DALI LEARNS RULES GENERATING SPATIOTEMPORAL TRANSCRIPTOMICS

Salil Bhate[*]
Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA

Sandeep Kambhampati[*]
Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA
Harvard Medical School
Boston, MA 02114, USA

Mehrtash Babadi
Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA

Fei Chen[†]
Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA
Harvard University
Cambridge, MA 02138, USA
chenf@broadinstitute.org

Caroline Uhler[†]
Broad Institute of MIT and Harvard
Cambridge, MA 02142, USA
Massachusetts Institute of Technology
Cambridge, MA 02139, USA
cuhler@broadinstitute.org

ABSTRACT

Learning how molecular interactions dictate the self-organization of biological systems is a longstanding goal in developmental biology. Time-series spatial transcriptomics (spatiotemporal transcriptomics) provides data to interrogate such dynamic tissue processes. Existing computational methods for spatiotemporal transcriptomics focus on aligning cells and spatial regions across timepoints and learning the spatiotemporal determinants of cell-fate transitions, but do not identify interactions between local spatial processes that generate developmental patterns. To address this, we introduce *Dynamical Analysis of Latent Interactions* (DALI), a framework that takes as input a time series of spatial transcriptomic snapshots and simultaneously fits a latent partial differential equation (PDE) and a spatial registration model to learn rules that generate tissue development. On two Stereo-seq datasets of Zebrafish and Mouse embryogenesis, DALI learns rules that predict gene expression from held-out tissue regions better than baseline approaches. We furthermore interpret these rules: the learned latent variables correspond to meaningful gene programs, and their learned interactions to developmentally important signaling interactions with previous causal evidence.

Code availability: Software is publicly available at `https://github.com/broadinstitute/latent-dynamical-systems`

1 INTRODUCTION

Biological systems display the remarkable property of self-organization: complex patterns emerge from simple, local interactions of individual components of the system. During the development of multicellular organisms, individual cells produce reproducible and robust spatial arrangements of diverse cell types through a process known as morphogenesis. Cells achieve morphogenesis

[*]Equal contribution as first authors
[†]Equal contribution as senior authors

by responding to signals in their local environment, such as secreted molecules ("morphogens") or direct contact with their neighbors. These local interactions collectively give rise to complex tissues, whole organs, and ultimately entire organisms.

Computational frameworks such as Conway's Game of Life Gardner (1970) have demonstrated how simple, local update rules are sufficient to generate complex spatiotemporal patterns. They have also inspired the design of artificial systems capable of growth and repair, such as Neural Cellular Automata (NCA) Mordvintsev et al. (2020). These models have offered conceptual insights into emergent behavior and self-organization, but they have largely been applied to toy models, such as RGB images.

Experimental advances have recently transformed the ability to observe self-organization. Spatial transcriptomics is a powerful data modality, which measures gene expression in tissues with cellular or even subcellular resolution. Although the modality is destructive, static snapshots can be profiled from multiple organisms along a time-series trajectory, revealing the spatiotemporal evolution of gene expression. Time-series spatial transcriptomics data can therefore enable the study of how molecular and cellular interactions give rise to dynamic tissue processes, such as embryonic development, tissue regeneration, and disease progression.

Models based on Optimal Transport (OT) have proven effective for aligning tissue states across time and inferring cellular transitions from time-series spatial transcriptomics data Halmos et al. (2025b;a); Klein et al. (2025); Qiu et al. (2024). Such approaches connect cells and spatial regions across timepoints, but they do not directly model how local interactions produce global spatiotemporal trajectories. Furthermore, generative models have been applied to learn factors underlying single-cell gene expression distributions, including dynamics Tong et al. (2020); von Bassewitz et al. (2025). However, it has not been determined whether generative models that constrain latent variables to exhibit simple spatiotemporal interactions can meaningfully recapitulate biological interactions underlying morphogenesis.

Here, we introduce *Dynamical Analysis of Latent Interactions* (DALI), a framework that simultaneously fits a partial differential equation (PDE) modeling spatiotemporal interactions between latent variables, a decoder into gene expression space, and spatial registration, to model time-course spatial transcriptomics data. We utilize a reaction-diffusion PDE as a minimal prior on the dynamics of the latent variables as signals that spread out and interact. We quantitatively evaluate DALI with interpolation and extrapolation using baselines that account for simple ways to combine gene expression patterns. We further show that the learned dynamics finds meaningful gene expression programs and interactions between them.

2 Related Work

Reaction-Diffusion between Morphogens The Chemical Basis of Morphogenesis Turing (1952), introduced a widely experimentally validated model for biological pattern formation using reaction-diffusion PDEs. In this model, a system consists of a "field", representing the spatial concentration of a particular morphogen. Interactions between diffusing fields can produce spatial patterns such as stripes and spots from a homogeneous initial condition.

Cellular Automata/Neural Cellular Automata Cellular automata (CA) are discrete dynamical systems composed of a grid of cells, each of which can exist in a finite number of states Ulam (1962); von Neumann & Burks (1966). The state of each cell evolves in discrete time steps according to a local update rule that depends on its own state and the states of its neighboring cells. Simple deterministic rules can give rise to complex self-organizing patterns. Neural Cellular Automata (NCA) extend the classical cellular automaton paradigm by replacing hand-crafted update rules with a learnable, differentiable rule parameterized by a neural network Mordvintsev et al. (2020).

Spatiotemporal transcriptomics registration. Previous work has modeled spatial transcriptomics data with continuous vector fields Clifton et al. (2023); Qiu et al. (2024), and shown that distinct samples can be aligned with a learnable spatial deformation of vector fields akin to template construction in biomedical imaging. Furthermore, optimal transport approaches learn couplings between cells across timepoints that preserve pairwise expression and spatial relationships Klein et al. (2025).

3 METHODS

3.1 PROBLEM STATEMENT

We consider time-series data of 2D spatially resolved gene-expression measurements of embryonic development: $X = (X^0, X^1, \ldots, X^T)$, where each $X^t = \{(x_i^t, g_i^t) : i = 1, \ldots, n_t\}$ denotes the measurements at absolute time t post-fertilization. Here, $x_i^t \in \mathbb{R}^2$ is the spatial sampling location and $g_i^t \in \mathbb{R}^G$ is the corresponding gene-expression vector, with G the total number of assayed genes. Because the assay is destructive, each timepoint X^t is obtained from a different individual organism.

Our goal is to find a "template" trajectory $\psi : \mathbb{R}^2 \times [0, T] \to \mathbb{R}^G$, where $\psi(x, t)$ is the gene expression at position x in space at timepoint t. This template trajectory models a generative process underlying development by satisfying the following core assumptions:

1. There exist organism-specific registration functions $R^t : \mathbb{R}^2 \to \mathbb{R}^2$ that spatially deform the trajectory to fit the data from each timepoint, i.e. for each $(x_i, g_i) \in X^t$, we have $g_i \approx \psi(R^t(x_i))$.

   Motivation: In the case of a highly stereotyped process (such as development), the spatial transcriptomics data from any individual organism should be obtainable from the template by spatial deformation.

2. There exists a *latent trajectory* $\tilde{\psi} : \mathbb{R}^2 \times [0, T] \to \mathbb{R}^d$ and *mixing* function $M : \mathbb{R}^d \to \mathbb{R}^G$ where $d \ll G$ such that:
   $$\psi = M \circ \tilde{\psi}$$
   Motivation: Embryonic development is canonically driven by a few key signals that control cellular processes Sanz-Ezquerro et al. (2017). The latent trajectory models the levels of a small ($d \ll G$) number of such signals putatively driving development. At any spatial location x at timepoint t, the mixing function M decodes the levels of the signals, $\tilde{\psi}(x, t)$, into the gene expression at that location.

3. The latent trajectory $\tilde{\psi}$ satisfies a reaction-diffusion partial differential equation (PDE) with drift of the following form,
   $$\frac{\partial \tilde{\psi}}{\partial t} = \Delta \tilde{\psi} + F\big(\tilde{\psi}, \frac{\partial \tilde{\psi}}{\partial y}, \frac{\partial \tilde{\psi}}{\partial y}\big).$$
   The function $F : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}^d$ is the combination of the local reaction and drift effects.

   Motivation: The assumed dynamics on $\tilde{\psi}$ enables us to interpret each latent component d as the concentration of an indirectly-observed signal that diffuses through space and interacts with other signals to specify the gene expression vectors at a particular location. The dynamics of the latent signals should not be interpreted as directly corresponding to the physical diffusion of cells or molecules: the resolution of the data is not enough fit such models. Instead, we aim to capture our prior assumption that developmental transitions are dictated by local interactions between signals that spread out through space.

In this work, we represent the registration R^t, the reaction-drift function F, and mixing function M with neural networks and estimate them with gradient descent. We evaluate models by reconstruction of held-out positions/timepoints as well as by interpreting the gene expression programs and interactions. Our overall architecture is depicted in the schematic Figure 1.

3.2 MODEL ARCHITECTURE: LATENT PDE IN DISCRETIZED SPACE AND TIME

We model the latent trajectory with a tensor describing its values on a grid

$$z_{abc}^t := \tilde{\psi}((a, b), t)_c \in \mathbb{R},$$

where $a, b \in 1, \ldots, L$ are grid points in the x and y dimensions respectively, $t \in 1, \ldots, nT$ are timepoints and $c \in 1, \ldots d$ are latent dimensions. We assume that in between each timepoint of the observed spatial transcriptomics data, there are n steps of the PDE.

