cases through mathematical derivations. We also highlight the benefit of adding known state-transition priors and spatial structure-preserving priors in case of limited observations through computations on simulated data. The application in 3D Drosophila datasets indicates the superiority of stVCR compared to existing methods based on static OT.

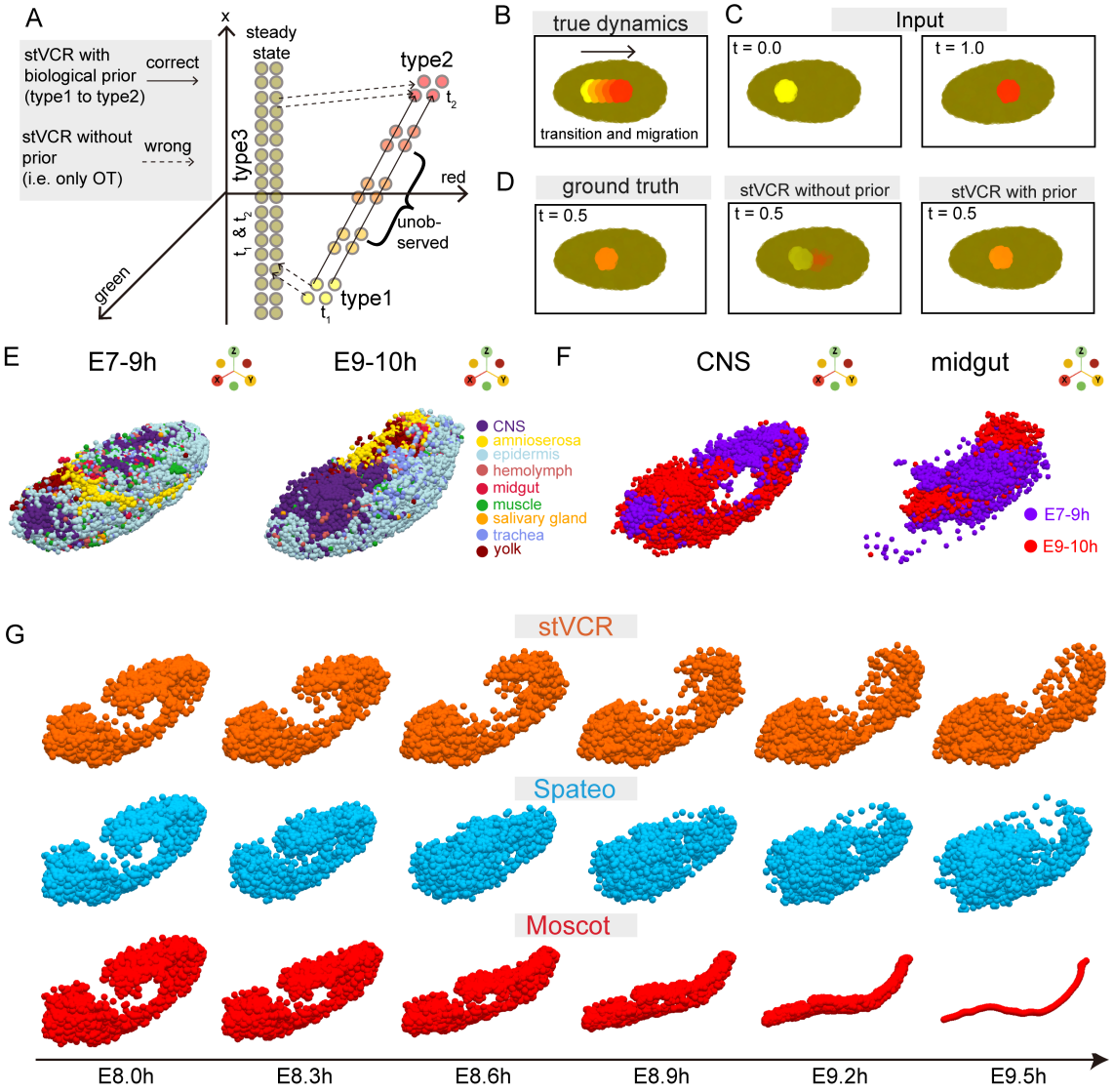

Fig. 5: **stVCR analysis of 3D Drosophila embryos and organs with biological state-transition prior and spatial structure preserving prior. A.** Dynamic diagram of cell evolution over time in simulated data. Longer observation intervals and lack of biological knowledge guidance will result in incorrect type 3 to type 1 and type 2 transitions rather than type 1 to type 2 transitions. **B.** True dynamics of simulated data. Yellow type 1 cells transition and migrate to become red type 2 cells. Brown cells in the background are in steady state. The color was determined by the expression of three genes *Red*, *Green* and *Blue*. **C.** Input data to stVCR at only two time points. Left: $t = 0.0$. Right: $t = 1.0$. **D.** Interpolation results at $t = 0.5$ of stVCR with and without biological prior compared to the ground truth. Left: ground truth. Middle: stVCR without prior. Right: stVCR with prior. **E.** Spatial coordinates of 3D Drosophila embryos after stVCR alignment. Left: E7-9h. Right: E9-10h. CNS, central nervous system. **F.** Spatial coordinates of 3D Drosophila organs after stVCR alignment. Left: CNS. Right: Midgut. **G.** Comparison of spatial migration trajectories of CNS cells. From left to right: E8.0h, E8.3h, E8.6h, E8.9h, E9.2h and E9.5h. From top to bottom: stVCR, Spateo [43] and Moscot [41].

Discussion

Time-series spatial transcriptomics data has made it possible to reconstruct the entire spatiotemporal dynamic process of cell fate determination. To dynamically connect unpaired snapshots and align temporal slices from various coordinate systems, we present stVCR to (1) simultaneously reconstruct and continuously generate cell differentiation, migration in physical space as well as division and apoptosis; (2) align spatial coordinates from data collected at different time points and (3) investigate the complex interactions between cell phenotype transitions, spatial migration, and proliferation.

Compared to existing methods that reconstruct trajectories from time-series spatial transcriptomics data [41, 43, 44], stVCR employs Rigid body transformation invariant OT, rather than GW-OT, making it the first algorithm to use dynamic OT modeling for time-series spatial transcriptomics data. Additionally, stVCR can model optional known biological priors and spatial structure-preserving priors. Once the entire dynamic process is reconstructed, stVCR can further perform a series of downstream analyses that are not feasible with static OT.

stVCR can be improved and extended from several aspects. For instance, learning both the low-dimensional representation of gene expression and the dynamics in the low-dimensional simultaneously may yield better results [18]. In addition, integrating ODE represented by neural networks is costly, and some work has attempted to construct simulation-free methods in time-series single-cell transcriptome data [50], which may also be generalized to spatial transcriptome data. Incorporating more intrinsic dynamics of gene expression (such as transcription, splicing, and degradation) and cellular interactions[51, 52] may yield deeper insights into the realistic biological processes. Lastly, integrating other modalities such as lineage information [53] and multi-omics measurements [54] could further enhance the trajectory inference. Due to the limitations of 3D time-series spatial transcriptome data availability, further applications and performance evaluations on more challenging datasets are needed and will be addressed in future research.

Overall, stVCR provides a unified and robust method for generative modeling of time-series spatial transcriptomics data, which reconstructs the entire spatiotemporal processes of single cells from a few given snapshots and investigates the complex space-gene regulatory mechanisms.

Methods

Basic optimal transport formulation setup

In essential aspects, we utilize the dynamical optimal transport (OT) formulation to reconstruct the spatiotemporal dynamics of single cells for snapshot spatial transcriptomics data. Let us first state the basic OT setup for simple case.

Let $\alpha = \sum_{i=1}^{n} a_i \delta(\boldsymbol{x} - \boldsymbol{x}_i)$ and $\beta = \sum_{j=1}^{m} b_j \delta(\boldsymbol{y} - \boldsymbol{y}_j)$ be two probability distributions with normalized positive vectors $\boldsymbol{a} = (a_i)$ and $\boldsymbol{b} = (b_j)$, where $\delta(\cdot)$ stands for the Dirac's δ-function. One typically couples α and β (or $\boldsymbol{a}$ and $\boldsymbol{b}$) through the Kantorovich's OT problem

$$\mathcal{L}_c(\alpha, \beta) = \min_{P \in U(\boldsymbol{a}, \boldsymbol{b})} \langle C, P \rangle := \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} p_{ij}, \tag{4}$$

where $U(\boldsymbol{a}, \boldsymbol{b}) := \left\{ P \in \mathbb{R}_+^{n \times m} : P \mathbb{1}_m = \boldsymbol{a} \text{ and } P^{\mathrm{T}} \mathbb{1}_n = \boldsymbol{b} \right\}$ and $C = (c_{ij})$ is the cost matrix. When $c_{ij} = c(\boldsymbol{x}_i, \boldsymbol{y}_j)$ with $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_p^p$, where $\|\boldsymbol{x}\|_p$ is the vector ℓ^p norm, the p-Wasserstein distance is defined as $W_p(\alpha, \beta) = (\mathcal{L}_c(\alpha, \beta))^{1/p}$. The optimal coupling matrix component p_{ij} characterizes the probability that $\boldsymbol{x}_i$ will be transported to $\boldsymbol{y}_j$ [55].

