stVCR: Reconstructing spatio-temporal dynamics of cell development using optimal transport

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

*** 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, growth and migration in physical space over time. Due to the destructive nature of sequencing, simultaneous reconstruction of cell differentiation, growth and migration trajectories remains challenging, especially migration in physical space, as the coordinates obtained at different temporal snapshots are not in the same coordinate system. To realize this potential, we developed stVCR, which is an optimal transport algorithm with dynamical form, unbalanced setting and invariance to rigid body transformations. stVCR extends the previous algorithm, which only reconstructs differentiation trajectories and population growth, to end-to-end simultaneously reconstruct cell differentiation, growth, migration in physical space and align spatial coordinates of multiple snapshots. In addition, stVCR allows the study of the interaction between gene expression and spatial migration and the effect of gene expression and spatial migration on growth. We verified the effectiveness of stVCR on simulated data and axolotl brain regeneration data.

Introduction

The development of a fertilized egg into a complete embryo is an extremely complex and fascinating process in biology. This process involves intricate interactions between the dynamic regulation of gene expression, cell differentiation, cell division, apoptosis, and cell migration within physical space. During early development, the fertilized egg undergoes a series of changes in gene expression patterns that vary over time and across different regions, resulting in diverse cells expressing distinct genes at specific locations and times. This dynamic regulation of gene expression and spatial organization determines the direction and fate of cell differentiation. Cell division and apoptosis are also critical in embryonic development. Through continuous division and apoptosis, cells form multicellular structures and gradually differentiate into various cell types to create different tissues and organs. Additionally, this process involves the organization of cells into different structures through migration and changes in their physical locations.

The advent of spatial transcriptome (ST) technology has made it possible to obtain both gene expression data and spatial coordinates [5, 29, 33, 34]. As technology advances, spatial resolution has reached the single-cell or even subcellular level, exemplified by methods such as Stereo-seq [5] and 10x Visium HD. Due to the destructive nature of sequencing, ST data can only provide snapshot data 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. Even sequencing at multiple time points during embryonic development to obtain ST data only provides multiple unpaired snapshot data.

Recovering cell developmental trajectories from single-cell data or ST data is a complex and challenging task. RNA velocity [19] utilizes unspliced/spliced RNA to infer the developmental direction of each cell. This inspired a series of subsequent work using unspliced/spliced RNA to more accurately infer RNA velocity [3, 9, 11, 13, 14, 20]. These methods suffer from scale invariance due to the lack of temporal information [21]. Metabolic labeling scRNA-seq introduce temporal information into the data by distinguishing new/old RNA [1, 4, 10, 15, 16, 22, 26, 38]. Dynamo [27] designed parameter inference methods for metabolic labeling scRNA-seq data based on steady-state assumptions and deterministic models, and Storm [25] 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 [31] 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. But it only approximates cell growth by using growth hallmark gene, which largely depends on the choice of database. TrajectoryNet [35] combines dynamical OT [2] and normalizing flows to infer continuous trajectories of cell development and takes cell growth into account through a separated static OT. MIOFlow [17] follows the geometry by operating in the latent space of an geodesic autoencoder. FBSDE [40] integrates the forward and backward movements of two SDEs in time to capture the cell developmental dynamics. TIGON [32] uses dynamical unbalanced OT [7] to reconstructs dynamic trajectories and population growth simultaneously. Since the usual scRNA-seq data do not include the spatial coordinates of the cells, none of the above methods can model cell migration in physical space. 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 to align 2D adjacent tissue ST slices to reconstruct the 3D structure of the tissue. Moscot [18] 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 [30] to accommodate larger data sizes. Spateo [28] 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 overlooks the interplay between gene expression and spatial location, and processes such as cell division and apoptosis. TopoVelo [23] 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.

Reconstructing dynamic trajectories of cell differentiation, growth, and migration in physical space simultaneously for time-series ST data without relying on a priori biological knowledge is challenging, especially migration in physical space, as the coordinates obtained at different temporal snapshots are not in the same coordinate system. Analogous to recovering a video from multiple photos, we aim to reconstruct the entire cellular developmental dynamics from multiple unpaired ST snapshot data, thus obtaining a continuous developmental trajectory. To achieve this goal, we developed an algorithm called spatio-temporal Video Cassette Recorder (stVCR), which is an optimal transport algorithm with dynamical form, unbalanced setting and invariance to rigid body transformations. stVCR can end-to-end simultaneously reconstruct dynamic trajectories of cell differentiation, growth, migration and align spatial coordinates of multiple ST snapshots. In addition, stVCR can reveal the complex regulatory mechanisms behind the overall cellular dynamics, including how gene expression and spatial location affect each other, and how gene expression and

spatial location affect growth.

Results

Overview of stVCR

In stVCR, we adopt different modeling strategies for different types of data in the spatial transcriptome (Fig. 1A). For gene expression data, since gene expression data obtained at different time points are in the same space, we use the usual Wasserstein OT to model them (Fig. 1A Left and Methods). For spatial coordinate data, since possible rotations and translations make the spatial coordinates obtained at different time points in different coordinate systems, we use the rigid-body transformation invariant OT to model them (Fig. 1A Middle and Methods). In addition, since cell division and apoptosis make the number of cells vary at different time points, we use the unbalanced OT to model this (Fig. 1A Right and Methods). We unified all three 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 and Methods). We take the spatial coordinates of the cells at the first time point as the reference coordinates system, and a group of cells is described as a Lagrangian view of the time-dependent probability density $\rho_t(x, z)$. Here $x \in \mathbb{R}^d$ is the gene expression data after dimensionality reduction. $z \in \mathbb{R}^{d_s}$ ($d_s = 2$ or 3) is the spatial coordinate in the reference coordinates system. stVCR finds the optimal rigid rotation matrix R_i and translation r_i of the coordinate system of the other time points that differ from the first time point, and uses the partial differential equation

$$\partial_t \rho_t(x, z) + \nabla \cdot ([v_{x,t}(x, z), v_{z,t}(x, z)]\rho_t(x, z)) = g_t(x, z)\rho_t(x, z) \tag{1}$$

to interpolate the probability density $\rho_i = \rho(R_i, r_i)$ of the transformed time-series spatial transcriptome data at the same time (Fig. 1B,C), where $v_{x,t}(x, z)$, $v_{z,t}(x, z)$, and $g_t(x, z)$ are parameterized using neural networks. $v_{x,t}(x, z) \in \mathbb{R}^d$ describes gene expression changes in cells, $v_{z,t}(x, z) \in \mathbb{R}^{d_s}$ describes the migration of cells in physical space, and $g_t(x, z) \in \mathbb{R}$ describes cell growth (Fig. 1C).

stVCR parameterizes the rotation matrix in 2D rigid body by using rotation angle (or Euler angle for 3D case), and simultaneously finds the optimal rotation angle and translation and parameterized dynamics by minimizing Wasserstein–Fisher–Rao (WFR) loss

$$\begin{aligned}\mathcal{L}_{WFR} = \int_{t_0}^{t_K} \int_{\mathbb{R}^{d+d_s}} \Big(\|v_{x,t}(x, z)\|^2 + \tau_1 \|v_{z,t}(x, z)\|^2 \\ + \tau_2 \|g_t(x, z)\|^2 \Big) \rho_t(x, z) \mathrm{d}(x, z)\ \mathrm{d}t\end{aligned} \tag{2}$$

and matching error loss

$$\mathcal{L}_{Match} = \Sigma_{i=1}^{K} d_{match}(\rho_{t=t_i}, \rho_i(R_i, r_i)). \tag{3}$$

where WFR loss promotes the evolution of cell dynamics in a way that consumes the least energy, and the matching error loss promotes the cell dynamics to match the observed data as much as possible at the observed time point (Methods). The optimization process involves integrating the ODEs represented by neural networks, which we solve using the torchdiffeq package [6].

Once we have completed the entire training process to obtain the optimal rigid-body transformation and parameterized dynamics, we can first use the optimal rigid-body transformation to align the spatial coordinates of different time points to the same coordinate system, and then we can perform a series of downstream analyses (Fig. 1D

and Methods). (1) Interpolation and prediction. For the interpolation problem, we evolve forward or backward from the nearest observations to the interested time point based on learned cellular dynamics. For the prediction problem, similarly evolve forward from the last time point. In addition, for already annotated data, we can infer developmental lineages by learning a time-dependent classifier to annotate unobserved interpolations or predictions (Methods). (2) Studying cell-specific gene-gene, gene-space, and space-space interactions. For clarity, we use superscripts to represent the components of a vector. To study gene-gene interactions, we calculate $\frac{\partial v_x^j}{\partial x^k}$, which describes the regulatory strength of the source gene k on the target gene j. Similarly we can compute $\frac{\partial v_x^j}{\partial z^k}$, $\frac{\partial v_z^j}{\partial x^k}$ and $\frac{\partial v_z^j}{\partial z^k}$, which describe the effect of cell movement in the z^k direction on the expression of gene j, the effect of gene k expression on the movement of the cell in the z^j direction, and the effect of cell movement in the z^k direction on the movement along the z^j direction, respectively. In addition to the directions represented by the axes, for any direction represented by a unit vector $\vec{n} = (n_1, n_2)$, we can compute the directional derivative $\frac{\partial v_x^j}{\partial \vec{n}}$, which describes the effect of moving along given direction on expression of gene j. (3) Studying the cell-specific effects of genes and spatial location on growth. The effect of each gene on cell growth can be obtained by calculating the partial derivative $\frac{\partial g}{\partial x^j}$. The effect of cell movement along any given direction on cell growth can be obtained by calculating the directional derivative $\frac{\partial g}{\partial \vec{n}}$. We can accordingly obtain the key genes and key moving directions that promote cell division or apoptosis.