The dynamics of z models the latent trajectory with the following equation:

$$z_{abc}^{t+1} = H[z^t]_{abc} := \max\left(0, z_{abc}^t + \sigma\lambda_c(\Delta z^t)_{abc} + (1-\sigma)F(z^t, \partial_x z^t, \partial_y z^t)_{abc}\right), \tag{1}$$

where:

- $\Delta z^t, \partial_x z^t$ and $\partial_y z^t$ are estimates of the Laplacian and spatial gradients respectively, obtained by convolution for each channel as:

  $$\partial_x z^t = K_x^{\text{Sobel}} * z^t, \qquad \partial_y z^t = K_y^{\text{Sobel}} * z^t, \qquad \Delta z_t = K_\Delta * z^t,$$

  where $*$ denotes 2D convolution applied independently to each channel, where K^{Sobel} is the Sobel filter and K_Δ is the discrete Laplacian:

  $$K_x^{\text{Sobel}} = \begin{bmatrix} 1 & 0 & -1 \\ 2 & 0 & -2 \\ 1 & 0 & -1 \end{bmatrix}, \qquad K_y^{\text{Sobel}} = \begin{bmatrix} 1 & 2 & 1 \\ 0 & 0 & 0 \\ -1 & -2 & -1 \end{bmatrix},$$

  $$K_\Delta = \begin{bmatrix} 0 & 1 & 0 \\ 1 & -4 & 1 \\ 0 & 1 & 0 \end{bmatrix}. \tag{2}$$

- F is a pointwise linear function of its three arguments followed by a tanh nonlinearity to limit the magnitude of the reaction updates:

  $$F(z^t, \partial_x z^t, \partial_y z^t)_{abc} = \tanh\left[\sum_{c'} W_{c,c'}^{(F;0)} z_{abc'}^t + \sum_{c'} W_{c,c'}^{(F;x)} \partial_x z_{abc'}^t + \sum_{c'} W_{c,c'}^{(F;y)} \partial_y z_{abc'}^t + b_c^{(F)}\right], \tag{3}$$

  where $W^{(F;0)}$, $W^{(F;x)}$, $W^{(F;y)}$, and $b^{(F)}$ are the learnable parameters.
- The initial condition z_{abc}^0 is learned freely to reconstruct the first time point in the spatial transcriptomics dataset.
- λ_c is a learned positive real number for each latent dimension capturing diffusion rates
- $\sigma \in [0, 1]$ is a learned global parameter that weighs the reaction and diffusion components of the model.

This dynamics mirrors Neural Cellular Automata (NCA) Mordvintsev et al. (2020) in that it is a convolutional recurrent neural network with fixed, biologically-inspired kernels. In contrast to NCA, each grid point is a spatial location; the dynamics is thus intended to model interactions between the concentrations of signals at each such location rather than rules operating within individual cells.

3.3 Model Architecture: Decoder

The decoder is a map $M : \mathbb{R}^d \to \mathbb{R}^G$ that predicts gene expression values from the values of the d latent trajectory. We quantitatively evaluate two architectures for the decoder:

- A linear nonnegative decoder (termed "NMF decoder") enables easily interpreting the latent dimensions but has limited expressivity to fit gene-expression values
- A more flexible multi-layer perceptron (MLP) decoder

We further evaluate regularization to (a) encourage a more biologically meaningful latent space for the decoder and (b) incorporate prior gene-level information, such as sequence features of the gene promoter region "sequence prior regularization". Further details on the decoder architecture and regularization can be found in Appendix A.2 and A.3.

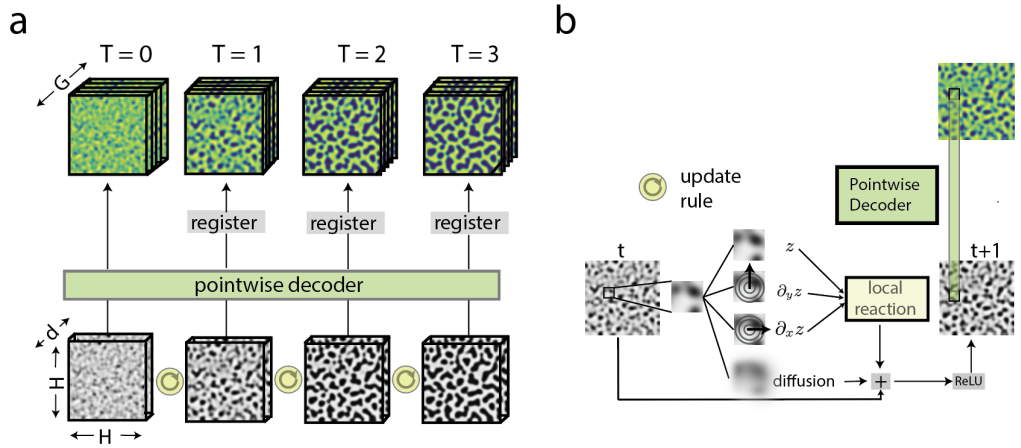

Figure 1: Schematic of DALI architecture for spatiotemporal data. **(a)** A PDE in the d-dimensional latent space is represented by the update rule in **(b)**. Latent variables are decoded to gene expression and transformed by an affine registration.

3.4 Model Architecture: Learned Affine Registration

The data is obtained by destructively sampling tissue sections from different organisms at each timepoint. Therefore, we do not expect a perfect physical alignment across timepoints. To account for this, we register the latent trajectory to each sample X^t end-to-end with an affine transformation R^t. We fix R^0, the registration map for the first timepoint, as the identity. Specifically, we learn

$$R^t : \mathbb{R}^2 \to \mathbb{R}^2,\ \mathbf{x} \mapsto Q^t\mathbf{x} + \mathbf{b},$$

where $Q^t \in \mathbb{R}^{2\times 2}$ is a matrix and $\mathbf{b} \in \mathbb{R}^2$ is a bias.

We implement registration using spatial grid sampling as implemented in spatial transformer networks Jaderberg et al. (2015). In addition, we manually initialize the registration parameters using the overall shape to avoid local minima.

3.5 Optimization

The set of learnable parameters, denoted by Θ, comprises: the initial condition of the latent trajectory z^0, the diffusion rates $\{\lambda_c\}$, the reaction-diffusion weight σ, the reaction network parameters Θ_F, the decoder parameters Θ_M, and the affine registration parameters $\{R^t\}$ for $t > 0$.

We estimate the gene expression for an observed spatial location $x_i^t \in \mathbb{R}^2$ at embryonic time t by evolving the latent dynamics, decoding to gene space, and sampling the resulting field at the registered coordinate. Let z^t denote the latent grid at embryonic time t, obtained by recursively applying the update $H(\Theta)$ (Eq. 1) for nt steps starting from z^0. The predicted expression is:

$$\hat{g}(x_i^t, z^t, t; \Theta) = M\left(\text{Sample}\left(z^t, R^t(x_i^t)\right); \Theta_M\right), \tag{4}$$

where $\text{Sample}(Z, x)$ denotes bilinear interpolation of the grid Z at the query coordinate x.

We optimize Θ by minimizing a composite loss function that penalizes reconstruction error on tissue samples as well as enforces background and regularization constraints:

$$\mathcal{L}(\Theta) = \mathcal{L}_{\text{tissue}} + \lambda_{\text{bg}}\mathcal{L}_{\text{bg}} + \lambda_{\text{reg}}\mathcal{L}_{\text{reg}}(\Theta). \tag{5}$$

The tissue reconstruction loss $\mathcal{L}_{\text{tissue}}$ is the sum of squared errors between predicted gene expression $\hat{g}(\cdot)$ and true gene expression g over all observed cells across the trajectory:

$$\mathcal{L}_{\text{tissue}} = \sum_{t=0}^{T} \sum_{(x_i^t, g_i^t)\in X^t} \|\hat{g}(x_i^t, z^t, t; \Theta) - g_i^t\|^2. \tag{6}$$

To constrain the overall shape of the embryo, we enforce zero expression in background regions. Let $\mathcal{G}$ be the set of discrete grid coordinates; we define the background set $\tilde{X}^t \subset \mathcal{G}$ as grid points that do not align with the registered spatial footprint of the tissue sample X^t. The background loss is:

$$\mathcal{L}_{\text{bg}} = \sum_{t=0}^{T} \sum_{u \in \tilde{X}^t} \|(M(z^t))_u\|^2. \tag{7}$$

$\mathcal{L}_{\text{reg}}(\Theta)$ denotes the regularization (see Appendix A.3 for further details). We minimize $\mathcal{L}(\Theta)$ with backpropagation in time using the Adam optimizerKingma & Ba (2015) with a learning rate of 1e-4, β_1 = 0.9, and β_2 = 0.999.