A special case is $p = 2$, i.e., $c(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$. In this case, the above OT formulation has an equivalent dynamic form (Benamou-Brenier form [32]) by minimizing the transport kinetic energy

$$\left(W_2(\alpha, \beta)\right)^2 = \min_{(\alpha_t(\boldsymbol{x}), \boldsymbol{v}_t(\boldsymbol{x}))} \int_0^1 \int_{\mathbb{R}^d} \|\boldsymbol{v}_t(\boldsymbol{x})\|^2 \alpha_t(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \mathrm{d}t, \tag{5}$$

where $\alpha_t, \boldsymbol{v}_t$ satisfies the continuity partial differential equation (PDE)

$$\partial_t \alpha_t(\boldsymbol{x}) + \nabla \cdot (\boldsymbol{v}_t(\boldsymbol{x}) \alpha_t(\boldsymbol{x}))) = 0, \text{ such that } \alpha_{t=0} = \alpha, \alpha_{t=1} = \beta. \tag{6}$$

The vector field $\{\boldsymbol{v}_t(\boldsymbol{x})\}_{t \in [0,1]}$ is to be optimized such that the boundary conditions are satisfied and the minimal kinetic energy is achieved.

The dynamic formulation can be generalized to the case when the total mass of α and β are not equal (unbalanced setting). A common approach is to consider the so-called Wasserstein-Fisher-Rao (WFR) distance [35, 56, 57]

$$\left(\mathrm{WFR}\left(\alpha, \beta\right)\right)^2 = \min_{(\alpha_t(\boldsymbol{x}), \boldsymbol{v}_t(\boldsymbol{x}), g_t(\boldsymbol{x}))} \int_0^1 \int_{\mathbb{R}^d} \left(\|\boldsymbol{v}_t(\boldsymbol{x})\|^2 + \tau \|g_t(\boldsymbol{x})\|^2 \right) \alpha_t(\boldsymbol{x}) \, \mathrm{d}\boldsymbol{x} \mathrm{d}t, \tag{7}$$

where $\alpha_t, \boldsymbol{v}_t, g_t$ satisfies the PDE

$$\partial_t \alpha_t(\boldsymbol{x}) + \nabla \cdot (\boldsymbol{v}_t(\boldsymbol{x})\alpha_t(\boldsymbol{x})) = g_t(\boldsymbol{x})\alpha_t(\boldsymbol{x}), \text{ such that } \alpha_{t=0} = \alpha, \alpha_{t=1} = \beta \tag{8}$$

See Supplementary Note 1 for a detailed description of the above algorithm.

Data preprocessing

For gene expression count matrix, we first normalized the raw counts data using size factor. Then we selected the top 2000 highly variable genes. Finally, we utilize an Autocoder to project highly variable genes to low dimensions. Specifically, we represent an encoder $\boldsymbol{q}_{\text{emb}} = f_{\text{enc}}(\boldsymbol{q}_{\text{ori}}, \theta_{\text{enc}})$ and a decoder $\tilde{\boldsymbol{q}}_{\text{ori}} = f_{\text{dec}}(\boldsymbol{q}_{\text{emb}}, \theta_{\text{dec}})$ using neural networks, where the input of the encoder is the original gene expression $\boldsymbol{q}_{\text{ori}}$ and the output is the low-dimensional embedding $\boldsymbol{q}_{\text{emb}}$, and the decoder is the opposite. The loss function is taken as

$$L(\theta_{\text{enc}}, \theta_{\text{dec}}) = \frac{1}{N}\sum_{i=1}^{N} \|\boldsymbol{q}_{\text{ori},i} - \tilde{\boldsymbol{q}}_{\text{ori},i}\|_2^2, \tag{9}$$

where N refers to the total number of cells. In actual computations, we take the dimension of the low-dimensional embedding $\boldsymbol{q}_{\text{emb}}$ to be 10. Also for simplicity of notation, we still use $\boldsymbol{q}$ to refer to $\boldsymbol{q}_{\text{emb}}$ to denote the low-dimensional embedding of gene expression unless otherwise stated.

Notation conventions in stVCR

We use the notation $\boldsymbol{x} = (x^1, x^2, \ldots, x^{d_s}) \in \mathbb{R}^{d_s}$, $\boldsymbol{q} = (q^1, q^2, \ldots, q^{d_g}) \in \mathbb{R}^{d_g}$ for spatial and gene expression variables, respectively. For the considered spatio-temporal transcriptome data, we assume there are n_k cells at time $t = t_k$ for $k = 0, 1, \ldots, K$. We denote the available datasets by

$$X^{(k)} = (\boldsymbol{x}_1^{(k)}, \boldsymbol{x}_2^{(k)}, \ldots, \boldsymbol{x}_{n_k}^{(k)}) \in \mathbb{R}^{d_s \times n_k}, \; Q^{(k)} = (\boldsymbol{q}_1^{(k)}, \boldsymbol{q}_2^{(k)}, \ldots, \boldsymbol{q}_{n_k}^{(k)}) \in \mathbb{R}^{d_g \times n_k} \tag{10}$$

at $t = t_k$, where d_s is the dimension of the spatial coordinates, usually 2 or 3, and d_g is the dimension of the embedded gene expression space ($d_g = 10$ in our setup). In the data analysis, we often need to consider the spatial coordinates after alignment with rigid body transformations at time points $t = t_1, \ldots, t_K$, which we denote by

$$\hat{X}^{(k)} = (\hat{\boldsymbol{x}}_1^{(k)}, \hat{\boldsymbol{x}}_2^{(k)}, \ldots, \hat{\boldsymbol{x}}_{n_k}^{(k)}), \; \hat{Q}^{(k)} = Q^{(k)}, \; k = 1, 2, \ldots, K. \tag{11}$$

When only rotation R and translation $\boldsymbol{r}$ are considered, $\hat{\boldsymbol{x}}^{(k)} = R\boldsymbol{x}^{(k)} + \boldsymbol{r}$. The data (empirical) distribution formed by the dataset $(X^{(k)}, Q^{(k)})$ is denoted by

$$\rho^{(k)}(\boldsymbol{x}, \boldsymbol{q}) = \frac{1}{n_k}\sum_{i=1}^{n_k} \delta(\boldsymbol{x} - \boldsymbol{x}_i^{(k)}, \boldsymbol{q} - \boldsymbol{q}_i^{(k)}), \tag{12}$$

where $\delta(\cdot)$ stands for the Dirac's δ-function in $(\boldsymbol{x}, \boldsymbol{q})$-space. Correspondingly, we can define $\hat{\rho}^{(k)}$ for the dataset $(\hat{X}^{(k)}, \hat{Q}^{(k)})$ after alignment. Both $\rho^{(k)}$ and $\hat{\rho}^{(k)}$ are probability distributions.

We use the notation $f_t(\boldsymbol{x}, \boldsymbol{q})$ or the shorthand f_t to denote the time dependence of a function $f(\boldsymbol{x}, \boldsymbol{q}, t)$, while for time dependent coordinates $(\boldsymbol{x}(t), \boldsymbol{q}(t))$, we use the notation $(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)})$ to be consistent with the notation $(X^{(k)}, Q^{(k)})$. All of the vectors are bold-faced, while the matrices are in normal fonts.