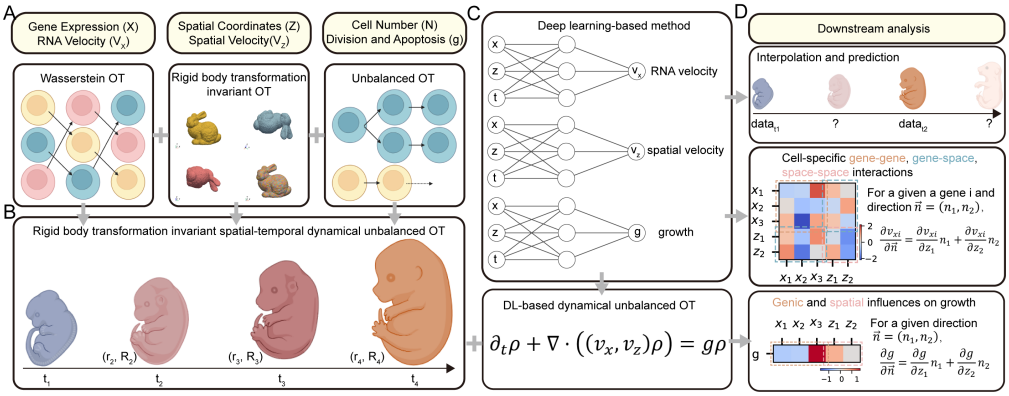

Figure 1. Overview of stVCR. A. stVCR adopts different modeling strategies for different types of data in the spatial transcriptome, more specifically using Wasserstein OT for gene expression data (Left), rigid body transformation invariant OT for spatial coordinates (Middle), and unbalanced OT for changes in cell numbers due to cell division and apoptosis (Right). **B.** stVCR unifies the three modules into the dynamical OT. The input time-series spatial transcriptome snapshots are described as the probability density $\rho(r_i, R_i)$ of spatial coordinates after a rigid body transformation. **C.** The density $\rho_t(x, z)$ is governed by a partial differential equation involving RNA velocity $v_{x,t}(x, z)$, spatial velocity $v_{z,t}(x, z)$ and growth $g_t(x, z)$ that are modelled by three neural networks. **D.** stVCR can perform downstream analyses, including interpolation and prediction (Top), studying cell-specific gene-gene, gene-space, space-space interactions (Middle), and studying the cell-specific effects of gene expression and spatial movement on growth (Down).

Benchmark on the simulated time-series ST data

We first tested the performance of stVCR on the simulated dataset of three genes and two spatial dimensions (Fig. 2 and Fig. S1; Methods section). There are regulatory relationships between different genes and different spatial coordinates, in addition, gene expression and cell migration also affect cell growth (Fig. 2A). The three genes are named *Red*, *Green* and *Blue* genes. *Blue* is the upstream gene, and high expression of *Blue* will inhibit *Red* and *Green*. *Red* and *Green* are a toggle-switch, which self-activate and inhibit each other. In addition, *Red* promotes cell growth and *Green* inhibits cell growth. The horizontal direction in spatial coordinates (i.e., the x-axis) is the key to the simulated data. Cells moving outward horizontally (i.e., $|x|$ is larger) will promote *Red* expression and cell growth, while the opposite will promote *Green* expression. In summary, the difference in spatial location makes the two genes in the toggle-switch unequal in status. The cells at initial time point included three groups of cells (Fig. 2 B). The gene expression of the second and third groups at initial time point is similar, all genes are underexpressed, but the spatial location of the second group is on the outside and the third group is on the inside (Fig. 2B Left). While the second group of cells migrated outwards in the horizontal direction, they continuously divided and the *Red* gene was gradually highly expressed. The third group of cells showed gradually high expression of *Green* gene and continued apoptosis. Without spatial information, it is not possible to distinguish between the second and third group of cells at the initial moment, which would lead to erroneous trajectories (Fig. 2B Left). The first group of cells had high *Blue* expression and were in steady state, which remained virtually unchanged over time. Notice that the second group of cells is constantly dividing and increasing in number, while the third group of cells is constantly apoptotic and decreasing in number. If balanced OT is used for modeling and cell growth is ignored it will lead to wrong cell trajectories of the small number group to the large number group (Fig. 2B Right). We took as input data at t=0.0, 0.5, 1.0, 1.5, 2.0 and 2.5 for a total of 6 time points. In addition, we rotated the spatial coordinates of the second to sixth time points by 8, 16, 24, 32, and 40 degrees counterclockwise, respectively, to simulate the possible rotation of tissues by spatial transcriptome sequencing (Fig. 2C and Fig. S1A).

We tested the performance of stVCR on this simulated data. First, we take the data from the first time point and evolve them according to the learned dynamics. The results show that the learned dynamics are close to the real dynamics (Supplementary Video 1; Fig. 2D and Fig. S1B). 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 the 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). We interpolate the unobserved moments t=0.25, 0.75, 1.25, 1.75 and 2.25 and the results are close to ground truth (Fig. 2E and Fig. S1C). We take the data at the last time point and make the prediction for t=2.75 based on the learned dynamics, which is also close to the ground truth (Fig. 2F). We investigated the ability of stVCR to restore the effects of gene interactions and cell migration on gene expression. First we calculate the partial derivatives of *Green* with respect to *Red* in the learned dynamics and true dynamics and visualize them in coordinate space (Fig. 2G Left). Qualitatively, they were close, and *Green* inhibited *Red* expression mainly in the second and third group of cells. Next, we calculated the directional derivative of *Red* gene expression for the given direction $n = (1, 0)$ (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. Finally, we investigated the spatial variability of cell growth and the effect of cell migration on growth. We compared true and learned cell growth rates and visualized them in coordinate space (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 $n = (1, 0)$, similarly this shows that cells moving outward in the horizontal direction will promote cell growth (Fig. 2H Right).

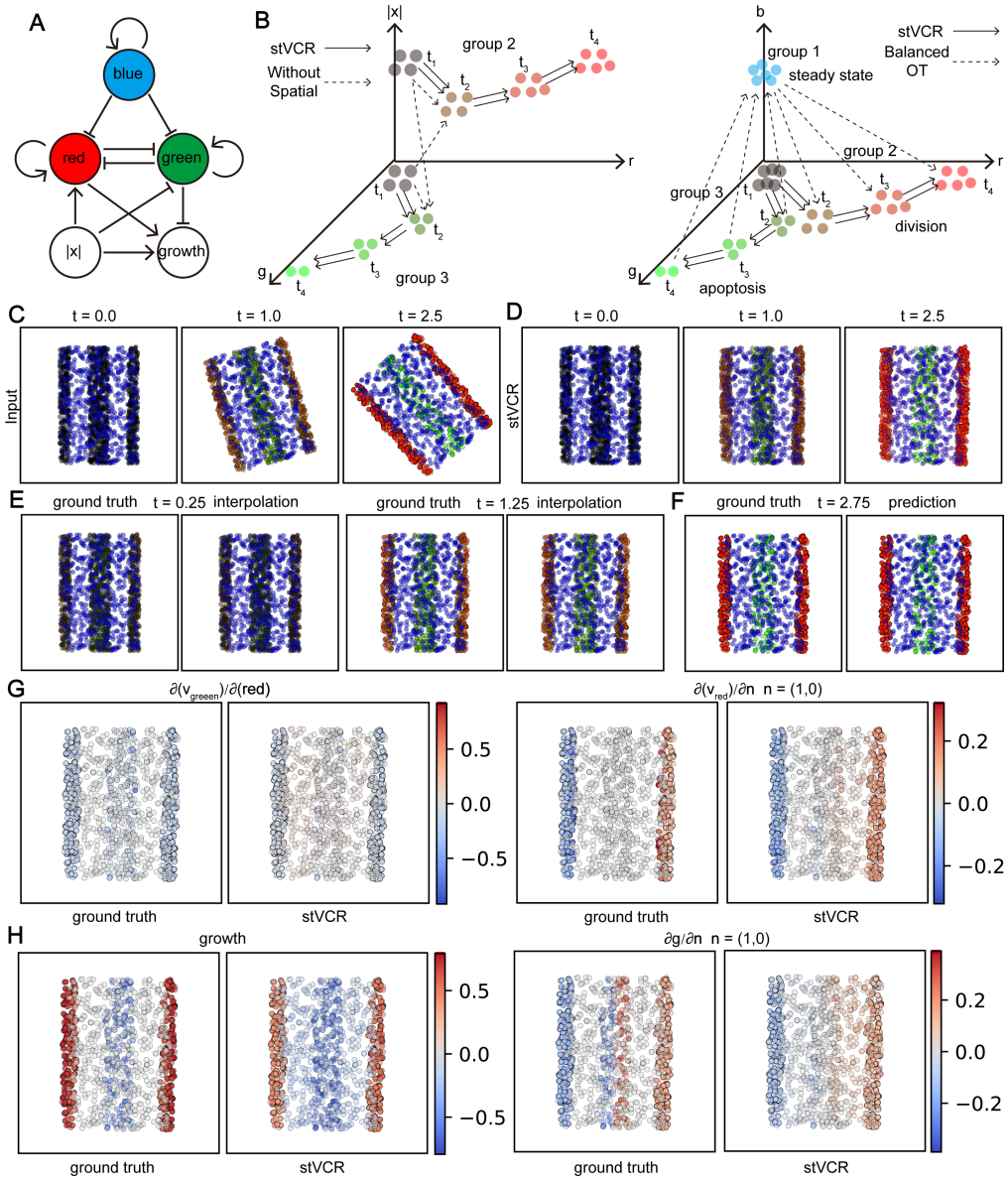

Figure 2. Benchmark on the simulated time-series ST data. A. Regulation relationship diagram of simulated data. **B.** Dynamic diagram of cell evolution over time. Left: Dynamics of the second group and third group of cells over time in r, g and $|x|$ dimension. Right: Dynamics of all cells over time in r, g and b dimension. **C.** Input data 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 evolution 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* with respect to *Red* 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 n, where $n = (1, 0)$. **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 growth with respect to given direction n, where $n = (1, 0)$.