3.6 Evaluation

We evaluate our learned parameters by whether they can reconstruct held-out (masked) data. From the spatiotemporal time-series $X = (X^1, X^2, ..., X^N)$, we consider two masking schemes:

1. Within a given timepoint t_m, we remove a set (or multiple sets) of spatially contiguous coordinates $\mathcal{M} = \{(i, j)\}$ to obtain

   $$X^{t_m}_{\setminus \mathcal{M}} = \{X^{t_m}_{i,j} \mid (i, j) \notin \mathcal{M}\}.$$

   The task is then to reconstruct the missing region(s) $X^{t_m}_{\mathcal{M}}$.
2. We remove an entire sample X^{t_m} from the sequence and train the model on the remaining data,

   $$X_{\setminus t_m} = \{X^t \mid t \neq t_m\}.$$

   The goal is to reconstruct X^{t_m} given the dynamics inferred from $X_{\setminus t_m}$.

We compare the predicted and true gene expression profiles in these masked areas/samples using mean squared error (MSE), Pearson Correlation Coefficient (PCC), and Wasserstein distance.

4 Results

4.1 Datasets

We evaluate DALI on two spatial transcriptomic time-series datasets.

- The first is of zebrafish development Liu et al. (2022) at somitogenesis stages (10 hours post fertilization (hpf), 12 hpf, 18 hpf, and 24 hpf). We evaluate reconstruction of the 18 hpf sample with masking schemes 1 and 2, testing the ability of the model to interpolate the held-out data.
- The second dataset was a subset of the mouse development dataset Chen et al. (2022) at embryonic stages E10.5 and E12.5. Here, we evaluate whether held-out regions from the E12.5 timepoint can be reconstructed. This task moves beyond interpolation towards extrapolation because there is no data for held-out regions at later timepoints, or for driving registration parameters.

4.2 Baselines

We established a baseline (termed "superposition") that measures the extent to which spatial gene expression patterns at a given timepoint can be explained by affinely transforming and overlaying patterns from other timepoints. Given a held-out timepoint (or a sample with held-out regions), we fit an autoencoder with architecture matched to the mixing function M. We affinely register and overlay data from other timepoints in the latent space (Appendix A.4). This baseline is stringent because it requires optimizing registration parameters on held-out data.

For masking scheme 2 (held out timepoints), we additionally consider an optimal transport-based baseline using MOSCOT Klein et al. (2025). MOSCOT solves a Fused Gromov-Wasserstein type

problem to learn an optimal transport coupling between spatial gene expression data from consecutive timepoints. Given a held-out timepoint t_m, we compute an OT coupling for the pair (X^{t_m-1}, X^{t_m+1}) using MOSCOT's `SpatioTemporalProblem`. We then approximate cells at X^{t_m} as a linear interpolation from this coupling. We report the Wasserstein distance between the true and predicted gene expression. We additionally report the interpolated distance from a random OT coupling between (X^{t_m-1}, X^{t_m+1}) as well as "timepoint-distances" (i.e. the Wasserstein distance between (X^{t_m-1}, X^{t_m}) and (X^{t_m}, X^{t_m-1}).

For masking scheme 1 (held out regions of a timepoint), we further implement a Deep Image Prior baseline (Appendix A.4) Ulyanov et al. (2018). This evaluates whether the local spatial correlation structure inherent to each sample is sufficient to define gene expression patterns in the held-out regions, as would be expected with, for example, linear spatial gradients.

4.3 Transportation of spatial gene expression patterns across timepoints

To evaluate whether the trajectory learned by DALI correctly reconstructs gene-expression, we hold out either tissue regions or entire samples. On the zebrafish embryogenesis dataset Liu et al. (2022), we hold out a large section of the embryo from the 18 hpf sample. We perform multiple independent holdouts ("mask modes") where the model is trained to generate the 10 hpf, 12 hpf, the unmasked half of the 18 hpf (as indicated by grey regions in Figure 2A, lower panel), and the 24 hpf sample (Figure 2A). We do not apply any constraints on the reconstruction in the masked half (as indicated by red regions in Figure 2A, lower panel).

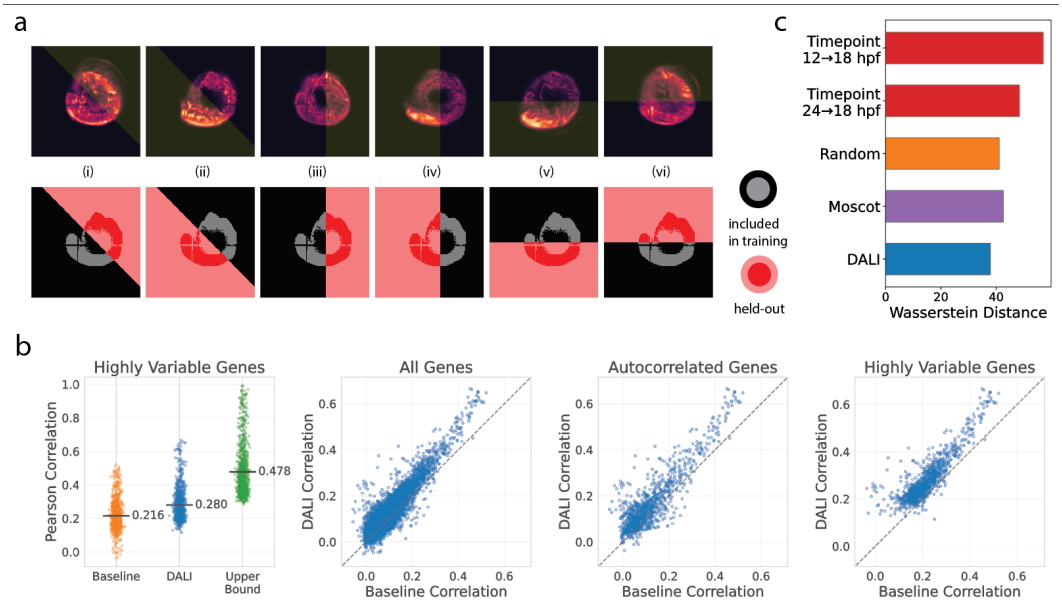

Figure 2: **(a)** DALI interpolation of 18hpf zebrafish sample across different masked regions. Top panel shows the sum of all latent activations i.e. $\sum_c z_{abc}$. **(b)** Pearson correlation (PCC) between true and predicted gene expression in the held-out region for masking mode (ii) for different gene subsets (Appendix A.1). Performance of superposition baseline ("baseline") and an autoencoder trained on the masked region ("upper bound") also shown. Bar indicates median PCC. **(c)** Wasserstein distance between true and predicted gene expression for various baselines and DALI interpolation when entire 18hpf zebrafish sample is held out.

We compute the Pearson Correlation Coefficient (PCC) between the true gene expression and the predicted gene expression within the masked section of the embryo. On mask mode (ii) for example, the DALI model exhibits higher median PCC than the superposition baseline when evaluated on all positions, i.e. both positions with gene expression and background positions (Figure 2B). Results for all mask modes are reported in Table S1, S2.

We hold out the entire 18 hpf sample and compute the Wasserstein distance between the true gene expression and predicted gene expression. We find that while DALI and MOSCOT both interpolate the held-out sample better than transporting spatial gene expression patterns from adjacent timepoints, DALI outperforms a random OT coupling while MOSCOT does not (Figure 2C).

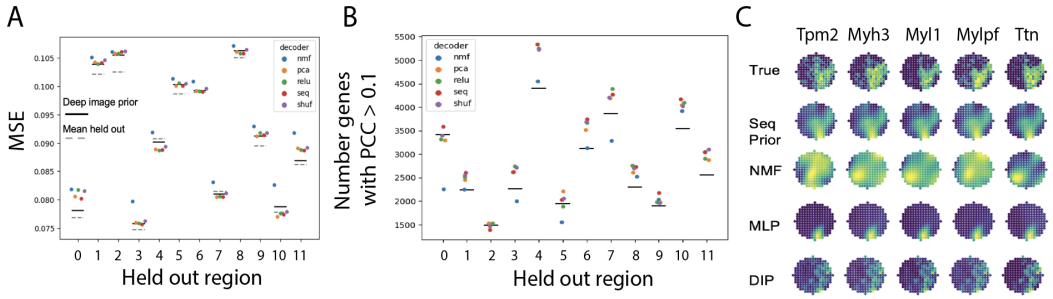

Figure 3: **(a)** MSE for reconstruction of gene expression across held out regions of E12.5 with different decoders (colors). NMF: non-negative linear decoder, relu: MLP decoder with ReLU nonlinearity, seq: sequence regularized MLP decoder, PCA: PCA regularized decoder, shuf: sequence regularized with features permuted across genes (see Appendix A.3). Solid line is Deep Image Prior, dashed line is mean gene expression in the held out region. **(b)** Number of genes with reconstructed/true expression PCC > 0.1 with different decoders. Solid line is Deep Image Prior **(c)** Qualitative visualization of top reconstructed genes in region 5. Reconstructions from different decoders as labeled on left, DIP is Deep Image Prior.

We next focused on the mouse spatiotemporal transcriptomics atlas Chen et al. (2022), held-out twelve regions from the E12.5 sample (Figure S2 and trained a model using the remaining portion of the E12.5 sample as well as the E10.5 sample. No constraints were applied to the positions that were held out from the E12.5 sample.