Spatial alignment for temporal snapshots

Due to the unknown deformation of the tissue during spatial transcriptome (ST) sequencing, we need to align the spatial coordinates of the ST data at different time points before the subsequent analysis. We assume the large scale deformation of the observed coordinates contains only rigid body transformations (i.e., rotations and translations) in this work. We do not follow the Gromov-Wasserstein (GW) optimal transport framework for its over-generality and large computational cost. Yet, we adopt the optimal transport approach in [58] by explicitly modeling the set of invariant

manipulations $\mathcal{G}$, and simultaneously finding the optimal matching of distributions and the optimal transformation through the optimization problem:

$$(P^\star, g^\star) = \underset{P\in U(\boldsymbol{a},\boldsymbol{b}), g\in\mathcal{G}}{\arg\min} \langle C(g), P\rangle \overset{\text{def.}}{=} \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij} d\big(\boldsymbol{x}_i, g(\boldsymbol{y}_j)\big). \tag{13}$$

It can be solved by an iterative algorithm

$$P^{(n)} = \underset{P\in U(\boldsymbol{a},\boldsymbol{b})}{\arg\min} \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij} d\big(\boldsymbol{x}_i, g^{(n)}(\boldsymbol{y}_j)\big), \tag{14}$$

$$g^{(n+1)} = \underset{g\in\mathcal{G}}{\arg\min} \sum_{i=1}^{M}\sum_{j=1}^{N} p_{ij}^{(n)} d\big(\boldsymbol{x}_i, g(\boldsymbol{y}_j)\big). \tag{15}$$

The subproblem (14) is to solve a static OT.

When we take the set $\mathcal{G} = \{(R, \boldsymbol{r})\}$ as the rigid body transformations, and the distance function d as the Euclidean distance, that is, $d(\boldsymbol{x}, \boldsymbol{y}) = \|\boldsymbol{x} - \boldsymbol{y}\|_2^2$, the subproblem (15) is a weighted Procrustes problem [59]

$$(R^{(n+1)}, \boldsymbol{r}^{(n+1)}) = \underset{\substack{R\in\mathbb{R}^{d_s\times d_s}, \boldsymbol{r}\in\mathbb{R}^{d_s}\\ R^TR=I,\ \det R=1}}{\arg\min} \sum_{i,j} p_{ij}^{(n)} \left\|\boldsymbol{x}_i - (R\boldsymbol{y}_j + \boldsymbol{r})\right\|_2^2, \tag{16}$$

where R refers to a rotation matrix, and $\boldsymbol{r}$ refers to a translation vector.

We call this problem *rigid body transformation invariant optimal transport*, and it explicitly models unknown rigid body transformation, making it possible to compute ground cost functions between distributions at different time points. It can be easily modified to model more general cases, e.g., if affine transformations are allowed in the tissue measuring process.

Dynamics loss in stVCR

We are concerned with the entire dynamics of cell population evolution and therefore use dynamical optimal transport. Since the number of cell population is constantly changing due to cell division and apoptosis during evolution, unbalanced setting is necessary.

stVCR reconstructs $\rho_t(\boldsymbol{x}, \boldsymbol{q})$ by interpolating the input population densities $\rho^{(k)}$ up to a normalization at $t = t_k$ using a transport-with-growth PDE

$$\partial_t \rho_t(\boldsymbol{x}, \boldsymbol{q}) + \nabla\cdot\Big(\big(\boldsymbol{v}_t(\boldsymbol{x}, \boldsymbol{q}), \boldsymbol{p}_t(\boldsymbol{x}, \boldsymbol{q})\big)\rho_t(\boldsymbol{x}, \boldsymbol{q})\Big) = g_t(\boldsymbol{x}, \boldsymbol{q})\rho_t(\boldsymbol{x}, \boldsymbol{q}), \tag{17}$$

where $\boldsymbol{v}_t(\boldsymbol{x}, \boldsymbol{q})$ refers to the spatial velocity, $\boldsymbol{p}_t(\boldsymbol{x}, \boldsymbol{q})$ refers to the RNA velocity of cells, and $g_t(\boldsymbol{x}, \boldsymbol{q})$ refers to the proliferation rate of cells. We take the coordinate system at t_0 as the reference, and assume that the coordinate system at t_k $(k = 1, 2, \ldots, K)$ differs from the coordinate system at t_0 by a rotation R_k and translation $\boldsymbol{r}_k$. Thus the feasible state space $\mathcal{S}$ for the arguments under constraints is

$$\begin{aligned}\mathcal{S}(X^{(0:K)}, Q^{(0:K)}) := \big\{(\rho_t, \boldsymbol{v}_t, \boldsymbol{p}_t, g_t; R_{1:K}, \boldsymbol{r}_{1:K})\big|\ \partial_t\rho_t + \nabla\cdot((\boldsymbol{v}_t, \boldsymbol{p}_t)\rho_t) = g_t\rho_t,\\ \rho_{t_0} = \rho^{(0)}, \|\rho_{t_k}\|_1 = n_k/n_0, \bar{\rho}_{t_k} = \hat{\rho}^{(k)}, R_k^TR_k = I, \det R_k = 1,\ k = 1, 2, \ldots, K\big\},\end{aligned} \tag{18}$$

where $\hat{\rho}^{(k)}$ refers to the distribution formed by $(\hat{X}^{(k)}, \hat{Q}^{(k)})$ after alignment through rigid body transformation $(R_k, \boldsymbol{r}_k)$ at t_k, $\bar{\rho}_t := \rho_t/\|\rho_t\|_1$ and $\|\rho_t\|_1 := \int \rho_t \mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{q}$ is the total mass of ρ_t. The notation $0{:}K$ is a shorthand for the indices $\{0, 1, \ldots, K\}$. Same rule applies to similar notations in other places.

Borrowing the idea of Wasserstein-Fisher-Rao (WFR) distance for unbalanced optimal transport [34], we obtain the optimal dynamics $(\rho_t, \boldsymbol{v}_t, \boldsymbol{p}_t, g_t)$ and the optimal rigid body transformation $(R_k, \boldsymbol{r}_k)$ for $k = 1, 2, \ldots, K$ by minimizing the kinetic and growth energy

$$\int_{t_0}^{t_K}\int_{\mathbb{R}^{d_g+d_s}} \left(\|\boldsymbol{v}_t\|^2 + \alpha_{\text{Exp}}\|\boldsymbol{p}_t\|^2 + \alpha_{\text{Gro}}\|g_t\|^2\right)\rho_t(\boldsymbol{x}, \boldsymbol{q})\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{q}\mathrm{d}t \tag{19}$$

for $(\rho_t, \boldsymbol{v}_t, \boldsymbol{p}_t, g_t; R_{1:K}, \boldsymbol{r}_{1:K}) \in \mathcal{S}(X^{(0:K)}, Q^{(0:K)})$. Direct derivation of the solution of the Feynman-Kac type PDE (17) by characteristics [34, 60] shows the above loss function can be rewritten as the following dynamics loss

$$\begin{aligned}\mathcal{L}_{\mathrm{Dyn}} = \mathbb{E}_{(\boldsymbol{x}^{(t_0)},\boldsymbol{q}^{(t_0)})\sim\rho^{(0)}} \int_{t_0}^{t_K} \Big(&\|\boldsymbol{v}_t(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)})\|^2 + \alpha_{\mathrm{Exp}}\|\boldsymbol{p}_t(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)})\|^2 \\ &+ \alpha_{\mathrm{Gro}}\|g_t(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)})\|^2 \Big) w_t[\boldsymbol{x}, \boldsymbol{q}]\mathrm{d}t,\end{aligned} \tag{20}$$

where $\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)}$ satisfies the characteristic ordinary differential equations (ODEs)

$$\frac{\mathrm{d}\boldsymbol{x}^{(t)}}{dt} = \boldsymbol{v}_t(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)}),\ \frac{\mathrm{d}\boldsymbol{q}^{(t)}}{\mathrm{d}t} = \boldsymbol{p}_t(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)}),\ (\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)})|_{t=t_0} = (\boldsymbol{x}^{(t_0)}, \boldsymbol{q}^{(t_0)}) \tag{21}$$

and the weights $w_t[\boldsymbol{x}, \boldsymbol{q}] = e^{\int_{t_0}^t g_s(\boldsymbol{x}^{(s)},\boldsymbol{q}^{(s)})\mathrm{d}s}$ satisfies the ODE

$$\frac{\mathrm{d}\ln w_t}{\mathrm{d}t} = g_t(\boldsymbol{x}^{(t)}, \boldsymbol{q}^{(t)}),\ w_t|_{t=t_0} = 1. \tag{22}$$

In fact, the density ρ_t has the representation $\rho_t(\boldsymbol{x}, \boldsymbol{q}) = \mathbb{E}_{(\boldsymbol{x}^{(t_0)},\boldsymbol{q}^{(t_0)})\sim\rho^{(0)}}\delta(\boldsymbol{x}-\boldsymbol{x}^{(t)}, \boldsymbol{q}-\boldsymbol{q}^{(t)})w_t[\boldsymbol{x}, \boldsymbol{q}]$ with the total mass $\|\rho_t\|_1 = \mathbb{E}_{(\boldsymbol{x}^{(t_0)},\boldsymbol{q}^{(t_0)})\sim\rho^{(0)}} w_t[\boldsymbol{x}, \boldsymbol{q}]$. The three parts in $\mathcal{L}_{\mathrm{Dyn}}$ correspond to the kinetic energy of spatial migration, $\mathcal{L}_{\mathrm{Spa}}$, kinetic energy of gene expression change, $\mathcal{L}_{\mathrm{Exp}}$, and growth energy, $\mathcal{L}_{\mathrm{Gro}}$, respectively, considered in Eq. (3).