stVCR reconstructs dynamic trajectory of axolotl brain regeneration

We applied stVCR for the axolotl brain regeneration dataset using single-cell Stereo-seq technology [37]. This datasets include sections 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, the regeneration process is mainly concentrated in 2 DPI to 20 DPI, so we took the data of 5 moments at 2, 5, 10, 15 and 20 and reconstructed the whole dynamic regeneration process using stVCR (Fig. 3).

stVCR reconstructs the dynamic process and aligns the spatial coordinates of the data collected at different time points to the same coordinate system (Fig. 3A Left). Fig. 3A and Fig. S2A show the aligned coordinate at different time points. We trained a classifier based on existing cell annotations using neural network, which allowed us to annotate cells at unobserved time points (see the Methods section). 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 2). We show the interpolation results at 3.5, 7.5, 12.5 and 17.5 DPI (Fig. 3B and Fig. S2B). In addition, 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 reaEGC and MCG cells (Fig. 3C). Next, we studied the spatial distribution of cell growth rates (Fig. 3D and Fig. S2C), which showed that cell growth rates in the injured hemispheres were significantly higher than those in the uninjured hemispheres, implying that cell division was more active in the injured hemispheres. We think this phenomenon may be due to the need to compensate for cells lost due to injury. We also investigated cell-specific gene regulation of growth and examined spatial variation in regulatory relationships. *Sfrp1* is a marker gene for sfrpEGC cells, and we observed that it positively regulates cell growth (Fig. 3E Left), i.e., that higher expression will lead to more active cell division. We note that this regulatory relationship is concentrated in sfrpEGC cells at the wound site, whereas it is not evident in sfrpEGC cells in the uninjured hemisphere (Fig. 3E Left). Correspondingly the neuron growth contributor *Atf3* showed a generalized promotion of cell growth (Fig. 3E Right). Finally, we studied the change in cell type of each cell over time. We took sfrpEGC cells and reaEGC cells at 2 DPI and evolved them to 15 DPI using learned dynamics, then annotated them using learned classifiers, and finally constructed developmental lineages based on the results (Fig. 3F; Methods). We followed the original study focusing on three EGC cells wntEGC, sfrpEGC and reaEGC cells. The results show that sfrpEGC cells can be transformed into wntEGC and reaEGC cells, and that wntEGC and reaEGC cells can transform into each other (Fig. 3G). We speculate the possible reason for this is that the degree of wound healing changes over time. Initially, when the wound is more severe, the sfrpEGC cells are directly transformed into reaEGC cells or into wntEGC cells and then into reaEGC cells. Later, as the wound heals, the reaEGC cells transform back into wntEGC cells or differentiate into other cells.

In summary, stVCR describes the dynamics of such axolotl brain regeneration. At the beginning of injury, sfrpEGC and wntEGC cells transform into reaEGC cells, and the proliferation of cells in the injured hemisphere become active, especially EGC cells in the vicinity of the wound. reaEGC cells move toward the wound. As the wound gradually healed, reaEGC cells transformed back to wntEGC cells or differentiated into other neurons.

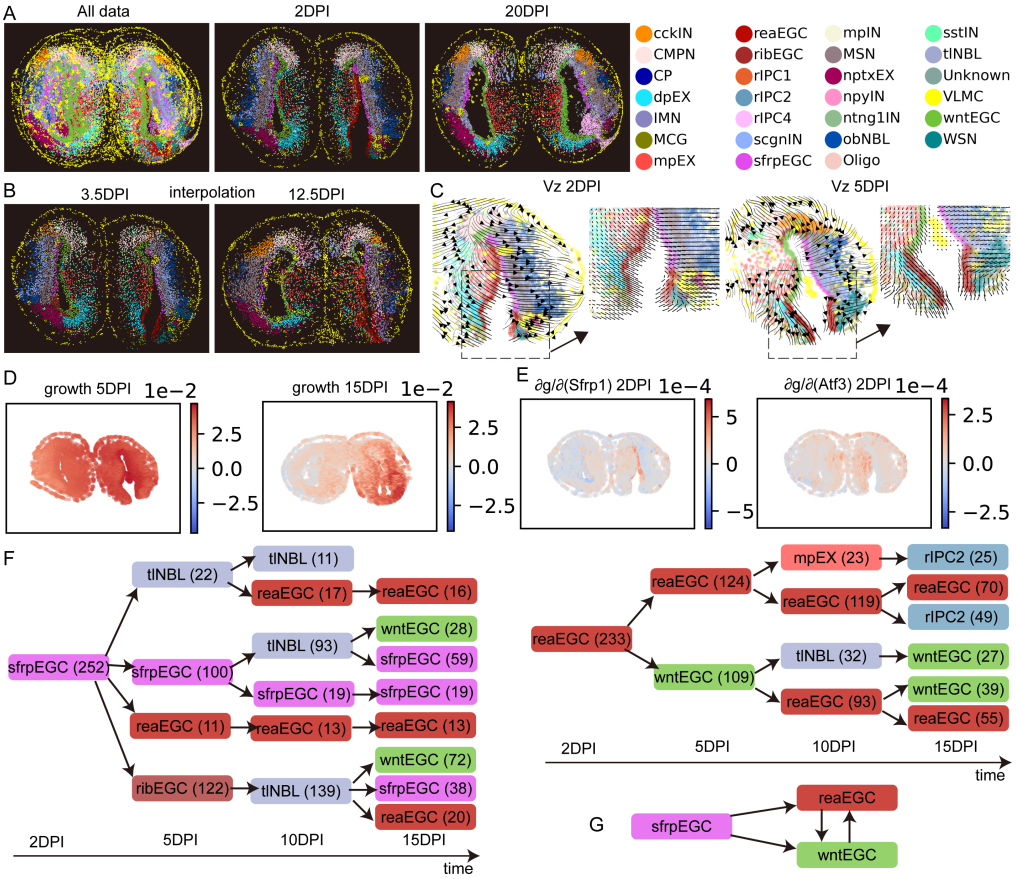

Figure 3. stVCR analysis of axolotl brain regeneration. A. stVCR aligns the spatial coordinates of data at different time points to the same coordinate system. Left: All data; Middle: 2DPI; Right: 20DPI. Cell type annotations come from the original study. **B.** stVCR interpolation at 3.5DPI and 12.5DPI. Cell type annotations come from the stVCR's time-dependent classifier. **C.** stVCR spatial velocity at 2DPI and 5DPI injured hemispheres. Left: Streamline plot; Right: Locally amplified grid velocity. **D.** stVCR cell growth rate at 5DPI and 15DPI data. **E.** stVCR restores regulation of cell growth by gene expression. Left: Partial derivative of cell growth g with respect to gene *Sfrp1* on 2DPI data; Right: Partial derivative of cell growth g with respect to gene *Atf3* on 2DPI data. **F.** stVCR reconstructs the time-varying developmental lineages of 2DPI sfrpEGC cells (Left) and reaEGC cells (Right). The number in parentheses is the number of cells. **G.** Transformational relationships between three EGC-type cells sfrpEGC, reaEGC, and wntEGC.

Discussion

stVCR models time series spatial transcriptome data based on OT and its variants, especially the rigid body transformation invariant OT, making it possible to simultaneously study changes in cell gene expression, spatial migration, and cell division and apoptosis. One possible limitation is that stVCR need to reduce gene expression to low dimensions in advance and then learn the dynamics in the low dimensional space. Learning both the low-dimensional representation of gene expression and the dynamics in the low-dimensional simultaneously may yield better results [11]. Another possible limitation is that 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 [36], which may also be generalized to spatial transcriptome data. Finally, current OT dynamics completely ignore the intrinsic dynamics of gene expression (such as transcription, splicing, and degradation), and combining OT dynamics with intrinsic dynamics may yield more accurate results.

Conclusion

We present stVCR for recovering the entire spatio-temporal dynamics of cell development from time series spatial transcriptome data, including changes in cell gene expression, spatial migration, division, and apoptosis. At the same time, stVCR also aligns the spatial coordinates at different time points so that they are in the same coordinate system. stVCR mainly utilizes the OT and its variants and generalizes the variants to the dynamical form, especially the rigid body transformation invariant OT for sptail coordinate. After learning the entire spatio-temporal dynamics, stVCR can perform a series of downstream analyses, including interpolation and prediction, studying gene-gene, gene-space and spaca-space interactions, and studying the effects of gene and spatial migration on growth. Due to the limitation of 3D time series spatial transcriptome data, further applications and performance evaluations for more challenging datasets are desired and it will be studied in the future.