DALI using a MLP decoder with sequence prior regularization (A.3) outperformed the mean over held-out regions in three of the twelve regions (Figure 3). In addition, it outperformed the interpolation baseline (with deep image prior Ulyanov et al. (2018)) in seven of the twelve held-out regions (Figure 3A), indicating that the model uses information across time points and not only spatial correlations within the E12.5 timepoint. Applying only the learned affine registration (i.e., without integrating the PDE) led a greater MSE, indicating that the performance could not be explained by directly matching regions across timepoints (Table S5). While MSE differences were small, there was a more substantial difference between decoders and baselines in the number of genes with a PCC greater than 0.1. With respect to this metric, the NMF decoder models outperformed the deep image prior baseline in three of the twelve regions (Figure 3B). Variations of the MLP decoder outperform both the NMF decoder and deep image prior in all twelve regions. We further qualitatively confirmed that the genes best constructed captured non-linear patterns that were consistent with true gene expression patterns (Figure 3C). Thus, the DALI model explains nonlinear spatial gene expression variation for some genes in some of the regions by integrating information across timepoints.

4.4 DALI IDENTIFIES BIOLOGICALLY MEANINGFUL SPATIOTEMPORAL GENE EXPRESSION PROGRAMS

Using the optimal hyperparameters, we obtained a model trained on all timepoints for the zebrafish and evaluated the learned programs. To identify spatiotemporal gene expression programs (i.e. a spatiotemporally coordinated, functional set of genes), we attribute the contribution of each latent to reconstructing each gene with the attribution method DeepLIFT Shrikumar et al. (2017). We run Gene Set Enrichment Analysis Subramanian et al. (2005) on the resulting latent x gene matrices to interpret these programs. The spatial localization of programs was consistent with known biology as well as the programs obtained with the existing method Hotspot (Figure S4). For example a program enriched for myofibril assembly and muscle fiber development (FDR q-val < 0.05 in all timepoints) occurs in the myotome (tail muscle tissue) while a program enriched for regulation of

RNA metabolic processes occurs throughout the central nervous system (FDR q-val < 0.05 in all timepoints), consistent with previous reports Liu et al. (2022).

On the mouse data, we additionally identify high-magnitude interactions between gene expression programs. Since the learned reaction weights ($W^{(F;0)}$, $W^{(F;x)}$, $W^{(F;y)}$) are linear, they can be directly interpreted as interactions between programs. We identify interactions whose sign was consistent across multiple random initializations (Figure S6). The interactions recapitulated biological signalling pathways with previously established causal effects, for example:

- There was a positive edge between Programs 24 and 31. Program 24 enriched Wnt and VEGFR pathway genes, while Program 31 enriched targets of key Wnt effectors LEF1, CTNNB1 and GATA2.
- There was a negative edge between Program 31 (enriching YAP1 TF targets) and Program 26 (enriching TBX20 targets and heart genes); this is a canonical developmental relationship critical for heart development

Taken together, these results suggest the capability for DALI to learn biologically plausible cross-tissue programs and relationships between them.

5 Conclusion

Here, we introduce *Dynamical Analysis of Latent Interactions* (DALI), inspired by neural cellular automata and reaction-diffusion systems. DALI learns generative rules that recapitulate observed spatiotemporal gene expression trajectories. Our model exhibits capacity to reconstruct gene expression patterns in held out spatial regions and time points. We compare DALI to stringent baselines which account for simple, linear changes in spatial gene expression patterns across timepoints, as well as to existing methods based on optimal transport. We find that our model outperforms these baselines in both the zebrafish and mouse datasets. By interpreting the learned parameters, we find that DALI finds biologically plausible gene expression programs as well as interactions between them.

However, the DALI model is constrained by data and modeling assumptions. These can be attributed to: (i) limited temporal resolution of the data, (ii) limited sample size, and (iii) modeling 3D dynamics with biased 2D sections (in terms of cell-type composition). With feasible additional data, the LDS framework can be extended to address these limitations by: (i) increasing resolution through integrating measurements across modalities (e.g., by using modality-specific decoders to bridge high temporal and high molecular resolution data), (ii) modeling distributions over trajectories (e.g., by learning sample-specific initial conditions), and (iii) modeling 3D trajectories (e.g., by implementing the reaction-diffusion framework with 3D CNNs).

Our results also demonstrate that the spatiotemporal prior on the latent space qualitatively and quantitatively alters the gene expression programs that are learned. However, our specific choice of spatiotemporal dynamics assumes local interactions, which may not be sufficiently expressive to capture the true interactions driving development. Thus, exploring different choices of spatiotemporal priors to find one that optimally interpolates in sparse data regimes will reveal context-specific relationships between gene expression programs that drive development.

Ultimately, we envision DALI as a general and flexible framework for modeling spatiotemporal biological data, enabling the study of how local molecular interactions give rise to emergent tissue organization.

REFERENCES

Ao Chen, Sha Liao, Mengnan Cheng, Kailong Ma, Liang Wu, Yiwei Lai, and et al. Spatiotemporal transcriptomic atlas of mouse organogenesis using dna nanoball-patterned arrays. *Cell*, 185(10): 1777–1792.e21, 2022. doi: 10.1016/j.cell.2022.04.003.

Kalen Clifton, Manjari Anant, Gohta Aihara, Lyla Atta, Osagie K. Aimiuwu, Justus M. Kebschull, and et al. Stalign: Alignment of spatial transcriptomics data using diffeomorphic metric mapping. *Nat. Commun.*, 14:8123, 2023. doi: 10.1038/s41467-023-43915-7.

Hugo Dalla-Torre, Liam Gonzalez, Javier Mendoza-Revilla, Nicolas Lopez Carranza, Adam Henryk Grzywaczewski, Francesco Oteri, Christian Dallago, Evan Trop, Bernardo P de Almeida, Hassan Sirelkhatim, et al. Nucleotide transformer: building and evaluating robust foundation models for human genomics. *Nature Methods*, 22(2):287–297, 2025.

Martin Gardner. Mathematical games: The fantastic combinations of john conway's new solitaire game "life". *Scientific American*, 223(4):120–123, 1970. URL https://web.stanford.edu/class/sts145/Library/life.pdf.

Peter Halmos, Julian Gold, Xinhao Liu, and Benjamin J. Raphael. Learning latent trajectories in developmental time series with hidden-markov optimal transport. In *Research in Computational Molecular Biology (RECOMB 2025)*, volume 15333 of *Lecture Notes in Computer Science*, pp. 367–370. Springer, Cham, 2025a. doi: 10.1007/978-3-031-90252-9_41.

Peter Halmos, Xinhao Liu, Julian Gold, Feng Chen, Li Ding, and Benjamin J. Raphael. Destot: Alignment of spatiotemporal transcriptomics data. *Cell Syst.*, 16(2):101160, 2025b. doi: 10.1016/j.cels.2024.12.001.

Max Jaderberg, Karen Simonyan, Andrew Zisserman, and Koray Kavukcuoglu. Spatial transformer networks. In C. Cortes, N. D. Lawrence, D. D. Lee, M. Sugiyama, and R. Garnett (eds.), *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://proceedings.neurips.cc/paper_files/paper/2015/file/33ceb07bf4eeb3da587e268d663aba1a-Paper.pdf.

Diederik P. Kingma and Jimmy Ba. Adam: A method for stochastic optimization. *International Conference on Learning Representations*, 2015.

Dominik Klein, Giovanni Palla, Marius Lange, Michal Klein, Zoe Piran, Manuel Gander, and et al. Mapping cells through time and space with moscot. *Nature*, 2025. doi: 10.1038/s41586-024-08453-2.

M. V. Kuleshov, M. R. Jones, A. D. Rouillard, N. F. Fernandez, Q. Duan, Z. Wang, S. Koplev, S. L. Jenkins, K. M. Jagodnik, A. Lachmann, M. G. McDermott, C. D. Monteiro, G. W. Gundersen, and A. Ma'ayan. Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Research*, 44(W1):W90–W97, 2016. doi: 10.1093/nar/gkw377.

Chang Liu, Rui Li, Young Li, Xiumei Lin, Kaichen Zhao, Qun Liu, and et al. Spatiotemporal mapping of gene expression landscapes and developmental trajectories during zebrafish embryogenesis. *Dev. Cell*, 57(10):1284–1298.e5, 2022. doi: 10.1016/j.devcel.2022.04.009.

Alexander Mordvintsev, Ettore Randazzo, Eyvind Niklasson, and Michael Levin. Growing neural cellular automata: Differentiable self-organising morphogenesis. *Distill*, 2020. URL https://distill.pub/2020/growing-ca/.

Giovanni Palla, Hannah Spitzer, Michaela Klein, David S Fischer, Anna C Talirz Schaar, Lukas B Kuemmerle, Sergei Rybakov, Ignacio L Ibarra, Olof Holmberg, Isaac Virshup, et al. Squidpy: a scalable framework for spatial omics analysis. *Nature Methods*, 19(2):171–178, 2022.

Xiaojie Qiu, Daniel Y. Zhu, Yifan Lu, Jiajun Yao, Zehua Jing, Kyung Hoi Min, and et al. Spatiotemporal modeling of molecular holograms. *Cell*, 187(26):7351–7373.e61, 2024. doi: 10.1016/j.cell.2024.11.012.

Juan J. Sanz-Ezquerro, Andrea E. Münsterberg, and Sigmar Stricker. Editorial: Signaling pathways in embryonic development. *Frontiers in Cell and Developmental Biology*, 5:76, 2017. doi: 10.3389/fcell.2017.00076. URL https://arxiv.org/pdf/2510.02903.