The formulation (20) is suitable for the numerical evaluation of the loss function through Monte Carlo particle simulations instead of density estimation in high dimensional space. We also remark that the evolution of (17) can be implemented in the forward (from t_0 to t_K) or backward (from t_K to t_0) way, and similar formulations as above can be obtained correspondingly. Both directions are taken in our computations for a more robust performance.

Matching loss in stVCR

The constraints $\|\rho_{t_k}\|_1 = n_k/n_0$ and $\bar{\rho}_{t_k} = \hat{\rho}^{(k)}$ in (18) are indeed realized as soft penalties to perform distribution matching. This matching between the total mass, $\{\bar{\rho}_{t_k}\}_{k=1:K}$ and $\hat{\rho}^{(1:K)}$, and the determination of $\{(R_{1:K}, \boldsymbol{r}_{1:K})\}$ are obtained simultaneously in terms of the rigid body transformation invariant OT.

Define the weights $w_{t,j} = w_t[\boldsymbol{x}_j, \boldsymbol{q}_j]$ for the cell j with initial state $(\boldsymbol{x}_j^{(0)}, \boldsymbol{q}_j^{(0)})$ at $t = t_0$. With this notation, we have the evolved distribution

$$\rho_t = \frac{1}{n_0}\sum_{j=1}^{n_0} w_{t,j}\delta(\boldsymbol{x} - \boldsymbol{x}_j^{(t)}, \boldsymbol{q} - \boldsymbol{q}_j^{(t)}).$$

The total mass of ρ_t is $\sum_{j=1}^{n_0} w_{t,j}/n_0$, which no longer corresponds to a probability distribution. This non-normalization is due to the cell proliferation.

The matchings between the total mass, and the normalized distributions $\{\bar{\rho}_{t_k}\}_k$ and $\{\hat{\rho}^{(k)}\}_k$ is performed through the loss function

$$\mathcal{L}_{\mathrm{Mch}} = \sum_{k=1}^{K}\left(W_2(\bar{\rho}_{t_k}, \hat{\rho}^{(k)})\right)^2 + \kappa_{\mathrm{Gro}}\sum_{k=1}^{K}\frac{|\sum_{j=1}^{n_0} w_{t_k,j} - n_k|}{n_k}, \tag{23}$$

where the second term penalizes the total mass mismatch through their relative error, and the first term penalizes the normalized distribution mismatch which we take as the square of the 2-Wasserstein distance between $\{\bar{\rho}_{t_k}\}_k$ and $\{\hat{\rho}^{(k)}\}_k$:

$$\left(W_2(\bar{\rho}_{t_k}, \hat{\rho}^{(k)})\right)^2 := \min_{P\in U(\bar{\rho}_{t_k}, \hat{\rho}^{(k)})}\langle C^{(k)}, P\rangle.$$

Naturally, we take the cost matrix $C^{(k)}$ by integrating gene expression and spatial coordinates with components

$$c_{ij}^{(k)} = \kappa_{\mathrm{Exp}}\|\boldsymbol{q}_i^{(t_k)} - \hat{\boldsymbol{q}}_j^{(k)}\|_2^2 + (1-\kappa_{\mathrm{Exp}})\|\boldsymbol{x}_i^{(t_k)} - \hat{\boldsymbol{x}}_j^{(k)}\|_2^2,\quad i = 1:n_0, j = 1:n_k$$

where $\hat{\boldsymbol{x}}_j^{(k)} = R_k \boldsymbol{x}_j^{(k)} + \boldsymbol{r}_k$ is the coordinate after the alignment manipulation, κ_{Exp} is an adjustable hyperparameter that measures the importance of gene expression or spatial coordinates.

Modeling known cell type transition prior

Modeling known biological prior can help to infer the spatiotemporal dynamics of single cells more accurately, especially for situations where there are few observed time points or long time intervals between adjacent time points. For known permissible cell type transitions, we grouped cells with permissible transitions at different time points into the same type. We use $\hat{\rho}_c^{(k)}$ to refer to the empirical distribution of type c cells in the observed data (after rotation and translation) at the time point t_k, and $\bar{\rho}_{t_k,c}$ to refer to the normalized distribution of type c cells evolving from t_0 to t_k. Assuming a total of C types, we realize the distribution matching for each known permissible cell type transitions, i.e.,

$$\|\rho_{t_k,c}\|_1 = n_{k,c}/n_{0,c},\ \bar{\rho}_{t_k,c} = \hat{\rho}_c^{(k)}, \quad k = 1:K,\ c = 1:C, \tag{24}$$

where $n_{0,c}$ and $n_{k,c}$ are the number of type c cells at time t_0 and t_k, respectively. In this circumstance, the feasible state space $\mathcal{S}(X^{(0:K)}, Q^{(0:K)})$ must be modified correspondingly, and the matching loss (23) has to be revised as

$$\mathcal{L}_{\text{Mch}} = \sum_{k=1}^{K}\sum_{c=1}^{C}\left(W_2(\bar{\rho}_{t_k,c}, \hat{\rho}_c^{(k)})\right)^2 + \kappa_{\text{Gro}}\sum_{k=1}^{K}\sum_{c=1}^{C}\frac{|\sum_{j=1}^{n_{0,c}} w_{t_k,j} - n_{k,c}|}{n_{k,c}}, \tag{25}$$

Modelling spatial structure preserving prior

Organ development obeys physical rules, and its spatial structure cannot change at will. For example, some organs remain connected as they develop without breaking into multiple parts. However, when the time interval between observations is long, usual OT-methods often produce results that violate the physical rules in order to minimize the energy, so we need to explicitly model this prior to maintain the spatial structure of the specified organ. For the organs that need to be spatially structure-preserved, we first construct a graph according to spatial coordinates and gene expression. More specifically, at $t = t_0$, we first construct a k_{spa} nearest-neighbor graph using the spatial coordinates, and then find the k_{nbr} cells with the closest gene expression as the final neighbors from the k_{spa} spatial neighbors of each cell in the specified structure to complete the graph construction. We denote the index set of these k_{nbr} neighbor cells, which is a subset of $\{1 : n_0\}$, by $\mathcal{N}(\boldsymbol{x}^{(t_0)}, \boldsymbol{q}^{(t_0)})$ for each cell $(\boldsymbol{x}^{(t_0)}, \boldsymbol{q}^{(t_0)})$ in the specified structure at $t = t_0$. In order to keep the specified organ development obeying the physical rules, we add the optional spatial structure preserving loss function

$$\mathcal{L}_{\text{SSP}}^{(\text{opt})} = \frac{n_{\text{ss}}}{n_0}\mathbb{E}_{(\boldsymbol{x}^{(t_0)},\boldsymbol{q}^{(t_0)})\sim\tilde{\rho}^{(0)}}\frac{1}{k_{\text{nbr}}}\sum_{i\in\mathcal{N}(\boldsymbol{x}^{(t_0)},\boldsymbol{q}^{(t_0)})}\int_{t_0}^{t_K}\left\|\frac{\mathrm{d}}{\mathrm{d}t}|\boldsymbol{x}_i^{(t)} - \boldsymbol{x}^{(t)}|\right\|^2 w_t[\boldsymbol{x},\boldsymbol{q}]\mathrm{d}t, \tag{26}$$

where $\tilde{\rho}^{(0)}$ is the probability distribution for the cells in the specified structure and n_{ss} is the is the number of cells in the specified structure. $\mathcal{L}_{\text{SSP}}^{(\text{opt})}$ preservers spatial structure by promoting as little change as possible in the distance between spatial trajectories of neighboring cells. It can be understood as a continuous time limit of the Gromov-Wasserstein OT distance.