Methods

Data preprocessing

In order to utilize stVCR more efficiently, we first normalized the raw counts data using size factor. And 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 $x_{emd} = f_{\theta_{enc}}(x)$ and a decoder $\tilde{x} = f_{\theta_{dec}}(x_{emd})$ using neural networks, where the input of the encoder is the gene expression x and the output is the low-dimensional embedding x_{emd}, and the decoder is the opposite. The loss function uses the mean square error, that is,

$$Loss = \frac{1}{N}\Sigma_{i=1}^{N}\|x - \tilde{x}\|_2^2, \tag{4}$$

where N refers to the total number of cells. In the actual calculation, we take the dimension of the low-dimensional embedding x_{emd} to be 10. Also for simplicity of notation, we still use x to refer to x_{emd} to denote the low-dimensional embedding of gene expression unless otherwise stated.

Optimal transport, dynamical optimal transport and dynamical unbalanced optimal transport

The Monge problem aims to find a mapping that assigns each point x_i to exactly one point y_j, directing the mass of measure α towards that of measure β, where $\alpha = \sum_{i=1}^{n} \mathbf{a}_i \delta x_i$ and $\beta = \sum_{j=1}^{m} \mathbf{b}_j \delta y_j$. The assignment problem is combinatorial, and the feasible set for the Monge problem is nonconvex. Kantorovich suggests an alternative approach where the mass located at any point x_i could potentially be dispatched across multiple locations. Kantorovich's optimal transport problem reads

$$\mathcal{L}_c(\alpha, \beta) = \min_{\mathbf{P} \in \mathbf{U}(\mathbf{a}, \mathbf{b})} \langle \mathbf{C}, \mathbf{P} \rangle \overset{\text{def.}}{=} \sum_{i,j} \mathbf{C}_{i,j} \mathbf{P}_{i,j}, \tag{5}$$

where $\mathbf{U}(\mathbf{a},\mathbf{b}) \overset{\text{def.}}{=} \left\{\mathbf{P} \in \mathbb{R}_+^{n\times m} : \mathbf{P}\mathbb{1}_m = \mathbf{a} \quad \text{and} \quad \mathbf{P}^{\mathrm{T}}\mathbb{1}_n = \mathbf{b}\right\}$.

When $c(x,y) = \|x-y\|^2$, the optimal transport distance $\mathcal{L}_c(\alpha_0,\alpha_1)$ as defined in Eq. (5) can be computed by finding a shortest path $(\alpha_t(x))_{t=0}^1$ between these two measures. This path is described by advecting the measure using a vector field $v_t(x)$ defined at each time point. The path $\alpha_t(x)$ and vector field $v_t(x)$ must satisfy the mass conservation formula, yielding

$$\partial_t \alpha_t(x) + \nabla \cdot (\alpha_t(x) v_t(x)) = 0 \quad \text{and} \quad \alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1. \tag{6}$$

Benamou and Brenier [2] proved that

$$\mathcal{L}_c(\alpha_0,\alpha_1) = \min_{(\alpha_t(x),v_t(x))_t \text{ sat. (6)}} \int_0^1 \int_{\mathbb{R}^d} \|v_t(x)\|^2 \alpha_t(x)\, \mathrm{d}x\mathrm{d}t. \tag{7}$$

To accommodate input measures with varying masses and address local mass discrepancies (unbalanced setting), many normalizations or relaxations have been proposed. A common approach involves incorporating a source term $s_t(x)$ into the continuity Eq. (6). Consequently, we consider

$$\begin{aligned}\overline{\mathcal{C}}(\alpha_0,\alpha_1) \overset{\text{def.}}{=} \{&(\alpha_t(x), v_t(x), s_t(x)) : \partial_t \alpha_t(x) + \nabla \cdot (\alpha_t(x) v_t(x)) = s_t(x), \\ &\alpha_{t=0} = \alpha_0, \alpha_{t=1} = \alpha_1\}.\end{aligned} \tag{8}$$

The crucial question is how to measure the cost associated to this source term and introduce it in the original dynamic formulation, Chizat et al. [7] proposed Wasserstein–Fisher–Rao (WFR) distance

$$\begin{aligned}\mathrm{WFR}^2(\alpha_0,\alpha_1) = \min_{(\alpha_t(x),v_t(x),s_t(x))_t \in \overline{\mathcal{C}}(\alpha_0,\alpha_1)} \int_0^1 \int_{\mathbb{R}^d} \Big(&\|v_t(x)\|^2 \\ &+ \tau \frac{\|s_t(x)\|^2}{(\alpha_t(x))^2}\Big) \alpha_t(x)\, \mathrm{d}x\mathrm{d}t.\end{aligned} \tag{9}$$

to avoid having mass which travels at infinite speed and suddenly grows in a region where there was no mass before.

Gromov–Wasserstein optimal transport and rigid body transformation invariant optimal transport

Previous work [18, 28, 39] used Gromov–Wasserstein (GW) optimal transport [24] to model the spatial coordinates of observations at different time points or different slices at the same time point. Optimal transport relies on the ground cost $\mathbf{C}$ to compare two distributions $\mathbf{a}$ and $\mathbf{b}$, but it does not apply when the two distributions are not defined on identical underlying spaces. To overcome this constraint, a weaker assumption is made, that two matrices $\mathbf{D} \in \mathbb{R}^{n\times n}$ and $\mathbf{D}' \in \mathbb{R}^{m\times m}$ measure the similarity among the points defining the own space. The GW problem reads

$$\mathrm{GW}((\mathbf{a},\mathbf{D}),(\mathbf{b},\mathbf{D}'))^2 \overset{\text{def.}}{=} \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \sum_{i,j,i',j'} \left|\mathbf{D}_{i,i'} - \mathbf{D}'_{j,j'}\right|^2 \mathbf{P}_{i,j}\mathbf{P}_{i',j'}. \tag{10}$$

Due to the unknown rotation and translation of the tissue during spatial transcriptome sequencing, the spatial coordinates of the data at different time points are in different coordinate systems that differ each other by a rigid body transformation (i.e., rotation and translation). When modeling spatial coordinates using GW-OT, the matrices $\mathbf{D}$ and $\mathbf{D}'$ typically use the Euclidean distance between points. This is a reasonable choice

in order to obtain a optimal mapping $\mathbf{P}$ because the GW-OT is rigid body transformation invariant in this setting. However, the GW-OT does not align the spatial coordinates of data at different time points to the same coordinate system, and it is still not possible to compute the ground cost between the spatial coordinates of the data at different time points, and therefore cannot be generalized to dynamical OT.

Cohen and Guibas [8] explicitly model optimal transport under an invariant set $\mathcal{G}$, defining the problem as simultaneously finding the optimal matching and the optimal transformation

$$\mathbf{P}^{\star}, g^{\star} = \arg \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b}), g\in\mathcal{G}} \langle \mathbf{C}(g), \mathbf{P} \rangle \overset{\text{def.}}{=} \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{P}_{ij} d\left(x_i, g\left(y_j\right)\right), \tag{11}$$

and design a monotonically iterative algorithm

$$\mathbf{P}^{(k)} = \arg \min_{\mathbf{P}\in\mathbf{U}(\mathbf{a},\mathbf{b})} \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{P}_{ij} d\left(x_i, g^{(k)}\left(y_j\right)\right), \tag{12}$$

$$g^{(k+1)} = \arg \min_{g\in\mathcal{G}} \sum_{i=1}^{m}\sum_{j=1}^{n} \mathbf{P}_{ij}^{(k)} d\left(x_i, g\left(y_j\right)\right). \tag{13}$$

Subproblem (12) is to solve a static OT. When we select invariant set $\mathcal{G}$ as a rigid body transformation sets in order to model spatial coordinates in spatial transcriptome data and the distance function d uses the Euclidean distance, that is, $d(x, y) = ||x - y||_2^2$, subproblem (13) is a weighted Procrustes problem

$$R^{(k+1)}, r^{(k+1)} = \arg \min_{\substack{R\in\mathbb{R}^{d_s\times d_s}, r\in\mathbb{R}^{d_s} \\ R^T R = I}} \sum_{i,j} \mathbf{P}_{ij}^{(k)} \left\| x_i - R^{k+1} y_j - v \right\|_2^2, \tag{14}$$

where d_s refers to the dimension of the coordinate space, usually 2 or 3, R refers to a rotation matrix, and r refers to a translation vector.

We call this problem rigid body transform invariant optimal transport, and it explicitly models unknown rigid body transformation, making it possible to compute ground cost functions between different distributions, which makes it generalizable to optimal transport in dynamic form.