Avanti Shrikumar, Peyton Greenside, and Anshul Kundaje. Learning important features through propagating activation differences. In *Proceedings of the 34th International Conference on Machine Learning (ICML)*, volume 70 of *Proceedings of Machine Learning Research*, pp. 3145–3153. PMLR, 2017. URL https://proceedings.mlr.press/v70/shrikumar17a.html.

A. Subramanian, P. Tamayo, V. K. Mootha, S. Mukherjee, B. L. Ebert, M. A. Gillette, A. Paulovich, S. L. Pomeroy, T. R. Golub, E. S. Lander, and J. P. Mesirov. Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences of the United States of America*, 102(43):15545–15550, 2005. doi: 10.1073/pnas.0506580102.

Christina V. Theodoris, Ling Xiao, Anant Chopra, Mark D. Chaffin, Zeina R. Al Sayed, Matthew C. Hill, and et al. Transfer learning enables predictions in network biology. *Nature*, 618:616–624, 2023. doi: 10.1038/s41586-023-06139-9.

Alexander Tong, Jessie Huang, Guy Wolf, David van Dijk, and Smita Krishnaswamy. Trajectorynet: A dynamic optimal transport network for modeling cellular dynamics. In *Proceedings of the 37th International Conference on Machine Learning*, volume 119 of *PMLR*, pp. 9526–9536, Vienna, Austria, 2020. PMLR. URL https://proceedings.mlr.press/v119/tong20a.html.

Alan M. Turing. The chemical basis of morphogenesis. *Phil. Trans. R. Soc. Lond. B*, 237:37–72, 1952. URL https://www.dna.caltech.edu/courses/cs191/paperscs191/turing.pdf.

Stanislaw Ulam. On some mathematical problems connected with patterns of growth of figures. In *Proceedings of Symposia on Applied Mathematics*, volume 14, pp. 215–224, Providence, 1962. American Mathematical Society.

Dmitry Ulyanov, Andrea Vedaldi, and Victor Lempitsky. Deep image prior. In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition (CVPR)*, June 2018.

Jan-Philipp von Bassewitz, Adeel Pervez, Marco Fumero, Matthew Robinson, Theofanis Karaletsos, and Francesco Locatello. Learning explicit single-cell dynamics using ode representations. *arXiv preprint*, arXiv:2510.02903, 2025. URL https://arxiv.org/abs/2510.02903.

John von Neumann and Arthur W. Burks. *Theory of Self-Reproducing Automata*. University of Illinois Press, Urbana, 1966.

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

A SUPPLEMENTARY NOTES

A.1 DATA PREPROCESSING

For the zebrafish dataset, read counts are normalized by total number of counts per spot (to normalize for sequencing depth) and log transformed. We perform an additional normalization, inspired from Theodoris et al. Theodoris et al. (2023) For each gene, we calculate its non-zero median expression value across all cells and normalize each gene expression value by this gene-specific factor. This has the effect of downweighting the expression value of genes that are widely and highly expressed housekeeping genes, while upweighting genes that are more lowly expressed but highly variable, such as transcription factors. We remove ribosomal and mitochondrial genes and further filter to genes expressed in at least 1% of all cells. After filtering, there are 8173 genes in the dataset.

To focus on biologically informative genes for certain analyses, the top 1000 spatially autocorrelated genes (autocorrelation defined by Moran's I score) in the 18 hpf sample are selected using the `gr.spatial_autcorr` function from the `squidpy` package Palla et al. (2022) with default parameters. The top 1000 highly variable genes are selected using the `pp.highly_variable_genes` function from the `scanpy` package Wolf et al. (2018) with default parameters.

For computational tractability, spot coordinates are additionally downsampled to a grid of predefined size. This size is a user hyperparameter and chosen as 100 for the experiments shown for the zebrafish dataset.

For the mouse dataset, log transformed gene expression was downloaded from Chen et al. Chen et al. (2022) and the model was fit directly to these measurements without downsampling spatial locations.

A.2 DECODER ARCHITECTURE

We consider three architectures for this decoder: a linear nonnegative decoder (termed 'NMF decoder'), a 3-layer MLP decoder with hidden dimensionality equal to the latent dimensionality, and a 3-layer MLP decoder of hidden dimensionality 200. The MLP decoders use either a ReLU or a CELU activation function (the CELU activation function was chosen so that the output is zero exactly when the input is zero).

A.3 REGULARIZATION

The regularization loss $\mathcal{L}_{\text{reg}}$ is the sum of various regularization terms on individual architectural components, as described in the following sections.

A.3.1 LATENT PDE REGULARIZATION

To encourage sparsity in the learned interactions, we add a L1 regularization term on the reaction weights F to the overall loss function. The weight of this penalty is a hyperparameter controlling the strength of PDE regularization.

A.3.2 DECODER REGULARIZATION

We evaluate whether regularizing the decoder encourages more biologically meaningful latent space for gene expression. Specifically, we consider:

- For the linear nonnegative decoder, we compute an orthogonality loss given by the average off-diagonal values of the pairwise cosine-similarity between rows of the decoder weight matrix.
- For MLP decoder, we add dropout.

The strength of the decoder regularization is a hyperparameter of the model.

We further evaluate incorporating prior information about each genes when given features $V_{\text{prior}} \in \mathbb{R}^{k \times G}$. Specifically, assuming that the weight of the final linear transformation of the decoder M

mapping the data into gene expression is a matrix $W \in \mathbb{R}^{m \times G}$, we learn an additional matrix $A : \mathbb{R}^m \times \mathbb{R}^k$ and bias $b \in \mathbb{R}^k$ to obtain a loss term

$$L_{\text{prior}}(W) := ||AW - V_{\text{prior}}||^2,$$

which encourages the model to learn a latent space that can be mapped to the prior features. This loss term acts as a regularization on the magnitude of the weights as well. We evaluate the value of the actual prior relationships by replacing V_{prior} with a permuted version. We separately consider and evaluate two kinds of prior information here. *Sequence prior regularization:* we use embeddings of the 1kb upstream of TSS using the nucleotide transformer foundation model (Dalla-Torre et al. (2025)); *PCA regularization:* we compute PCs across the entire dataset and use their loadings as gene features.

A.3.3 Registration regularization

Because we manually initialize the registration, we expect the learned registration parameters to be close to zero. We therefore add a L2 regularization term on the registration parameters to the overall loss function.

A.4 Baselines

A.4.1 Superposition Baseline

We obtain this baseline in two stages by fitting an autoencoder and then learning registrations and weights.

First, we fit an autoencoder to reconstruct each X^i in $X_{\setminus t_m}$. The decoder is constrained to have the same architecture and latent dimension as the decoder representing the mixing function M defined above. We learn affine registrations and weights to overlay other timepoints in a latent space such that when the decoder is applied, we reconstruct the held-out sample.

We freeze the encoder $E(\cdot)$ and decoder $D(\cdot)$, with encoder $E : \mathbb{R}^G \to \mathbb{R}^d$, and decoder $D : \mathbb{R}^d \to \mathbb{R}^G$.

We then learn a set of weights $\alpha = \{\alpha^t \mid t \neq t_m\}$ with $\sum_{t:t\neq t_m} \alpha_t = 1$, and a set of affine registrations that optimally reconstruct the held-out timepoint. The registration parameters are $(Q, b) = \{(Q^t, b^t) \mid t \neq t_m\}$. These parameters are optimized to reconstruct the held-out timepoint ($X^{t_m}_{\setminus \mathcal{M}}$ or X^{t_m}) in latent space as

$$\hat{Z}^{t_m} = \sum_{t \neq t_m} \alpha^t \big(Q^t E(X^t) + b^t\big),$$

which is subsequently decoded to pixel or gene-expression space:

$$\hat{X}^{t_m} = D(\hat{Z}^{t_m})$$

The registration parameters (Q, b) are optimized jointly with the combination weights α to minimize the reconstruction error over the masked regions in scheme 1 or the entire held-out sample in scheme 2:

$$\mathcal{L}_{\text{superpos}} = \big\|\hat{X}^{t_m}_{\setminus \mathcal{M}} - X^{t_m}_{\setminus \mathcal{M}}\big\|_2^2 + \lambda \big\|b^t\big\|_2^2,$$

or

$$\mathcal{L}_{\text{superpos}} = \big\|\hat{X}^{t_m} - X^{t_m}\big\|_2^2 + \lambda \big\|b^t\big\|_2^2.$$

After the baseline is trained, the evaluation metrics are then computed between $\hat{X}^{t_m}_{\mathcal{M}}$ and $X^{t_m}_{\mathcal{M}}$ (in masking scheme 1) or $\hat{X}^{t_m}$ and X^{t_m} (in masking scheme 2).

A.4.2 MOSCOT

For masking scheme 1, with entire held-out samples, the MOSCOT interpolation is obtained with the `compute_interpolated_distance` function, setting `source` as the 12 hpf sample, `intermediate` as the 18 hpf sample, and `target` as the 24 hpf sample. The interpolation parameter is set to 0.5 (i.e. linear interpolation). The interpolated cells are sampled as

weighted combinations of the source and target cells based on the coupling matrix and the interpolation parameter. The interpolation from a random OT coupling is similarly obtained using the `compute_random_distance` function, and the timepoint distances are obtained using the `compute_time_point_distances` function. Due to high computational complexity, MOSCOT is run by default on the top 30 principal components of gene expression. The Wasserstein distance for the DALI interpolation is computed by projecting the interpolated gene expression onto the principal components used by MOSCOT.