Deep learning-based solver in stVCR

Optimizing the total loss $\mathcal{L}$ in (2) is generally difficult, we use deep learning to find an approximate solution. For the arguments $(\rho_t, \boldsymbol{v}_t, \boldsymbol{p}_t, g_t; R_{1:K}, \boldsymbol{r}_{1:K})$ in the optimization, ρ_t is indeed determined by $(\boldsymbol{v}_t, \boldsymbol{p}_t, g_t)$. We suppose that changes in gene expression, cell migration, and cell proliferation depend on current gene expression and spatial location, and are parameterized by neural networks, i.e., $\boldsymbol{v}_t(\boldsymbol{x},\boldsymbol{q}) = \boldsymbol{v}(\boldsymbol{x},\boldsymbol{q},t;\theta_{\boldsymbol{v}})$, $\boldsymbol{p}_t(\boldsymbol{x},\boldsymbol{q}) = \boldsymbol{p}(\boldsymbol{x},\boldsymbol{q},t;\theta_{\boldsymbol{p}})$ and $g_t(\boldsymbol{x},\boldsymbol{q}) = g(\boldsymbol{x},\boldsymbol{q},t;\theta_g)$, where $\theta := (\theta_{\boldsymbol{v}}, \theta_{\boldsymbol{p}}, \theta_g)$ are the parameters of the neural networks. For the rigid body transformations, the rotation matrix R can be explicitly parameterized. In 2D case, we take

$$R = \begin{pmatrix}\cos\alpha & -\sin\alpha\\ \sin\alpha & \cos\alpha\end{pmatrix},$$

where α is the rotation angle. In 3D case, we parameterize the rotation matrix R by the Euler angles α, β, and γ with $R = R_x(\alpha)R_y(\beta)R_z(\gamma)$, where

$$R_x(\alpha) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos\alpha & -\sin\alpha \\ 0 & \sin\alpha & \cos\alpha \end{pmatrix}, \; R_y(\beta) = \begin{pmatrix} \cos\beta & 0 & \sin\beta \\ 0 & 1 & 0 \\ -\sin\beta & 0 & \cos\beta \end{pmatrix}, \; R_z(\gamma) = \begin{pmatrix} \cos\gamma & -\sin\gamma & 0 \\ \sin\gamma & \cos\gamma & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

Therefore, the overall parameters we need to optimize are the neural network parameters θ, as well as the rotation angles α_k (or Euler angles α_k, β_k and γ_k in 3D case) and translation vectors $\boldsymbol{r}_k$ for $k = 1 : K$.

With the above parameterization, the constraint PDE $\partial_t \rho_t + \nabla \cdot ((\boldsymbol{v}_t, \boldsymbol{p}_t)\rho_t) = g_t \rho_t$ with initial value $\rho_{t=t_0} = \rho^{(0)}$ can be solved by the particle approximations through the ODEs (21) and (22) by replacing the functions $(\boldsymbol{v}_t, \boldsymbol{p}_t, g_t)$ with $(\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{q}, t; \theta_{\boldsymbol{v}}), \boldsymbol{p}(\boldsymbol{x}, \boldsymbol{q}, t; \theta_{\boldsymbol{p}}), g(\boldsymbol{x}, \boldsymbol{q}, t; \theta_g))$. The evaluation of the integral in (20) can be also performed by numerical quadrature in time with the parameterized $(\boldsymbol{v}, \boldsymbol{p}, g)$. Finally, the overall loss

$$\mathcal{L}(\theta, \alpha_{1:K}, \beta_{1:K}, \gamma_{1:K}, \boldsymbol{r}_{1:K}) = \mathcal{L}_{\text{Dyn}} + \lambda_{\text{Mch}} \mathcal{L}_{\text{Mch}} + \lambda_{\text{SSP}} \mathcal{L}_{\text{SSP}}^{(\text{opt})} \tag{27}$$

can be evaluated through the deep learning approximations. The optimization of the parameters are achieved by the Adam optimizer [61]. In (27), $\mathcal{L}_{\text{Dyn}}$ and $\mathcal{L}_{\text{Mch}}$ are required and $\mathcal{L}_{\text{SSP}}^{(\text{opt})}$ is optional. Whether or not to include cell type transition prior in $\mathcal{L}_{\text{Mch}}$ is also optional.

Parameter initialization and training details

The structure of our neural networks $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{q}, t; \theta_{\boldsymbol{v}})$, $\boldsymbol{p}(\boldsymbol{x}, \boldsymbol{q}, t; \theta_{\boldsymbol{p}})$ and $g(\boldsymbol{x}, \boldsymbol{q}, t; \theta_g)$ use multilayer perceptron (MLP) with 128 neurons per layer for a total of 6 layers. To obtain an initial rotation and translation, We downsample 5000 cells in the data, and then use static rigid-body transformation invariant OT on the downsampled data. The training process involves solving the ODEs represented by the neural network, that is neural ODE, which we implement using the torchdiffeq package [48]. It also involves computing the static OT distance, which we implement using the POT package [47]. Finally, in order to enhance the matching with the observed data at each time point and improve the robustness of the algorithm (the new organs or cell types may appear in later time points), we not only compute the loss function defined by (27) by sampling the data from t_0 to later time points, but also compute a similar loss function by sampling the data from other time points t_k, evolving forward and backward. In actual computations, we choose to sample from the first time point t_0 and the last time point t_K to balance the accuracy and computational overhead.

Time-dependent cell type classifier

After we recovered the entire cell dynamics, we could obtain the gene expression and spatial location of cells at unobserved moments. In order to obtain type annotations for these cells at unobserved moments, we train a time-dependent cell type classifier using a neural network in cells that already have cell type annotations, so that we can use it for cell type annotation at unobserved moments. Specifically, we represent a classifier by a neural network $\boldsymbol{f} = \boldsymbol{f}_{\text{type}}(\boldsymbol{x}, \boldsymbol{q}, t; \theta_{\text{type}})$. The inputs are gene expression $\boldsymbol{q}$, spatial coordinates $\boldsymbol{x}$, and time t, and the outputs are probability distributions indicating the probability with which the cell $(\boldsymbol{x}, \boldsymbol{q})$ at time t belongs to each cell type. The loss function is taken as

$$\begin{aligned} L(\theta_{\text{type}}) &= \frac{1}{N} \sum_{i=1}^{N} \left(H(\boldsymbol{y}_i, \boldsymbol{f}_i) + \lambda \left\| \frac{\partial \boldsymbol{f}_i}{\partial t} \right\|_1 \right) \\ &= \frac{1}{N} \sum_{i=1}^{N} \left(\left(-\sum_{j=1}^{K_t} y_{i,j} \log(f_{i,j}) \right) + \lambda \left\| \frac{\partial \boldsymbol{f}_i}{\partial t} \right\|_1 \right), \end{aligned} \tag{28}$$

where $H(\cdot, \cdot)$ is the cross entropy, $\boldsymbol{f}_i := \boldsymbol{f}(\boldsymbol{x}_i, \boldsymbol{q}_i, t)$, $\boldsymbol{y}_i$ is the annotated cell type for cell i, K_t is the total number of cell types, and $y_{i,j}, f_{i,j}$ is the annotated and trained probabilities that the cell i is of type j. The first term in (28) is to force the trained classifier to be consistent with the known cell type annotations, while the second term is a regularization to promote the smoothness of the neural network classifier in time through an ℓ^1 norm of the time derivative on observation points.

Downstream analysis

For notational simplicity, we omit the notation for the parameters of the neural networks in the following description. When we use the stVCR to obtain the spatial velocity $\boldsymbol{v}(\boldsymbol{x}, \boldsymbol{q}_{\text{emb}}, t)$, the gene expression velocity $\boldsymbol{p}_{\text{emb}}(\boldsymbol{x}, \boldsymbol{q}_{\text{emb}}, t)$ in the embedded space, the cell proliferation rate $g(\boldsymbol{x}, \boldsymbol{q}_{\text{emb}}, t)$, and the time-dependent classifier $f_{\text{type}}(\boldsymbol{x}, \boldsymbol{q}_{\text{emb}}, t)$, we can perform a series of downstream analyses, including interpolation, prediction, and study of cell-specific gene-gene, gene-space and space-space interaction and the effects of gene and spatial migration on cell proliferation, etc. Below we give a detailed description of some downstream analysis tasks.