Dynamical, unbalanced, rigid body transform invariant optimal transport modeling time-series spatial transcriptome data 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 or apoptosis during evolution, unbalanced setting is used. Since the spatial coordinates obtained from sequencing at different time points are not in the same coordinate system, we need to consider the rigid-body transformations invariance of spatial coordinates at different time points. We use $t_0, t_1 \dots t_K$ to refer to time points, $X^0 \in \mathbb{R}^{n_0\times d}, X^1 \in \mathbb{R}^{n_1\times d} \dots X^K \in \mathbb{R}^{n_K\times d}$ to refer to gene expression data after dimensionality reduction, and $Z^0 \in \mathbb{R}^{n_0\times d_s}, Z^1 \in \mathbb{R}^{n_1\times d_s} \dots Z^K \in \mathbb{R}^{n_K\times d_s}$ to refer to spatial coordinates, where $n_0, n_1, \dots n_K$ refers to the number of cells.

stVCR reconstructs $\rho_t(x, z)$, by interpolating the input time-series densities ρ_i using a partial differential equation

$$\partial_t \rho_t(x, z) + \nabla \cdot ([v_{x,t}(x, z), v_{z,t}(x, z)]\rho_t(x, z)) = g_t(x, z)\rho_t(x, z), \tag{15}$$

where $v_{x,t}(x,z)$ refers to the velocity of gene expression, $v_{z,t}(x,z)$ refers to the spatial velocity and $g_t(x,z)$ refers to velocity of cell growth. We take the coordinate system at the t_0 as the base, and assume that the coordinate system at $t_i (i = 1, 2 \dots K)$ differs from the coordinate system at t_0 by one rotation R_i and one translation r_i. We consider

$$\begin{aligned} \overline{\mathcal{C}}\left(X^1, Z^1 \dots X^k, Z^k\right) &\overset{\text{def.}}{=} \{(\rho_t(x,z), v_{x,t}(x,z), v_{z,t}(x,z), g_t(x,z), R_1, r_1, \dots \\ &\dots R_K, r_K) : \partial_t \rho_t(x,z) + \nabla \cdot ([v_{x,t}(x,z), v_{z,t}(x,z)]\rho_t(x,z)) = g_t(x,z)\rho_t(x,z), \\ &\rho_{t=t_0} = \rho_0,\, \rho_{t=t_i} = \rho_i(R_i, r_i), R_i^T R_i = I, i = 1, 2 \dots K\}, \end{aligned} \tag{16}$$

where $\rho_i(R_i, r_i)$ refers to the density of the data expression X^i and spatial coordinates Z^i after being rotated by R_i and translated by r_i at t_i.

We obtain the optimal dynamics $v_{x,t}, v_{z,t}, g_t$ and the optimal rigid body transformation $(R_i, r_i), i = 1, 2 \dots K$ by minimizing the WFR distance

$$\begin{aligned} &\mathrm{WFR}^2\left(X^1, Z^1, \dots X^k, Z^k\right) = \min_{(\rho_t(x,z), v_{x,t}(x,z), v_{z,t}(x,z), g_t(x,z), R_2, r_2, \dots R_k, r_k)\text{sat.}(16)} \\ &\int_{t_0}^{t_K} \int_{\mathbb{R}^{d+d_s}} \left(\|v_{x,t}(x,z)\|^2 + \tau_1 \|v_{z,t}(x,z)\|^2 + \tau_2 \|g_t(x,z)\|^2\right) \rho_t(x,z) \mathrm{d}(x,z)\ \mathrm{d}t. \end{aligned} \tag{17}$$

According to the dimensionless form of the WFR distance derived by Sha et al. [32], we have

$$\begin{aligned} &\mathrm{WFR}^2\left(X^1, Z^1, \dots X^k, Z^k\right) = \min_{(\rho_t(x,z), v_{x,t}(x,z), v_{z,t}(x,z), g_t(x,z), R_2, r_2, \dots R_k, r_k)\text{sat.}(16)} \\ &\mathbb{E}_{(x_0,z_0)\sim\rho_0} \int_{t_0}^{t_K} \left(\|v_{x,t}(x,z)\|^2 + \tau_1 \|v_{z,t}(x,z)\|^2 + \tau_2 \|g_t(x,z)\|^2\right) e^{\int_{t_0}^{t} g_s(x,z)\ \mathrm{d}s}\ \mathrm{d}t \end{aligned} \tag{18}$$

Deep learning-based solver in stVCR

Solving problem (17) or (18) directly is difficult, we use deep learning to find an approximate solution. We suppose that changes in gene expression, cell migration, and cell growth depend on current gene expression and spatial location, and are parameterized by neural networks, i.e., $v_{x,t}(x,z) = v_x(x,z,t;\theta_1)$, $v_{z,t}(x,z) = v_z(x,z,t;\theta_2)$ and $g_t(x,z) = g(x,z,t;\theta_3)$, where θ_1, θ_2 and θ_3 are the parameter of the neural network. Exactly satisfying all the constraints in (16) is difficult, but some of the constraints can be modeled explicitly. R is a rotation matrix that can also be modeled exactly. For the 2D case, consider a rotation angle α, we let

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

to ensure that R is a 2D rotation matrix. For the 3D case, we parameterize a rotation matrix R using the Euler angles α, β, and γ, where α, β, and γ denote the angles of rotation around each of the three coordinate axes. Formally,

$$R_\alpha = \begin{pmatrix} 1 & 0 & 0 \\ 0 & \cos(\alpha) & -\sin(\alpha) \\ 0 & \sin(\alpha) & \cos(\alpha) \end{pmatrix}, \quad R_\beta = \begin{pmatrix} \cos(\beta) & 0 & \sin(\beta) \\ 0 & 1 & 0 \\ -\sin(\beta) & 0 & \cos(\beta) \end{pmatrix},$$

$$R_\gamma = \begin{pmatrix} \cos(\gamma) & -\sin(\gamma) & 0 \\ \sin(\gamma) & \cos(\gamma) & 0 \\ 0 & 0 & 1 \end{pmatrix}, \qquad \text{and} \quad R = R_\alpha R_\beta R_\gamma.$$

Therefore, the parameters we need to optimize are the neural network parameters θ_1, θ_2 and θ_3, as well as the rotation angles α_i (or Euler angles α_i, β_i and γ_i for the 3D case) and the translation r_i, $i = 1, 2 \dots K$.

We take the Lagrangian view, which is to think of the cell population as a group of discrete particles, that is $\rho_i(R_i, r_i) = \frac{1}{n_0}\Sigma_{j=1}^{n_i}\delta_{x_j^i, z_j^i}$, where z_j^i is the coordinate after rotation and translation. Note that the total mass at different time points is not the same, depending on the total number of cells at that time point, and the total mass at t_i is n_i/n_0. We look at the change per particle rather than the change in the probability density, which eliminates the need to do a probability density estimate. Without loss of generality, let's assume that the first time point t_0 is equal to 0. The constraints $\partial_t \rho_t(x,z) + \nabla \cdot ([v_{x,t}(x,z), v_{z,t}(x,z)]\rho_t(x,z)) = g_t(x,z)\rho_t(x,z)$ and $\rho_{t=t_0} = \rho_0$ are equivalent to

$$\begin{cases} \dfrac{\mathrm{d}x(t)}{\mathrm{d}t} = v_x(x,z,t;\theta_1) \\ \dfrac{\mathrm{d}z(t)}{\mathrm{d}t} = v_z(x,z,t;\theta_2) \\ \dfrac{\mathrm{d}\ln w(t)}{\mathrm{d}t} = g(x,z,t;\theta_3) \\ x(0), z(0) = x_0, z_0 \sim \rho_0 \\ w(0) = 1/n_0, \end{cases} \tag{19}$$

which refers to sampling data from the first time point t_0 as an initial value, set the initial weight of each particle is uniform and evolves according to the parameterized dynamics $v_x(x,z,t;\theta_1)$, $v_z(x,z,t;\theta_2)$ and $g(x,z,t;\theta_3)$.

The remaining constraints $\rho_{t=t_i} = \rho_i(R_i, r_i), i = 1, 2 \ldots K$ are difficult to satisfy exactly, and we add them as penalty terms to the loss function. Notice that $\rho_{t=t_i} = \Sigma_{j=1}^{n_0} w_j(t_i)\delta_{x_j(t_i), z_j(t_i)}$, which implies that the total mass of $\rho_{t=t_i}$ and $\rho_i(R_i, r_i)$ are $\Sigma_{j=1}^{n_0} w_j(t_i)$ and n_i/n_0, and are not always equal during training. A natural approach would be to use a static unbalanced OT to measure their distance, but this is costly. We use the

$$d_{match}(\rho_{t=t_i}, \rho_i(R_i, r_i)) = \min_{\mathbf{P}_i \in \mathbf{U}(w(t_i)^{\mathrm{n}}, w_i^{\mathrm{n}})} \langle \mathbf{C}_i, \mathbf{P}_i \rangle + \kappa_0 |\Sigma_{j=1}^{n_0} w_j(t_i) - n_i/n_0| \tag{20}$$

where $w(t_i)^{\mathrm{n}}$ and w_i^{n} refer to the particle weights of the two measures after normalization. $\mathbf{C}$ is the cost function for integrating gene expression and spatial coordinates, usually we use $c_{ij} = c((x_i, z_i), (x_j, z_j)) = ||x_i - x_j||_2^2 + \kappa_1 ||z_i - z_j||_2^2$, where κ_1 an adjustable hyperparameter that measures the importance of gene expression or spatial coordinates. Intuitively, the distance defined in Eq. (20) includes two terms. The first term is the static OT distance after normalization of the two distributions, which promotes the match between the gene expression and spatial coordinates evolved according to dynamics and the observed data; the second term is the absolute value of the total mass difference between the two distributions, which promotes the match between the number of cells evolved according to dynamics and the observed data. The importance of these two terms is weighed by the adjustable hyperparameter κ_0.