A.4.3 Deep Image Prior

We consider masking scheme 2, for a given timepoint, X^{t_m}, with masked regions $X^{t_m}_{\mathcal{M}}$. Following Ulyanov et al. (2018), we initialize a tensor with random uniform noise. We train a U-net with skip connections, composed with a pointwise decoder, to directly predict the gene expression $X^{t_m}_{\setminus\mathcal{M}}$ from this fixed input noise, and we restrict number of iterations during the optimization process. We then evaluate the capability of this U-net to reconstruct the masked regions $X^{t_m}_{\mathcal{M}}$ from the same input. Due to the inductive bias of convolutional neural nets, the Deep Image Prior learns natural image structure such as local spatial correlations. Therefore, this baseline explains the capability to reconstruct $X^{t_m}_{\mathcal{M}}$ simply due to local spatial patterns learned from $X^{t_m}_{\setminus\mathcal{M}}$. A sample defined by a linear spatial gradient, for example, would achieve high reconstruction performance with the Deep Image Prior.

B Supplementary Results

B.1 Ablations

We evaluated components of the architecture by testing the impact of removing them on the task of reconstruction held-out regions in both datasets (Table S3, S4, S6). Our findings were:

- Enforcing the PDE to learn denoising rules improved reconstruction error in both datasets.
- The more expressive MLP decoder consistently improved reconstruction in both datasets.
- We observed that including spatial gradient terms in the reaction improved reconstruction for the zebrafish but worsened it for the mouse dataset. Thus, including gradients may improve performance when the PDE must capture significant nonlinear pattern changes in shape.
- A timescale of around 10-20 steps was necessary to obtain successful reconstruction in both datasets.

B.2 Identified programs and interactions are biologically meaningful

For the mouse dataset, gene expression programs are interpreted from the model with the best performance on the reconstruction task. Eight of the forty programs for this model were consistently identified (spatial correlation after matching > 0.4) across three or more initializations. Aggregating across the previously defined cellular annotations, we saw that programs generally occurred across many tissues, but were enriched differently within them. Five programs were active predominantly in ectoderm derived tissue (brain, head and surface ectoderm), one program was active in endoderm tissues (highest in liver and primitive gut tube), and one was active in mesoderm tissues (heart and sclerotome) (Figure S5A-B).

We interpreted the programs with gene-set enrichment analysis Kuleshov et al. (2016), measuring enrichment of developmental disorder, transcription factor target, and signaling pathway gene sets among the top 200 genes spatially correlated with each program; these recapitulated expected tissue specific expression patterns (Figure S6A).

We identified high-magnitude interactions between programs from the learned reaction weights whose sign was consistent across initializations, plotting them as a graph (Figure S6D). The interactions recapitulated biological signalling pathways with previously established causal effects, as follows.

- A negative edge between Program 39 (enriching Notch and Smoothened pathways) and Program 31 (enriching the intrinsic apoptotic pathway).
- A positive edge between Program 14 (enriching targets of REST genes) and Program 31 (enriching the apoptosis pathway).
- There was an edge between Programs 24 and 31. Whereas program 24 enriched wnt and vegfr pathway genes, 31 enriched targets of key wnt effectors LEF1, CTNNB1 and GATA2.
- There was a negative edge between Program 31 (enriching YAP1 TF targets) and Program 26 (enriching TBX20 targets and heart genes); this is a canonical developmental relationship critical for heart development

B.3 DALI FINDS SPATIOTEMPORAL GENE EXPRESSION PROGRAMS CONSISTENTLY

We evaluated whether incorporating temporal dynamics affects the spatial programs that are learned on the zebrafish dataset. We fit a non-negative matrix factorization (NMF) model on the gene expression data from all the timepoints. We then train a DALI model with a non-negative linear decoder (Methods) that is initialized as the decoder from the pre-trained factor matrix from the NMF model. We then fit the DALI model to the data, allowing the model to both learn the PDE parameters and update the decoder weights. We then compare the programs learned by the DALI and NMF models.

Across random seeds, DALI consistently alters the intialization towards developmentally meaningful programs. We determined consistency by evaluating Pearson correlation over spatial locations as well as gene rankings between the NMF-intialized and DALI-fit decoder weight matrices (Figure S7,S8). While some programs (such as P1) are preserved with the NMF initialization, other programs such as P11 are reproducibly altered and compared to the initialization, exhibit greater coherence across time points (in terms of their spatial distribution relative to the head, tail and center) (Figure S9).

We additionally run gene set enrichment analysis Subramanian et al. (2005) on the difference between the initialized decoder and the fit decoder to identify gene pathways driving this alteration. We observe developmentally related gene sets that are reproducibly enriched in all 3 replicates, for example:

- In Program 3, "regulation of Wnt signaling" (Normalized enrichment score (NES) > 1.5 and FDR q-val < 0.15 in all 3 replicates)
- In Program 5, "skeletal myofibril assembly " (NES > 1.8 and FDR q-val < 0.031 in all 3 replicates)
- In Program 13, "positive regulation of MAPK cascade" (NES > 1.6 and FDR q-val < 0.016 in 2 of the 3 replicates)

We further evaluated the consistency of spatial program identification across model initializations using the mouse dataset. We matched programs across initializations using earth-mover distance to minimize spatial correlation distances averaged across time points. Following matching, models trained with the MLP decoder exhibited a spatial correlation of 0.6, indicating consistency of the spatial distribution of underlying programs (Figure S10A) that had been found.

We evaluated whether the learned dynamics were consistent, beyond only the learned programs. We first confirmed that the spatial locations in the earlier time point that contributed to gene expression in the held out regions were consistent across initializations. Given any gene in the later time point, we attribute its expression to spatial locations in the earlier timepoint by taking gradients with respect to each spatial location and latent dimension and taking the norm across the latent dimension. Applying this to the average expression of the top differentially expressed gene in each held-out region, we saw a high correlation in absolute gradients (considering only regions with high absolute gradient) (Figure S10B). As an example, the gradient for the top differentially expressed gene in held-out region 4 (Dbx4) is consistently high in the center of a patch at the the earlier timepoint (Figure S11B).

We further confirmed consistency of the learned dynamics by evaluating the learned reaction weights. Here, we evaluated the sign of the thresholded reaction weights between matched programs and observed that they agreed on average 55% of the time with all three decoders (Figure S10C).

Taken together, these results indicate that the DALI model identifies programs and interactions consistently, but requires aggregating across initializations to account for variability.

B.4 DALI IDENTIFIES SMOOTH DYNAMICS TRANSPORTING SPATIAL GENE EXPRESSION PATTERNS ACROSS TIMEPOINTS

We analyze the per-timepoint reconstruction loss as a function of the number of discrete PDE steps (i.e. "iterations") from the initial condition. Because each timepoint is associated with a specific iteration at which its loss is backpropagated, we expect its reconstruction error to reach a minimum near the iteration corresponding to that timepoint. We track how reconstruction loss evolves across iterations for both observed and held-out regions.

We perform this analysis on the zebrafish data under three settings: (i) no data held out (Fig. S12C), (ii) masking scheme 1 with held-out regions (Fig. S12A-B), and (iii) masking scheme 2, in which an entire timepoint is held out (Fig. S13C). Across all conditions, the learned PDE dynamics exhibit smooth transitions between neighboring timepoints. Moreover, the qualitative shape of these loss-dynamics curves is consistent across random initializations of the model, suggesting that the model learns a stable and coherent spatiotemporal trajectory.

C SUPPLEMENTARY TABLES/FIGURES

C.1 INTERPOLATION (SEE SECTIONS 4.3 AND B.1)

Mask mode	All positions			GEX positions		
	Baseline	DALI	Upper Bound	Baseline	DALI	Upper Bound
$x <$ centroid (vi)	-0.0102	**0.1331**	0.2383	0.0008	**0.0089**	0.1475
$x >$ centroid (v)	0.1122	**0.1177**	0.2361	**0.0207**	0.0106	0.1490
$y <$ centroid (iv)	0.0401	**0.1385**	0.2295	**0.0041**	0.0029	0.1407
$y >$ centroid (iii)	0.0505	**0.1342**	0.2471	0.0135	**0.0184**	0.1591
$y > x$ (ii)	0.0026	**0.1371**	0.2448	0.0020	**0.0094**	0.1562
$y < x$ (i)	0.0372	**0.1193**	0.2307	-0.0062	**0.0046**	0.1415

Table S1: Interpolation Results on Zebrafish for Masking Scheme 1. For each masked region ("mask mode", corresponding to the labels in Figure 2A, the mean Pearson Correlation Coefficient (PCC) is computed between true and predicted gene expression. Baseline=superposition, upper bound = autoencoder. Bold values indicate cases where the DALI model outperforms the baseline. Left: all positions within mask (dark + light red in Figure 2A); Right: Positions within mask with true gene expression i.e. "GEX positions" (dark red in Figure 2A).