Recovery of cell evolution rates in original space

As stated in the data preprocessing step, the original gene expression and its dimension-reduced expression value in embedded space, $\boldsymbol{q}_{\text{ori}}$ and $\boldsymbol{q}_{\text{emb}}$, are related by the trained encoder-decoder neural network f_{enc}, f_{dec} as $\boldsymbol{q}_{\text{emb}} = f_{\text{enc}}(\boldsymbol{q}_{\text{ori}})$ and $\boldsymbol{q}_{\text{ori}} = f_{\text{dec}}(\boldsymbol{q}_{\text{emb}})$. With this representation, we can easily recover the spatial velocity, gene expression velocity and proliferation rate in original gene expression space:

$$\begin{aligned}\boldsymbol{p}_{\text{ori}}(\boldsymbol{x}, \boldsymbol{q}_{\text{ori}}, t) =& \frac{1}{2\delta_t}\Big(f_{\text{dec}}\big(f_{\text{enc}}(\boldsymbol{q}_{\text{ori}}) + \delta_t \boldsymbol{p}_{\text{emb}}(\boldsymbol{x}, f_{\text{enc}}(\boldsymbol{q}_{\text{ori}}), t)\big) \\ & - f_{\text{dec}}\big(f_{\text{enc}}(\boldsymbol{q}_{\text{ori}}) - \delta_t \boldsymbol{p}_{\text{emb}}(\boldsymbol{x}, f_{\text{enc}}(\boldsymbol{q}_{\text{ori}}), t)\big)\Big), \\ \boldsymbol{v}(\boldsymbol{x}, \boldsymbol{q}_{\text{ori}}, t) =& \boldsymbol{v}(\boldsymbol{x}, f_{\text{enc}}(\boldsymbol{q}_{\text{ori}}), t),\ \ g(\boldsymbol{x}, \boldsymbol{q}_{\text{ori}}, t) = g(\boldsymbol{x}, f_{\text{enc}}(\boldsymbol{q}_{\text{ori}}), t).\end{aligned} \tag{29}$$

Interpolation and prediction

For some interpolation time t_{int} which the user is interested in, we choose the observation time t_{obs} that is closest to t_{int}. Without losing generality, we assume that $t_{\text{obs}} < t_{\text{int}}$. We take the observed data $\boldsymbol{q}_i$ and $\boldsymbol{x}_i$ at t_{obs} as the initial values, and then evolve according to the dynamics learned by stVCR to obtain the interpolation result. More specifically, for the i_{th} cell to evolve from time t to $t + \delta t$, we first compute

$$\boldsymbol{x}_i^{(t+\delta t)} = \boldsymbol{x}_i^{(t)} + \boldsymbol{v}_t(\boldsymbol{x}_i^{(t)}, \boldsymbol{q}_{i,\text{emb}}^{(t)})\delta t,\ \ \boldsymbol{q}_{i,\text{emb}}^{(t+\delta t)} = \boldsymbol{q}_{i,\text{emb}}^{(t)} + \boldsymbol{p}_t(\boldsymbol{x}_i^{(t)}, \boldsymbol{q}_{i,\text{emb}}^{(t)})\delta t, \tag{30}$$

then generate a random number $U \sim \text{Uniform}[0, 1]$: if $g_t(\boldsymbol{x}_i^{(t)}, \boldsymbol{q}_{i,\text{emb}}^{(t)}) > 0$ and $U < g_t(\boldsymbol{x}_i^{(t)}, \boldsymbol{q}_{i,\text{emb}}^{(t)})\delta t$, perform cell division; if $g_t(\boldsymbol{x}_i^{(t)}, \boldsymbol{q}_{i,\text{emb}}^{(t)}) < 0$ and $U < -g_t(\boldsymbol{x}_i^{(t)}, \boldsymbol{q}_{i,\text{emb}}^{(t)})\delta t$, then perform cell apoptosis. The prediction task is completely similar to the interpolation task, which only needs to take the data of the last time point as the initial value.

Cell specific gene-gene, gene-space and space-space interaction

We can study the interaction between genes and space from the learned quantities. For a cell i, as well as the target gene j and the source gene k of interest, we can calculate $\partial p_t^j / \partial q^k|_{(\boldsymbol{x},\boldsymbol{q})=(\boldsymbol{x}_i,\boldsymbol{q}_i)}$, which represents how increased expression of gene k changes the velocity of gene j in cell i. If it is positive, it means that gene k promotes gene j in cell i, otherwise gene k inhibits gene j.

Similarly, for cell i, and axes j and k of interest, we can calculate $\partial v_t^j / \partial x^k|_{(\boldsymbol{x},\boldsymbol{q})=(\boldsymbol{x}_i,\boldsymbol{q}_i)}$ to study space-space interaction. This is used in Spateo [43], although their spatial vector fields $\boldsymbol{v}(\boldsymbol{x})$ are independent of gene expression $\boldsymbol{q}$ and time t.

In stVCR, gene expression $\boldsymbol{q}$ and spatial coordinates $\boldsymbol{x}$ interact, which means we can study how cell migration affects gene expression. For cell i, a target gene j, and a given unit direction $\boldsymbol{n} = (n^1, n^2, n^3)$ (or $\boldsymbol{n} = (n^1, n^2)$ for 2D case), we can calculate the directional derivative

$$\frac{\partial p_t^j}{\partial \boldsymbol{n}} = \frac{\partial p_t^j}{\partial x^1} n^1 + \frac{\partial p_t^j}{\partial x^2} n^2 + \frac{\partial p_t^j}{\partial x^3} n^3$$

at $(\boldsymbol{x}, \boldsymbol{q}) = (\boldsymbol{x}_i, \boldsymbol{q}_i)$, which describes how the migration of cell i to the given direction $\boldsymbol{n}$ affects the expression of gene j, with positive values representing promotion and negative values the opposite. In addition, we can study how gene expression affects cell migration. For cell i and gene j of interest, we can define

$$\boldsymbol{n}_{i,q^j} := \Big(\frac{\partial v_t^1}{\partial q^j}, \frac{\partial v_t^2}{\partial q^j}, \frac{\partial v_t^3}{\partial q^j}\Big),$$

which describes that increased expression of gene j will promote the cell i migration in the direction $\boldsymbol{n}_{i,q^j}/\|\boldsymbol{n}_{i,q^j}\|$ and $\|\boldsymbol{n}_{i,q^j}\|$ indicates the promotion intensity.

Finally, we can define partial derivative of the norm of cell migration velocity with respect to gene $\partial\|\boldsymbol{v}_i\|/\partial q^j$ of cell i and gene j as migration driver gene score based on the above calculations

$$\frac{\partial\|\boldsymbol{v}_i\|}{\partial q^j} = \frac{1}{\|\boldsymbol{v}_i\|}\Big(v_i^1\frac{\partial v_i^1}{\partial q^j} + v_i^2\frac{\partial v_i^2}{\partial q^j} + v_i^3\frac{\partial v_i^3}{\partial q^j}\Big),$$

where $\boldsymbol{v}_i := \boldsymbol{v}_t(\boldsymbol{x_i}, \boldsymbol{q_i})$.

Cell specific effects of gene and spatial migration on growth

For cell i and gene j of interest, we can calculate $\partial g_t/\partial q^j|_{(\boldsymbol{x},\boldsymbol{q})=(\boldsymbol{x}_i,\boldsymbol{q}_i)}$, which describes the effect of gene j on cell proliferation, where a positive value means promoting, and a negative value the opposite. This concept is used in TIGON [34]. However, their growth function $g_t(\boldsymbol{q})$ does not depend on spatial coordinates $\boldsymbol{x}$, so the effect of cell spatial migration on growth cannot be studied.

In stVCR, cell proliferation depends on both its gene expression and its spatial location, so we can also study how cell migration affects its growth. For cell i and a given unit direction $\boldsymbol{n} = (n^1, n^2, n^3)$, we can calculate the directional derivative

$$\frac{\partial g_t}{\partial \boldsymbol{n}} = \frac{\partial g_t}{\partial x^1}n^1 + \frac{\partial g_t}{\partial x^2}n^2 + \frac{\partial g_t}{\partial x^3}n^3,$$

which describes how cell migration in the direction $\boldsymbol{n}$ affects its proliferation.

Finally, we can define partial derivatives of the cell proliferation rate with respect to gene j for cell i, $\partial g_t/\partial q^j|_{(\boldsymbol{x},\boldsymbol{q})=(\boldsymbol{x}_i,\boldsymbol{q}_i)}$, as growth driver gene score based on the above calculations.

Temporal Developmental lineage construction

Since we can interpolate for any time points of interest and can annotate cells at these unobserved time points with the time-dependent classifier $\boldsymbol{f}_{\text{type}}(\boldsymbol{x}, \boldsymbol{q}_{\text{emb}}, t)$, we can construct temporal developmental lineages of cells of interest.

Simulated data setup

In this paper, two simulation data are included. The first corresponding to Fig. 2 and the second corresponding to Fig. 5 and Fig. S11. Below, we will introduce how these three simulation data are generated.