In addition, we noticed the dimensionless form of WFR distance in Eq. (18) can be expressed more succinctly by using the notation $w(t)$ for the weight of the particle at time t

$$\begin{aligned} \mathcal{L}_{WFR} = \mathbb{E}_{(x^0, z^0) \sim \rho_0} \int_0^{t_K} \Big(& \|v_x(x,z,t;\theta_1)\|^2 + \tau_1 \|v_z(x,z,t;\theta_2)\|^2 \\ & + \tau_2 \|g(x,z,t;\theta_3)\|^2 \Big) n_0 w(x,z,t) \, \mathrm{d}t. \end{aligned} \tag{21}$$

Summarizing the above discussion, the composition of the loss function includes the

WFR distance $\mathcal{L}_{WFR}$ and the matching error $\mathcal{L}_{Match}$

$$\begin{aligned} &Loss(\theta_1, \theta_2, \theta_3, \alpha_i(\text{or include } \beta_i, \gamma_i), r_i) = \mathcal{L}_{WFR} + \lambda \mathcal{L}_{Match} \\ &= \mathbb{E}_{(x^0, z^0) \sim \rho_0} \int_0^{t_K} \Big(\|v_x(x, z, t; \theta_1)\|^2 + \tau_1 \|v_z(x, z, t; \theta_2)\|^2 \\ &+ \tau_2 \|g(x, z, t; \theta_3)\|^2 \Big) n_0 w(x, z, t) \, \mathrm{d}t + \lambda \Sigma_{i=1}^K d(\rho_{t=t_i}, \rho_i(R_i, r_i)) \end{aligned} \tag{22}$$

Parameter initialization and training details

The structure of our neural networks $v_x(x, z, t; \theta_1)$, $v_z(x, z, t; \theta_2)$ and $g(x, z, t; \theta_3)$ 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 ordinary differential equations (ODE) represented by the neural network, that is neural ODE, which we implement using the torchdiffeq package [6]. It also involves computing the static OT distance, which we implement using the POT package [12]. Finally, in order to enhance the matching with the observed data at each time point, we not only compute the loss function defined by Eq. (22) by sampling the data from the t_0 to evolve to subsequent time points, but likewise compute a similar loss function for Eq. (22) by sampling the data from other time points, evolving forward and backward. For example, for the t_j, we have

$$\begin{aligned} &Loss(\theta_1, \theta_2, \theta_3, \alpha_i(\text{or include } \beta_i, \gamma_i), r_i) = \mathcal{L}_{WFR} + \lambda \mathcal{L}_{Match} \\ &= \mathbb{E}_{(x^j, z^j) \sim \rho_j} - \int_{t_j}^{t_0} + \int_{t_j}^{t_K} \Big(\|v_x(x, z, t; \theta_1)\|^2 + \tau_1 \|v_z(x, z, t; \theta_2)\|^2 \\ &+ \tau_2 \|g(x, z, t; \theta_3)\|^2 \Big) n_0 w(x, z, t) \, \mathrm{d}t + \lambda \Sigma_{i=0, i \neq j}^K d(\rho_{t=t_i}, \rho_i(R_i, r_i)) \end{aligned} \tag{23}$$

In the actual calculation, we choose to sample from the first time point and the last time point to balance accuracy and computation overhead.

Simulated data

The dynamics of our simulated data 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

$$\begin{aligned} \frac{dr}{dt} &= f_1(x)\left(\frac{r^n}{1 + r^n} + \frac{1}{1 + g^n + 10b^n} - r\right) + 0.05 w_t \\ \frac{dg}{dt} &= f_2(x)\left(\frac{g^n}{1 + g^n} + \frac{1}{1 + r^n + 10b^n} - g\right) + 0.05 w_t \\ \frac{db}{dt} &= \frac{b^2}{1 + b^2} - 0.4b + 0.01 w_t \\ \frac{dx}{dt} &= \operatorname{sign}(x) \exp(-4b) \exp(-4g)(r - 2)^2 r^2 + 0.001 w_t \\ \frac{dy}{dt} &= 0 \end{aligned} \tag{24}$$

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 standard Brownian motion. In the calculation, 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. The specific definitions of $f_1(x)$ and $f_2(x)$ are

$$f_1(x) = \begin{cases} \max\left(\exp(0.75(x-1)), 1\right), & x > 0 \\ -f_1(-x), & x < 0 \end{cases} \tag{25}$$

and

$$f_2(x) = \begin{cases} \max\left(\exp(0.75(1-x)), 1\right), & x > 0 \\ -f_2(-x), & x < 0. \end{cases} \tag{26}$$

When $|x| > 1$, $f_1(x) > 1$ 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. Cell growth modeled as division and apoptosis

$$\text{growth} = g_{\text{division}} - g_{\text{apoptosis}} = \frac{2r}{1+r}\frac{|x|}{1+|x|} - \frac{g}{1+g}, \tag{27}$$

which means that the gene *Red* and migrating outward in the horizontal direction will promote cell growth while the gene *Green* will inhibit cell growth.

We discretized time in order to obtain data through numerical simulations. Within the time step δ_t, we simulate gene expression and spatial coordinates after δ_t according to the forward Euler scheme. Cell division or apoptosis is determined by the growth function, specifically within the δ_t, the probability of cell division is $p_{\text{division}} = g_{\text{division}}\delta_t$, the probability of apoptosis is $p_{\text{apoptosis}} = g_{\text{apoptosis}}\delta_t$, and the probability of remaining unchanged is $1 - p_{\text{division}} - p_{\text{apoptosis}}$. When a cell divides into two, we add white noise with a standard deviation of 0.01 to the gene expression and spatial coordinates of the original cell and become the data of the new cell. In addition, at each step of the simulation when there is a negative value of gene expression, we force it to be corrected to 0. At the initial time point we randomly sampled three groups of cells. The gene expression (r, g, b) of the first group of cells was sampled from the normal distribution with mean $(0, 0, 2)$ and covariance matrix of $0.1I$, and the spatial coordinates (x, y) were sampled from the uniform distribution $[-2, 2] \times [-3, 3]$. Gene expression in both the second and third groups of cells was sampled from the normal distribution with mean $(0, 0, 0)$ and covariance matrix of $0.1I$, in addition the spatial coordinates of the second group of cells were sampled from a uniform distribution $[-2, -1.5] \cap [1.5, 2] \times [-3, 3]$, and the spatial coordinates of the third group of cells were sampled from a uniform distribution $[-0.5, 0.5] \times [-3, 3]$. We evolved the cell 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.

Time-dependent cell type classifier in stVCR

After we recovered the entire cell developmental 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 $p = f_\theta(x, z, t)$. The inputs are gene expression x, spatial coordinates z, and time t, and the outputs are a probability distribution p indicating

the probability that the current cell is of each cell type. The loss function is

$$\begin{aligned} Loss &= \frac{1}{N}\Sigma_{i=0}^{N}\left(CrossEntropyLoss(p_i, y_i) + \lambda\left|\frac{\partial p_i}{\partial t_i}\right|_1\right) \\ &= \frac{1}{N}\Sigma_{i=0}^{N}\left(\left(-\Sigma_{j=0}^{K-1} y_{i,j}\log(p_{i,j})\right) + \lambda\left|\frac{\partial p_i}{\partial t_i}\right|_1\right), \end{aligned} \tag{28}$$

where the first term is the standard cross-entropy loss function in multiclassification tasks, and the second term is the 1-norm of the partial derivative of the neural network output with respect to time t to promote the neural network's predictions to be as smooth as possible with respect to changes in time t.

Downstream analysis

For notational simplicity, we omit the notation for the parameters of the neural network in the following description. When we use the stVCR to obtain the gene expression velocity in the embedded space $v_{x_{\text{emd}}}(x_{\text{emd}}, z, t)$, the spatial velocity $v_z(x_{\text{emd}}, z, t)$, the cell growth rate $g(x_{\text{emd}}, z, t)$, and the time-dependent classifier $f_{\text{classifier}}(x_{emd}, z, t)$, 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 growth. In addition, since we have already obtained the $x_{emd} = f_{\text{enc}}(x)$ and a decoder $x = f_{\text{dec}}(x_{emd})$ of the cell's gene expression in advance, we can easily switch between the embedding space and the original gene expression space

$$\begin{aligned} v_x(x, z, t) &= f_{\text{dec}}(f_{\text{enc}}(x) + v_{x_{\text{emd}}}(f_{\text{enc}}(x), z, t)), \\ v_z(x, z, t) &= v_z(f_{\text{enc}}(x), z, t), \\ g(x, z, t) &= g(f_{\text{enc}}(x), z, t). \end{aligned} \tag{29}$$

Below we give a more detailed description of the downstream analysis task.