Mask mode	All positions			GEX positions		
	Baseline	DALI	Upper Bound	Baseline	DALI	Upper Bound
$x <$ centroid (vi)	-0.0124	**0.1119**	0.2014	-0.0015	**0.0098**	0.1261
$x >$ centroid (v)	0.0913	**0.0994**	0.2012	**0.0175**	0.0101	0.1314
$y <$ centroid (iv)	0.0346	**0.1158**	0.1950	0.0012	**0.0017**	0.1232
$y >$ centroid (iii)	0.0372	**0.1138**	0.2097	0.0124	**0.0190**	0.1360
$y > x$ (ii)	-0.0001	**0.1152**	0.2072	-0.0040	**0.0091**	0.1328
$y < x$ (i)	0.0329	**0.1000**	0.1960	-0.0072	**0.0032**	0.1238

Table S2: Interpolation Results on Zebrafish for Masking Scheme 1. For each masked region (corresponding to the labels in Figure 2A, the median Pearson Correlation Coefficient (PCC) is computed between true and predicted gene expression. Baseline=superposition, upper bound = autoencoder. Bold values indicate cases where the DALI model outperforms the baseline. Left: all positions within mask; Right: GEX positions within mask only.

Mask mode	Baseline	DALI	DALI (No Diffusion)	DALI (No Grads)	DALI (Learnable Kernels)	Upper Bound
Mean PCC						
y>x	0.0026	**0.1371**	<u>0.1278</u>	0.0967	0.0966	0.2448
y<x	0.0372	<u>0.1193</u>	**0.1349**	0.1064	0.0905	0.2307
Median PCC						
y>x	-0.0001	**0.1152**	<u>0.1083</u>	0.0818	0.0818	0.2072
y<x	0.0329	<u>0.1000</u>	**0.1134**	0.0895	0.0768	0.1960

Table S3: Mean and median PCC for **All positions** in diagonal segment masks with ablations. Baseline = Superposition; Upper Bound = Autoencoder; DALI columns use the Latent PDE model (per ablation). DALI = full model, No Diffusion = removing the diffusion term in the update rule, No grads = removing spatial gradients in reaction input, Learnable Kernels = replacing fixed Laplace and Sobel kernels with free parameters. Bold values indicate, per mask mode, best performing model, underlined values indicate second-best performing model

Mask mode	Baseline	DALI	DALI (No Diffusion)	DALI (No Grads)	DALI (Learnable Kernels)	Upper Bound
Mean PCC						
y>x	0.0020	<u>0.0094</u>	0.0090	**0.0136**	-0.0017	0.1562
y<x	-0.0062	<u>0.0046</u>	**0.0136**	0.0059	-0.0058	0.1415
Median PCC						
y>x	-0.0040	0.0091	<u>0.0119</u>	**0.0129**	-0.0002	0.1328
y<x	-0.0072	0.0032	**0.0147**	<u>0.0059</u>	-0.0044	0.1238

Table S4: Mean and median PCC for **GEX positions** in diagonal segment masks with ablations. Baseline = Superposition; Upper Bound = Autoencoder; DALI columns use the Latent PDE model (per ablation). DALI = full model, No Diffusion = removing the diffusion term in the update rule, No grads = removing spatial gradients in reaction input, Learnable Kernels = replacing fixed Laplace and Sobel kernels with free parameters. Bold values indicate, per mask mode, best performing model, underlined values indicate second-best performing model

Table S5: Quantitative results for interpolating mouse held out regions using only registration (best scores taken across all trained registration models)

Region	MSE	# genes, r>0.1
0	0.105	2144
1	0.150	1495
2	0.170	1383
3	0.110	1634
4	0.115	3186
5	0.162	1617
6	0.147	1998
7	0.106	3196
8	0.138	1766
9	0.150	1614
10	0.148	2259
11	0.152	2826

Table S6: Hyperparameter selection for mouse dataset: average interpolation mean-squared error (MSE) across held-out regions in mouse task across three initializations of the model

use grad	latent dim	timescale	dropout	mix	MSE
False	40	25	True	seq	0.091483
False	40	20	True	seq	0.091524
False	40	50	True	relu	0.091630
False	40	20	True	relu	0.091734
False	40	30	True	relu	0.091743
False	40	15	True	nmf	0.093690
False	40	20	True	nmf	0.093980
False	40	30	True	nmf	0.094563
False	40	10	False	nmf	0.096699
False	40	15	False	nmf	0.097411
False	40	5	False	nmf	0.098356
True	40	10	False	nmf	0.098921
True	40	5	False	nmf	0.099746
False	40	0	False	nmf	0.123164
True	40	0	False	nmf	0.123471

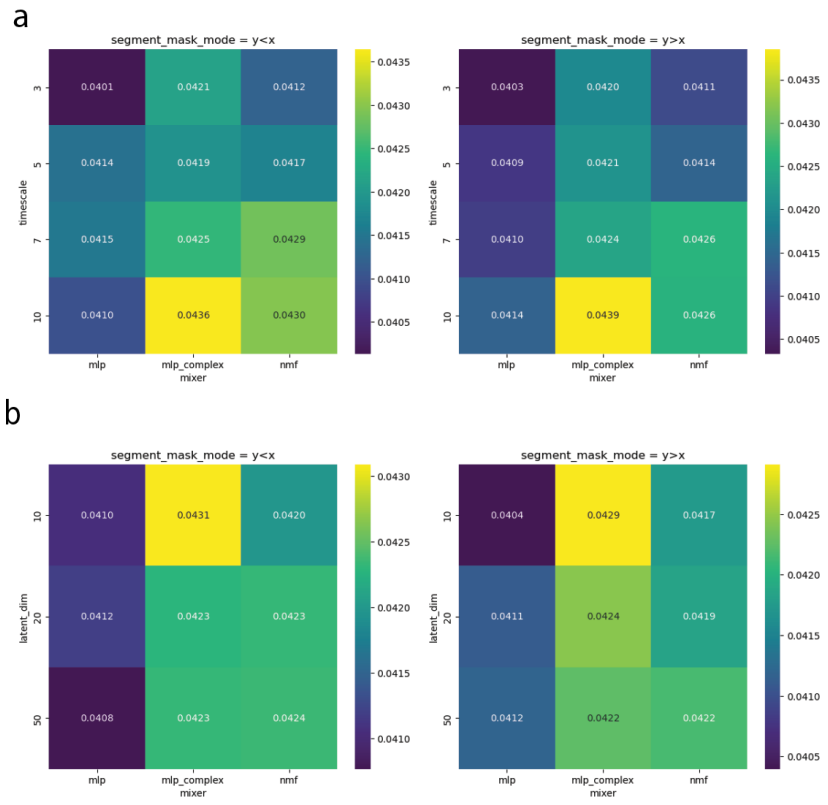

Figure S1: Hyperparameter selection for zebrafish dataset: average interpolation mean-squared error (MSE) for held-out regions in zebrafish task, averaged across three random initializations of the model

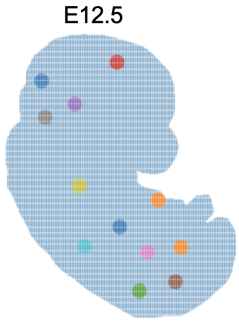

Figure S2: Held-out regions in the mouse E12.5 timepoint used for evaluating reconstruction

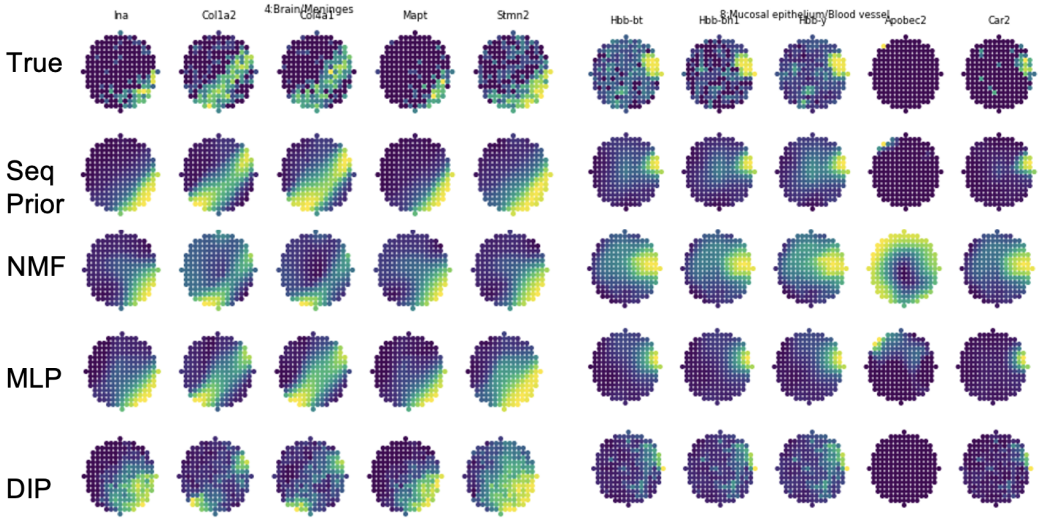

Figure S3: Qualitative visualization of top reconstructed genes in two held-out regions for the mouse reconstruction task

C.2 Interpreting Learned Programs (see Sections 4.4 and B.2)