For the first simulation data corresponding to Fig. 2, the dynamics consists of three genes *Red*, *Green* and *Blue* and two spatial coordinates x and y, whose regulation is shown in Fig. 2A. Such a regulatory relationship can be described by a system of stochastic differential equations for gene expression and spatial coordinates

$$\begin{aligned}
\frac{dr}{dt} &= f_1(x)\Big(\frac{r^n}{1+r^n} + \frac{1}{1+g^n+10b^n} - r\Big) + 0.05w_t,\\
\frac{dg}{dt} &= f_2(x)\Big(\frac{g^n}{1+g^n} + \frac{1}{1+r^n+10b^n} - g\Big) + 0.05w_t,\\
\frac{db}{dt} &= \frac{b^2}{1+b^2} - 0.4b + 0.01w_t,\\
\frac{dx}{dt} &= \text{sign}(x)\exp(-4b)\exp(-4g)(r-2)^2r^2 + 0.001w_t,\\
\frac{dy}{dt} &= 0,
\end{aligned} \tag{31}$$

where r, g and b refer to gene *Red*, *Green* and *Blue*, $f_1(x)$ and $f_2(x)$ refer to the factors that depend on the coordinates x, and w_t is a standard Brownian motion. In the computation, we take $n = 4$ to simulate the nonlinear regulation between genes. If we ignore these two x-related factors $f_1(x)$ and $f_2(x)$, r and g are a toggle switch of equal status. We hope $|x| > 1$, $f_1(x) > 1$ and $f_2(x) = 1$, this will promote *Red* expression. Conversely, when $|x| < 1$, $f_2(x) > 1$ and $f_1(x) = 1$, this will promote *Green* expression. The specific forms of $f_1(x)$ and $f_2(x)$ are detailed in the Supplementary Note 2.

cell proliferation modeled as division and apoptosis:

$$\text{proliferation rate} = g_{\text{division}} - g_{\text{apoptosis}} = \frac{2r}{1+r}\frac{|x|}{1+|x|} - \frac{g}{1+g}, \tag{32}$$

which means that the gene *Red* and migrating outward in the horizontal direction will promote cell proliferation while the gene *Green* will inhibit cell proliferation. We sampled three groups of cells at the initial moment and discretized time in order to obtain data through numerical simulations. We simulate gene expression and spatial coordinates according to the forward Euler scheme and simulate cell division and apoptosis by numerically simulating a special Markov process, the birth and death process. Specific details of the sampling of initial values and numerical simulation can be found in the Supplementary Note 2. We evolved the cells at the initial time point from $t = 0$ to $t = 3.0$ according to the given dynamics and took a total of six time points at $t = 0, 0.5, 1.0, 1.5, 2.0$ and 2.5 as observations. Considering that the spatial coordinates obtained at different time points using spatial transcriptome sequencing are not in the same coordinate system, we rotated the spatial coordinates of the second to sixth time points counterclockwise by 8, 16, 24, 32 and 40 degrees, respectively.

For the second simulation data corresponding to Fig. 4 and Fig. S11, similar to the first, the dynamics consists of three genes *Red*, *Green* and *Blue* and two spatial coordinates x and y. There are two types of cells in this simulation data, background cells and migratory cells. The background cells are in steady state, and their spatial coordinates and gene expression do not change with time. The migratory cells transited from high expression of *Red* and *Green* gene to high expression of only *Red* gene while moving to the right. Initial values for background cells and migratory cells can be found in the Supplementary Note 2. Gene expression and spatial coordinates of migrating cells evolve over time and obey stochastic dynamical systems

$$\begin{aligned}
\frac{dr}{dt} &= 0.05w_t, \\
\frac{dg}{dt} &= 1.5\Big(\frac{g^n}{1+g^n} + \frac{1}{1+r^n+10b^n} - g\Big) + 0.05w_t, \\
\frac{db}{dt} &= \frac{b^2}{1+b^2} - 0.4b + 0.01w_t, \\
\frac{dx}{dt} &= 1 + 0.001w_t, \\
\frac{dy}{dt} &= 0,
\end{aligned} \tag{33}$$

where we take $n = 4$ to model non-linear regulatory relationships. Unlike the first simulation data, we do not consider growth in the second simulation data. We evolved the cell at the initial time point from $t = 0$ to $t = 1.0$ according to the given dynamics. In Fig. 4, we took only two time points at $t = 0$ and 1.0 as observations. Additionally we rotated the spatial coordinates of the second time point by 8 degrees counterclockwise. In Fig. S11, we took a total of five time points at $t = 0, 0.25, 0.5, 0.75$ and 1 as observations and rotated the spatial coordinates of the second to fifth time points counterclockwise by 8, 16, 24 and 32 degrees, respectively.

Details of GO enrichment analysis

We used the python package GSEApy=1.0.3 [62] to perform GO enrichment analyses on migration genes and growth genes. In addition, the gene sets used were GO Biological Process 2018 (https://maayanlab.cloud/Enrichr/#libraries).

Data Availability

All the datasets used in this paper are publicly available. The simulation datasets of synthetic circuits are available at https://github.com/QiangweiPeng/stVCR/tree/main/tutorial. The axolotl brain regeneration datasets are freely accessible in CNGB Nucleotide Sequence Archive under accession code CNP0002068. Processed data can be downloaded from https://db.cngb.org/stomics/artista/ [49]. The processed 3D Drosophila embryo datasets can be downloaded from the Spateo package [43] (https://www.dropbox.com/s/bvstb3en5kc6wui/E7-9h_cellbin_tdr_v2.h5ad?dl=1 and https://www.dropbox.com/s/q02sx6acvcqaf35/E9-10h_cellbin_tdr_v2.h5ad?dl=1).

Code Availability

stVCR is implemented in Python and is available at https://github.com/QiangweiPeng/stVCR. The notebooks to reproduce all the results in the manuscript are available at https://github.com/QiangweiPeng/stVCR/tree/main/tutorial.

Supporting Information

Supplementary Notes 1-3

Supplementary Figures S1-S11

Supplementary Videos S1-S15

Acknowledgments

The authors thank the anonymous referees for careful reading and constructive suggestions. TL and QP acknowledge the support from National Key R&D Program of China under grant 2021YFA1003301, and National Science Foundation of China under grant 12288101. PZ acknowledge the support from National Science Foundation of China under grants 12288101, 8206100646, and the Fundamental Research Funds for the Central Universities. We also thank the High-performance Computing Platform of Peking University .

Contribution

All authors conceived the project. QP designed and implemented the algorithm, and performed data analysis. All authors interpreted the results and wrote the manuscript. QP wrote the supplementary materials. PZ and TL supervised the research.

References

[1] Gerri, C., Menchero, S., Mahadevaiah, S. K., Turner, J. M. & Niakan, K. K. Human embryogenesis: a comparative perspective. *Annual Review of Cell and Developmental Biology* **36**, 411–440 (2020).

[2] Briggs, J. A. *et al.* The dynamics of gene expression in vertebrate embryogenesis at single-cell resolution. *Science* **360**, eaar5780 (2018).

[3] Farrell, J. A. *et al.* Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science* **360**, eaar3131 (2018).

[4] Tyser, R. C. *et al.* Single-cell transcriptomic characterization of a gastrulating human embryo. *Nature* **600**, 285–289 (2021).

[5] Lauffenburger, D. A. & Horwitz, A. F. Cell migration: a physically integrated molecular process. *Cell* **84**, 359–369 (1996).

[6] Yamada, K. M. & Sixt, M. Mechanisms of 3d cell migration. *Nature Reviews Molecular Cell Biology* **20**, 738–752 (2019).

[7] Ståhl, P. L. *et al.* Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science* **353**, 78–82 (2016).

[8] Rodriques, S. G. *et al.* Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science* **363**, 1463–1467 (2019).

[9] Stickels, R. R. *et al.* Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqv2. *Nature Biotechnology* **39**, 313–319 (2021).

[10] Chen, A. *et al.* Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays. *Cell* **185**, 1777–1792 (2022).

[11] Oliveira, M. F. *et al.* Characterization of immune cell populations in the tumor microenvironment of colorectal cancer using high definition spatial profiling. *BioRxiv* 2024–06 (2024).

[12] La Manno, G. *et al.* RNA velocity of single cells. *Nature* **560**, 494–498 (2018).

[13] Bergen, V., Lange, M., Peidli, S., Wolf, F. A. & Theis, F. J. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nature Biotechnology* **38**, 1408–1414 (2020).

[14] Gao, M., Qiao, C. & Huang, Y. Unitvelo: temporally unified RNA velocity reinforces single-cell trajectory inference. *Nature Communications* **13**, 6586 (2022).