Interpolation and prediction

For some interpolation time t_{int} that the user cares about, 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 x_i and z_i at t_{obs} as the initial values, and then evolve according to the dynamics learned by stVCR, so as to obtain the interpolation result. For the i_{th} cell to evolve from time t to $t + \delta t$, more specifically,

$$\begin{aligned} x_{i,\text{emd}}(t + \delta t) &= x_{i,\text{emd}}(t) + v_{x_{\text{emd}}}(x_{i,\text{emd}}(t), z_i(t), t)\delta_t; \\ z_i(t + \delta t) &= z_i(t) + v_z(x_{i,\text{emd}}(t), z_i(t), t)\delta_t; \\ &\text{Generate random number } \gamma \sim U[0, 1]; \\ \text{If } g(x_{i,\text{emd}}(t), z_i(t), t) > 0 \ \&\ \gamma &< g(x_{i,\text{emd}}(t), z_i(t), t)\delta t, \text{ cell division;} \\ \text{If } g(x_{i,\text{emd}}(t), z_i(t), t) < 0 \ \&\ \gamma &< -g(x_{i,\text{emd}}(t), z_i(t), t)\delta t, \text{ cell apoptosis.} \end{aligned} \tag{30}$$

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 use subscript i to denote the sample i and superscript j to denote the jth component of a vector. For a cell i, as well as the target gene j and the source gene k of interest, we can calculate $\frac{\partial v_x(x_i, z_i, t)^j}{\partial x^k}$, which represents how increased expression of gene

k changes the velocity of gene j in cell i. If this is a positive value, it means that gene k promotes gene j in cell i, and vice versa it inhibits gene j. The study of gene interactions by calculating the Jacobian matrix between genes is widely used in many methods, like Dynamo [27] and TIGON [32].

Similarly, for cell i, and axes j and k of interest, we can calculate $\frac{\partial v_z(x_i,z_i,t)^j}{\partial z^k}$ to study space-space interaction. This is used in Spateo [28], although their spatial vector fields $v_z(z)$ are independent of gene expression x and time t.

In stVCR, gene expression x and spatial coordinates z interact, which means we can study how cell migration affects gene expression. For cell i, a target gene j, and a given unit direction $\vec{n} = (n_1, n_2, n_3)$ (or $\vec{n} = (n_1, n_2)$ for 2D case), we can calculate the directional derivative

$$\frac{\partial v_x(x_i, z_i, t)^j}{\partial \vec{n}} = \frac{\partial v_x(x_i, z_i, t)^j}{\partial z^1} n_1 + \frac{\partial v_x(x_i, z_i, t)^j}{\partial z^2} n_2 + \frac{\partial v_x(x_i, z_i, t)^j}{\partial z^3} n_3,$$

which describes how the migration of cell i to the given direction $\vec{n}$ 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 calculate

$$(\frac{\partial v_z(x_i, z_i, t)^1}{\partial x^j}, \frac{\partial v_z(x_i, z_i, t)^2}{\partial x^j}, \frac{\partial v_z(x_i, z_i, t)^2}{\partial x^j}) := \vec{n}_{i,x^j},$$

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

Cell specific effects of gene and spatial migration on growth

For cell i and gene j of interest, we can calculate $\frac{\partial g(x_i,z_i,t)}{\partial x^j}$, which describes the effect of gene j on cell growth, where a positive value means promoting a negative value and the opposite. This is used in TIGON [32]. However, their growth function $g(x, t)$ does not depend on spatial coordinates z, so the effect of cell spatial migration on growth cannot be studied.

In stVCR, cell growth 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 $\vec{n} = (n_1, n_2, n_3)$, we can calculate directional derivative

$$\frac{\partial g(x_i, z_i, t)}{\partial \vec{n}} = \frac{\partial g(x_i, z_i, t)}{\partial z^1} n_1 + \frac{\partial g(x_i, z_i, t)}{\partial z^2} n_2 + \frac{\partial g(x_i, z_i, t)}{\partial z^3} n_3,$$

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

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 $f_{\text{classifier}}(x_{emd}, z, t)$, we can construct developmental lineages of cells of interest.

Data and code availability statement

In this study, we used the public axolotl brain regeneration datasets [37]. stVCR is implemented in Python and is available at `https://github.com/QiangweiPeng/stVCR`.

Supporting Information

Supplementary Figure S1

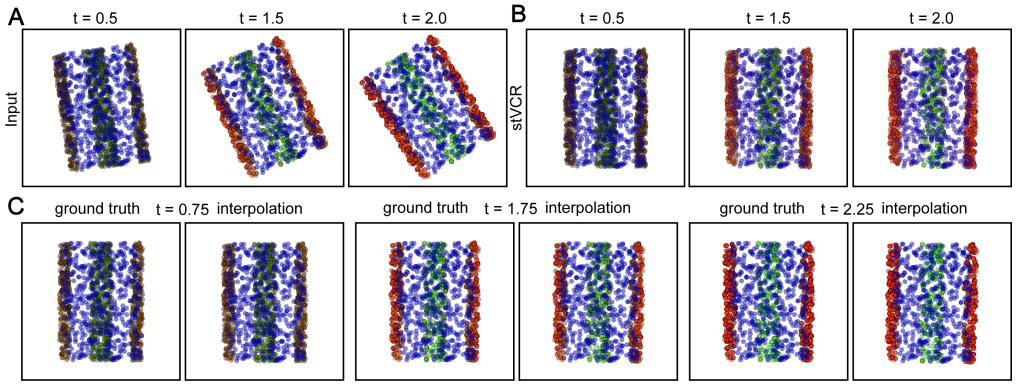

Figure S1. Benchmark on the simulated time-series ST data, related to Fig. 2 **A.** Input data at t=0.5, 1.5 and 2.0. The color was determined by the expression of three genes *Red*, *Green* and *Blue*. **B.** The evolution results at t=0.5, 1.5 and 2.0 of cells at t=0.0 according to learned dynamics using stVCR. **C.** Results of stVCR interpolation at t=0.75, t=1.75 and t=2.25 and comparison with ground truth. Left: ground truth; Right: stVCR.

Supplementary Figure S2

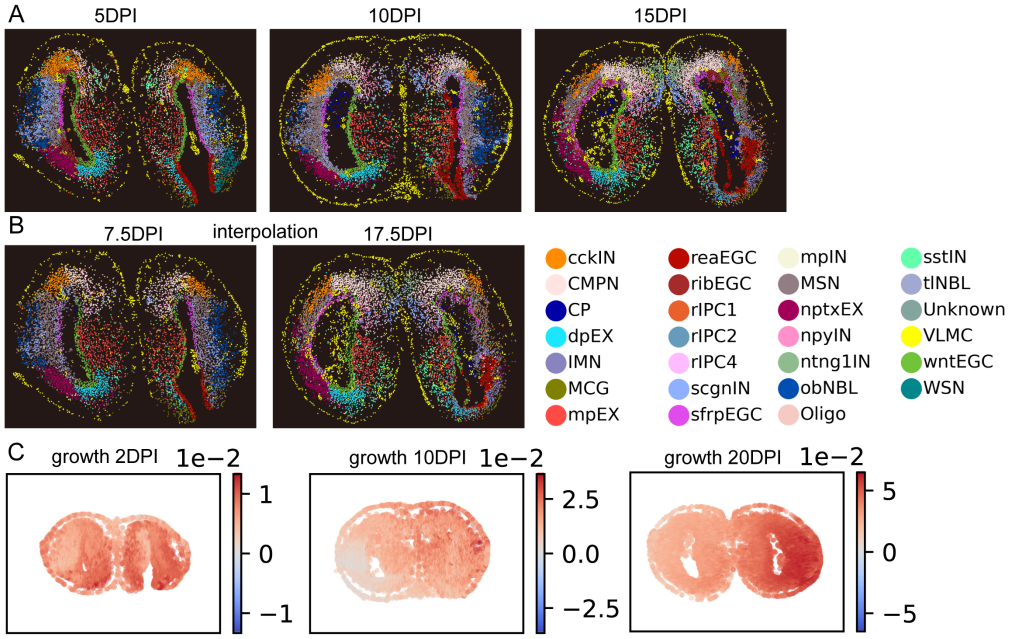

Figure S2. stVCR analysis of axolotl brain regeneration, related to Fig. 3 **A.** stVCR aligns the spatial coordinates of data at different time points to the same coordinate system. Left: 5DPI; Middle: 10DPI; Right: 15DPI. Cell type annotations come from the original study. **B.** stVCR interpolation at 7.5DPI and 17.5DPI. Cell type annotations come from the stVCR's time-dependent classifier. **C.** stVCR cell growth rate at 2DPI, 10DPI and 20DPI data.

Supplementary Video 1

Comparison of the true dynamics of simulated data with those learned by stVCR.

Supplementary Video 2

Axolotl brain regeneration dynamics learned by stVCR.

Acknowledgments

References

1. N. Battich, J. Beumer, B. De Barbanson, L. Krenning, C. S. Baron, M. E. Tanenbaum, H. Clevers, and A. Van Oudenaarden. Sequencing metabolically labeled transcripts in single cells reveals mrna turnover strategies. *Science*, 367(6482):1151–1156, 2020.

2. J.-D. Benamou and Y. Brenier. A computational fluid mechanics solution to the monge-kantorovich mass transfer problem. *Numerische Mathematik*, 84(3):375–393, 2000.

3. V. Bergen, M. Lange, S. Peidli, F. A. Wolf, and F. J. Theis. Generalizing rna velocity to transient cell states through dynamical modeling. *Nature biotechnology*, 38(12):1408–1414, 2020.

4. J. Cao, W. Zhou, F. Steemers, C. Trapnell, and J. Shendure. Sci-fate characterizes the dynamics of gene expression in single cells. *Nature biotechnology*, 38(8):980–988, 2020.