[15] Cui, H. *et al.* Deepvelo: deep learning extends RNA velocity to multi-lineage systems with cell-specific kinetics. *Genome Biology* **25**, 27 (2024).

[16] Li, S. *et al.* A relay velocity model infers cell-dependent RNA velocity. *Nature Biotechnology* **42**, 99–108 (2024).

[17] Gayoso, A. *et al.* Deep generative modeling of transcriptional dynamics for RNA velocity analysis in single cells. *Nature Methods* **21**, 50–59 (2024).

[18] Farrell, S., Mani, M. & Goyal, S. Inferring single-cell transcriptomic dynamics with structured latent gene expression dynamics. *Cell Reports Methods* **3** (2023).

[19] Li, T., Shi, J., Wu, Y. & Zhou, P. On the mathematics of RNA velocity I: theoretical analysis. *BioRxiv* 2020–09 (2020).

[20] Battich, N. *et al.* Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science* **367**, 1151–1156 (2020).

[21] Erhard, F. *et al.* scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* **571**, 419–423 (2019).

[22] Qiu, Q. *et al.* Massively parallel and time-resolved RNA sequencing in single cells with scNT-seq. *Nature Methods* **17**, 991–1001 (2020).

[23] Cao, J., Zhou, W., Steemers, F., Trapnell, C. & Shendure, J. Sci-fate characterizes the dynamics of gene expression in single cells. *Nature Biotechnology* **38**, 980–988 (2020).

[24] Hendriks, G.-J. *et al.* NASC-seq monitors RNA synthesis in single cells. *Nature Communications* **10**, 3138 (2019).

[25] Xu, Z., Sziraki, A., Lee, J., Zhou, W. & Cao, J. Dissecting key regulators of transcriptome kinetics through scalable single-cell RNA profiling of pooled CRISPR screens. *Nature Biotechnology* 1–6 (2023).

[26] Holler, K. *et al.* Spatio-temporal mRNA tracking in the early zebrafish embryo. *Nature Communications* **12**, 3358 (2021).

[27] Lin, S. *et al.* Well-temp-seq as a microwell-based strategy for massively parallel profiling of single-cell temporal RNA dynamics. *Nature Communications* **14**, 1272 (2023).

[28] Qiu, X. *et al.* Mapping transcriptomic vector fields of single cells. *Cell* **185**, 690–711 (2022).

[29] Peng, Q., Qiu, X. & Li, T. Storm: Incorporating transient dynamics to infer the RNA velocity with metabolic labeling information. *BioRxiv* 2023–06 (2023).

[30] Schiebinger, G. *et al.* Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell* **176**, 928–943 (2019).

[31] Tong, A., Huang, J., Wolf, G., Van Dijk, D. & Krishnaswamy, S. Daumé III, H. & Singh, A. (eds) *Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics.* (eds Daumé III, H. & Singh, A.) *International Conference on Machine Learning*, Vol. 119, 9526–9536 (PMLR, 2020).

[32] Benamou, J.-D. & Brenier, Y. A computational fluid mechanics solution to the Monge-Kantorovich mass transfer problem. *Numerische Mathematik* **84**, 375–393 (2000).

[33] Huguet, G. *et al.* Manifold interpolating optimal-transport flows for trajectory inference. *Advances in Neural Information Processing Systems* **35**, 29705–29718 (2022).

[34] Sha, Y., Qiu, Y., Zhou, P. & Nie, Q. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence* **6**, 25–39 (2024).

[35] Chizat, L., Peyré, G., Schmitzer, B. & Vialard, F.-X. An interpolating distance between optimal transport and Fisher–Rao metrics. *Foundations of Computational Mathematics* **18**, 1–44 (2018).

[36] Zhang, K., Zhu, J., Kong, D. & Zhang, Z. Modeling single cell trajectory using forward-backward stochastic differential equations. *PLOS Computational Biology* **20**, e1012015 (2024).

[37] Jiang, Q. & Wan, L. A physics-informed neural SDE network for learning cellular dynamics from time-series scRNA-seq data. *Bioinformatics* **40**, ii120–ii127 (2024).

[38] Zhang, Z., Li, T. & Zhou, P. Learning stochastic dynamics from snapshots through regularized unbalanced optimal transport. *arXiv preprint arXiv:2410.00844* (2024).

[39] Zeira, R., Land, M., Strzalkowski, A. & Raphael, B. J. Alignment and integration of spatial transcriptomics data. *Nature Methods* **19**, 567–575 (2022).

[40] Mémoli, F. Gromov–Wasserstein distances and the metric approach to object matching. *Foundations of Computational Mathematics* **11**, 417–487 (2011).

[41] Klein, D. *et al.* Mapping cells through time and space with moscot. *BioRxiv* 2023–05 (2023).

[42] Scetbon, M. & Cuturi, M. Low-rank optimal transport: Approximation, statistics and debiasing. *Advances in Neural Information Processing Systems* **35**, 6802–6814 (2022).

[43] Qiu, X. *et al.* Spateo: multidimensional spatiotemporal modeling of single-cell spatial transcriptomics. *BioRxiv* 2022–12 (2022).

[44] Halmos, P. *et al.* Ma, J. (ed.) *Dest-ot: Alignment of spatiotemporal transcriptomics data.* (ed.Ma, J.) *International Conference on Research in Computational Molecular Biology*, 434–437 (Springer, 2024).

[45] Liu, J., Gu, Y., Li, C. & Welch, J. D. Mapping cell fate transition in space and time. *BioRxiv* 2024–02 (2024).

[46] Zhou, P., Bocci, F., Li, T. & Nie, Q. Spatial transition tensor of single cells. *Nature Methods* 1–10 (2024).

[47] Flamary, R. *et al.* POT: Python optimal transport. *Journal of Machine Learning Research* **22**, 1–8 (2021). URL http://jmlr.org/papers/v22/20-451.html.

[48] Chen, R. T., Rubanova, Y., Bettencourt, J. & Duvenaud, D. K. Neural ordinary differential equations. *Advances in Neural Information Processing Systems* **31** (2018).

[49] Wei, X. *et al.* Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science* **377**, eabp9444 (2022).

[50] Tong, A. *et al.* Conditional flow matching: Simulation-free dynamic optimal transport. *arXiv preprint arXiv:2302.00482* **2** (2023).

[51] Jin, S. *et al.* Inference and analysis of cell-cell communication using CellChat. *Nature Communications* **12**, 1088 (2021).

[52] Cang, Z. *et al.* Screening cell–cell communication in spatial transcriptomics via collective optimal transport. *Nature Methods* **20**, 218–228 (2023).

[53] Wang, K. *et al.* PhyloVelo enhances transcriptomic velocity field mapping using monotonically expressed genes. *Nature Biotechnology* **42**, 778–789 (2024).

[54] Li, C., Virgilio, M. C., Collins, K. L. & Welch, J. D. Multi-omic single-cell velocity models epigenome–transcriptome interactions and improves cell fate prediction. *Nature Biotechnology* **41**, 387–398 (2023).

[55] Peyré, G. & Cuturi, M. *Computational Optimal Transport: With Applications to Data Science* Vol. 11 of *Foundations and Trends in Machine Learning* (now Publishers Inc., 2019).

[56] Liero, M., Mielke, A. & Savaré, G. Optimal transport in competition with reaction: The Hellinger–Kantorovich distance and geodesic curves. *SIAM Journal on Mathematical Analysis* **48**, 2869–2911 (2016).

[57] Kondratyev, S., Monsaingeon, L. & Vorotnikov, D. A new optimal transport distance on the space of finite radon measures. *Advances in Differential Equations* **21**, 1117 – 1164 (2016).

[58] Cohen, S. & Guibasm, L. Werner, B. (ed.) *The earth mover's distance under transformation sets*. (ed.Werner, B.) *Proceedings of the Seventh IEEE International Conference on Computer Vision*, Vol. 2, 1076–1083 (IEEE, 1999).

[59] Schönemann, P. H. A generalized solution of the orthogonal procrustes problem. *Psychometrika* **31**, 1–10 (1966).

[60] Shi, J., Aihara, K., Li, T. & Chen, L. Energy landscape decomposition for cell differentiation with proliferation effect. *National Science Review* **9**, nwac116 (2022).

[61] Bishop, C. M. & Bishop, H. *Deep Learning: Foundations and Concepts* (Springer Nature Switzerland, Switzerland, 2024).

[62] Fang, Z., Liu, X. & Peltz, G. GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics* **39**, btac757 (2023).