5. A. Chen, S. Liao, M. Cheng, K. Ma, L. Wu, Y. Lai, X. Qiu, J. Yang, J. Xu, S. Hao, et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10):1777–1792, 2022.

6. R. T. Chen, Y. Rubanova, J. Bettencourt, and D. K. Duvenaud. Neural ordinary differential equations. *Advances in neural information processing systems*, 31, 2018.

7. L. Chizat, G. Peyré, B. Schmitzer, and F.-X. Vialard. An interpolating distance between optimal transport and fisher–rao metrics. *Foundations of Computational Mathematics*, 18:1–44, 2018.

8. S. Cohen and L. Guibasm. The earth mover's distance under transformation sets. In *Proceedings of the Seventh IEEE International Conference on Computer Vision*, volume 2, pages 1076–1083. IEEE, 1999.

9. H. Cui, H. Maan, M. C. Vladoiu, J. Zhang, M. D. Taylor, and B. Wang. Deepvelo: deep learning extends rna velocity to multi-lineage systems with cell-specific kinetics. *Genome Biology*, 25(1):27, 2024.

10. F. Erhard, M. A. Baptista, T. Krammer, T. Hennig, M. Lange, P. Arampatzi, C. S. Jürges, F. J. Theis, A.-E. Saliba, and L. Dölken. scslam-seq reveals core features of transcription dynamics in single cells. *Nature*, 571(7765):419–423, 2019.

11. S. Farrell, M. Mani, and S. Goyal. Inferring single-cell transcriptomic dynamics with structured latent gene expression dynamics. *Cell Reports Methods*, 3(9), 2023.

12. R. Flamary, N. Courty, A. Gramfort, M. Z. Alaya, A. Boisbunon, S. Chambon, L. Chapel, A. Corenflos, K. Fatras, N. Fournier, L. Gautheron, N. T. Gayraud, H. Janati, A. Rakotomamonjy, I. Redko, A. Rolet, A. Schutz, V. Seguy, D. J. Sutherland, R. Tavenard, A. Tong, and T. Vayer. Pot: Python optimal transport. *Journal of Machine Learning Research*, 22(78):1–8, 2021.

13. M. Gao, C. Qiao, and Y. Huang. Unitvelo: temporally unified rna velocity reinforces single-cell trajectory inference. *Nature Communications*, 13(1):6586, 2022.

14. A. Gayoso, P. Weiler, M. Lotfollahi, D. Klein, J. Hong, A. Streets, F. J. Theis, and N. Yosef. Deep generative modeling of transcriptional dynamics for rna velocity analysis in single cells. *Nature methods*, 21(1):50–59, 2024.

15. G.-J. Hendriks, L. A. Jung, A. J. Larsson, M. Lidschreiber, O. Andersson Forsman, K. Lidschreiber, P. Cramer, and R. Sandberg. Nasc-seq monitors rna synthesis in single cells. *Nature communications*, 10(1):3138, 2019.

16. K. Holler, A. Neuschulz, P. Drewe-Boß, J. Mintcheva, B. Spanjaard, R. Arsiè, U. Ohler, M. Landthaler, and J. P. Junker. Spatio-temporal mrna tracking in the early zebrafish embryo. *Nature communications*, 12(1):3358, 2021.

17. G. Huguet, D. S. Magruder, A. Tong, O. Fasina, M. Kuchroo, G. Wolf, and S. Krishnaswamy. Manifold interpolating optimal-transport flows for trajectory inference. *Advances in neural information processing systems*, 35:29705–29718, 2022.

18. D. Klein, G. Palla, M. Lange, M. Klein, Z. Piran, M. Gander, L. Meng-Papaxanthos, M. Sterr, A. Bastidas-Ponce, M. Tarquis-Medina, et al. Mapping cells through time and space with moscot. *bioRxiv*, pages 2023–05, 2023.

19. G. La Manno, R. Soldatov, A. Zeisel, E. Braun, H. Hochgerner, V. Petukhov, K. Lidschreiber, M. E. Kastriti, P. Lönnerberg, A. Furlan, et al. Rna velocity of single cells. *Nature*, 560(7719):494–498, 2018.

20. S. Li, P. Zhang, W. Chen, L. Ye, K. W. Brannan, N.-T. Le, J.-i. Abe, J. P. Cooke, and G. Wang. A relay velocity model infers cell-dependent rna velocity. *Nature biotechnology*, 42(1):99–108, 2024.

21. T. Li, J. Shi, Y. Wu, and P. Zhou. On the mathematics of rna velocity i: theoretical analysis. *bioRxiv*, pages 2020–09, 2020.

22. S. Lin, K. Yin, Y. Zhang, F. Lin, X. Chen, X. Zeng, X. Guo, H. Zhang, J. Song, and C. Yang. Well-temp-seq as a microwell-based strategy for massively parallel profiling of single-cell temporal rna dynamics. *Nature Communications*, 14(1):1272, 2023.

23. J. Liu, Y. Gu, C. Li, and J. D. Welch. Mapping cell fate transition in space and time. *bioRxiv*, pages 2024–02, 2024.

24. F. Mémoli. Gromov–wasserstein distances and the metric approach to object matching. *Foundations of computational mathematics*, 11:417–487, 2011.

25. Q. Peng, X. Qiu, and T. Li. Storm: Incorporating transient dynamics to infer the rna velocity with metabolic labeling information. *bioRxiv*, pages 2023–06, 2023.

26. Q. Qiu, P. Hu, X. Qiu, K. W. Govek, P. G. Cámara, and H. Wu. Massively parallel and time-resolved rna sequencing in single cells with scnt-seq. *Nature methods*, 17(10):991–1001, 2020.

27. X. Qiu, Y. Zhang, J. D. Martin-Rufino, C. Weng, S. Hosseinzadeh, D. Yang, A. N. Pogson, M. Y. Hein, K. H. J. Min, L. Wang, et al. Mapping transcriptomic vector fields of single cells. *Cell*, 185(4):690–711, 2022.

28. X. Qiu, D. Y. Zhu, J. Yao, Z. Jing, L. Zuo, M. Wang, K. H. Min, H. Pan, S. Wang, S. Liao, et al. Spateo: multidimensional spatiotemporal modeling of single-cell spatial transcriptomics. *BioRxiv*, pages 2022–12, 2022.

29. S. G. Rodriques, R. R. Stickels, A. Goeva, C. A. Martin, E. Murray, C. R. Vanderburg, J. Welch, L. M. Chen, F. Chen, and E. Z. Macosko. Slide-seq: A scalable technology for measuring genome-wide expression at high spatial resolution. *Science*, 363(6434):1463–1467, 2019.

30. M. Scetbon and M. Cuturi. Low-rank optimal transport: Approximation, statistics and debiasing. *Advances in Neural Information Processing Systems*, 35:6802–6814, 2022.

31. G. Schiebinger, J. Shu, M. Tabaka, B. Cleary, V. Subramanian, A. Solomon, J. Gould, S. Liu, S. Lin, P. Berube, et al. Optimal-transport analysis of single-cell gene expression identifies developmental trajectories in reprogramming. *Cell*, 176(4):928–943, 2019.

32. Y. Sha, Y. Qiu, P. Zhou, and Q. Nie. Reconstructing growth and dynamic trajectories from single-cell transcriptomics data. *Nature Machine Intelligence*, 6(1):25–39, 2024.

33. P. L. Ståhl, F. Salmén, S. Vickovic, A. Lundmark, J. F. Navarro, J. Magnusson, S. Giacomello, M. Asp, J. O. Westholm, M. Huss, et al. Visualization and analysis of gene expression in tissue sections by spatial transcriptomics. *Science*, 353(6294):78–82, 2016.

34. R. R. Stickels, E. Murray, P. Kumar, J. Li, J. L. Marshall, D. J. Di Bella, P. Arlotta, E. Z. Macosko, and F. Chen. Highly sensitive spatial transcriptomics at near-cellular resolution with slide-seqv2. *Nature biotechnology*, 39(3):313–319, 2021.

35. A. Tong, J. Huang, G. Wolf, D. Van Dijk, and S. Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. In *International conference on machine learning*, pages 9526–9536. PMLR, 2020.

36. A. Tong, N. Malkin, G. Huguet, Y. Zhang, J. Rector-Brooks, K. Fatras, G. Wolf, and Y. Bengio. Conditional flow matching: Simulation-free dynamic optimal transport. *arXiv preprint arXiv:2302.00482*, 2(3), 2023.

37. X. Wei, S. Fu, H. Li, Y. Liu, S. Wang, W. Feng, Y. Yang, X. Liu, Y.-Y. Zeng, M. Cheng, et al. Single-cell stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration. *Science*, 377(6610):eabp9444, 2022.

38. Z. Xu, A. Sziraki, J. Lee, W. Zhou, and J. Cao. Dissecting key regulators of transcriptome kinetics through scalable single-cell rna profiling of pooled crispr screens. *Nature Biotechnology*, pages 1–6, 2023.

39. R. Zeira, M. Land, A. Strzalkowski, and B. J. Raphael. Alignment and integration of spatial transcriptomics data. *Nature Methods*, 19(5):567–575, 2022.

40. K. Zhang, J. Zhu, D. Kong, and Z. Zhang. Modeling single cell trajectory using forward-backward stochastic differential equations. *PLOS Computational Biology*, 20(4):e1012015, 2024.