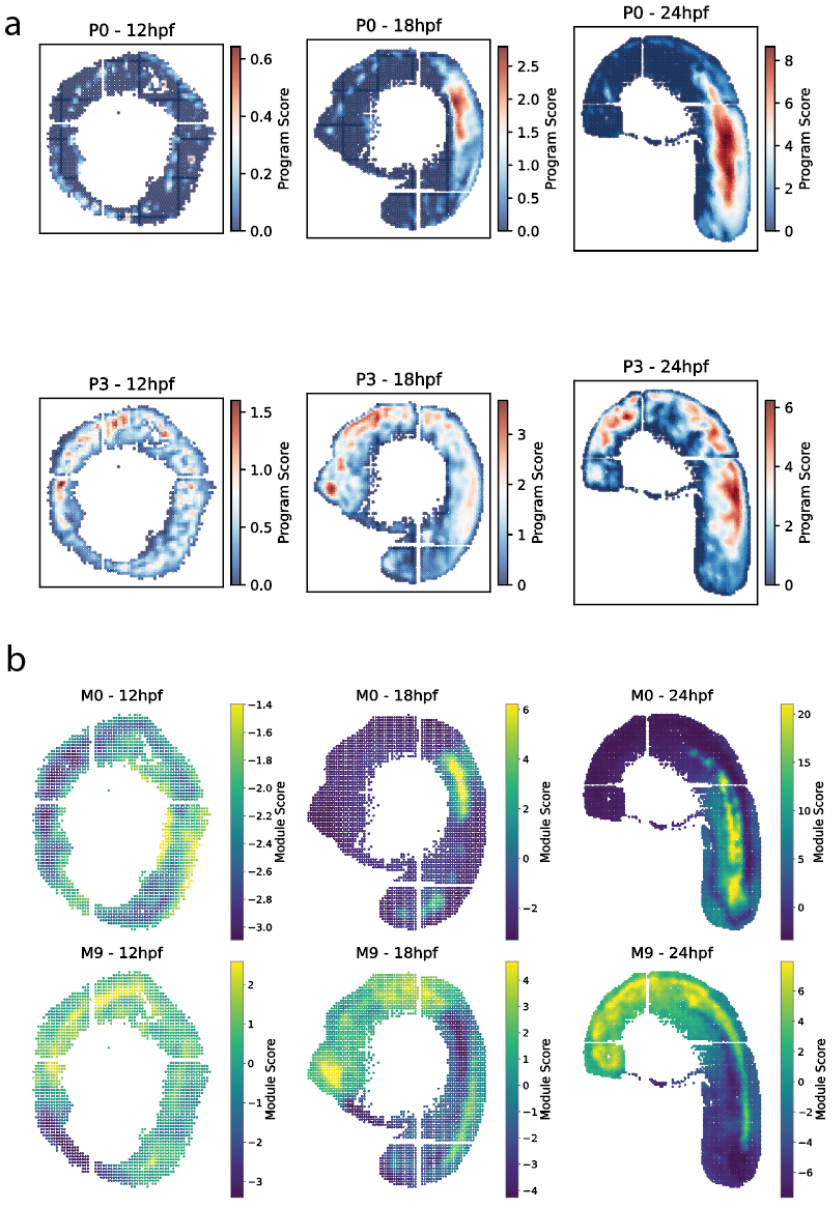

Figure S4: a. Visualization of representative programs learned by DALI by plotting spatial distribution of program activations b. Visualization of the matched spatial modules learned by Hotspot. Modules are matched by comparing the Spearman correlation of gene rankings between a given program/module.

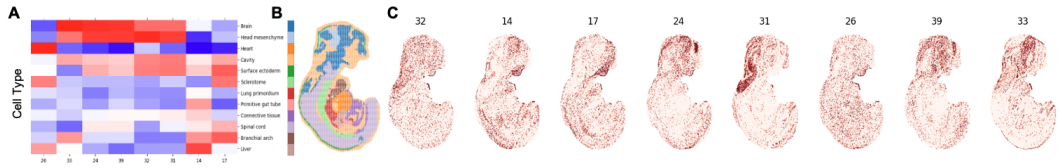

Figure S5: **(a)** Average activation of eight programs consistently identified across initializations within cell types from original publication. **(b)** Spatial distribution of cell types. **(c)** Spatial distribution of program activations.

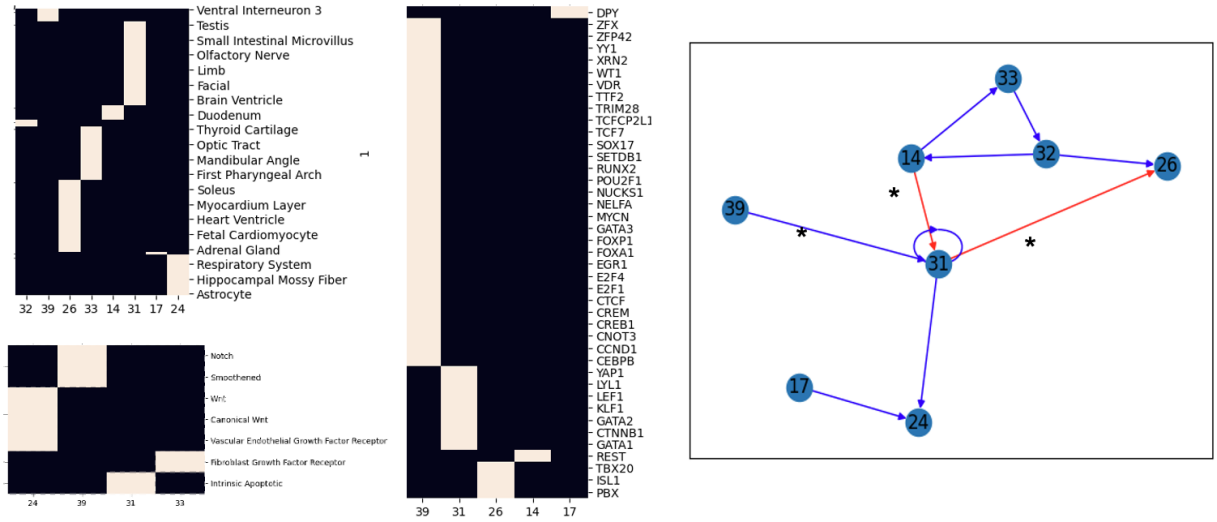

Figure S6: Interpretation of mouse programs. **(a-c)** Gene sets significantly enriched for each program; white: FDR<0.1; black: N.S. **(a)** Mouse Phenotype Ontology **(b)** Signaling pathways **(c)** TF-target gene sets. **(d)** Directed graph depicting a signed effect in learned reaction weight that is consistent across initializations (red is positive blue is negative). * indicates interactions with previous causal evidence.

C.3 Consistency of Program Identification (See B.3)

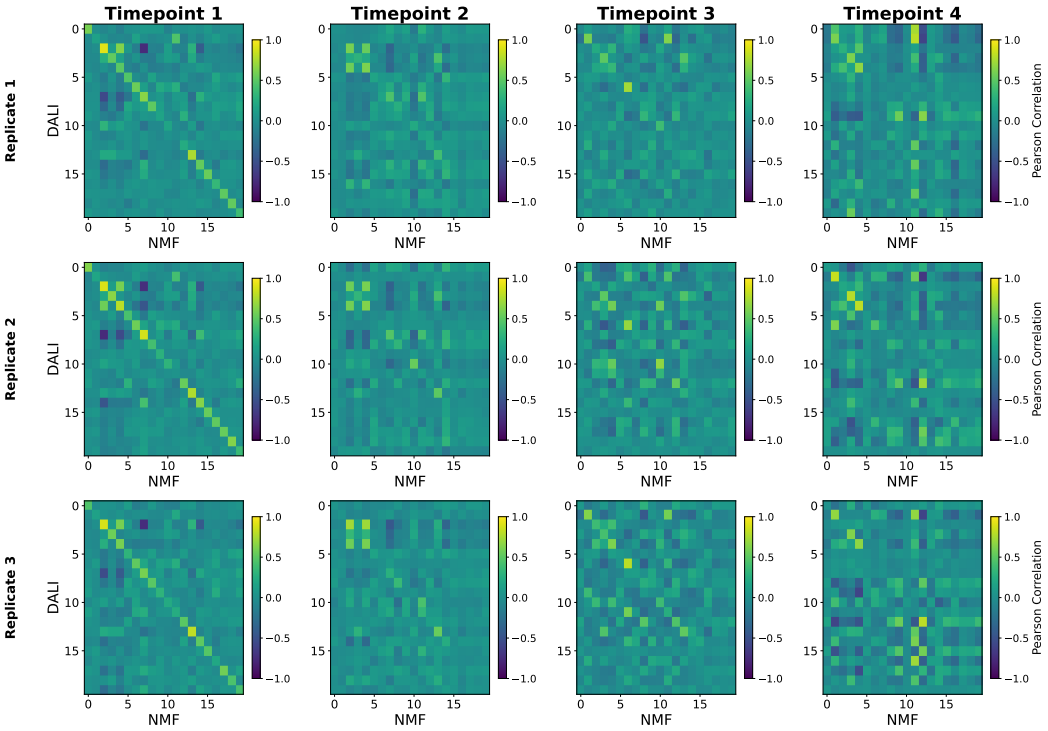

Figure S7: Spatial correlation of latent activations for NMF-initialized DALI vs NMF factors. Replicates correspond to independent random initializations of the PDE parameters in the DALI model.

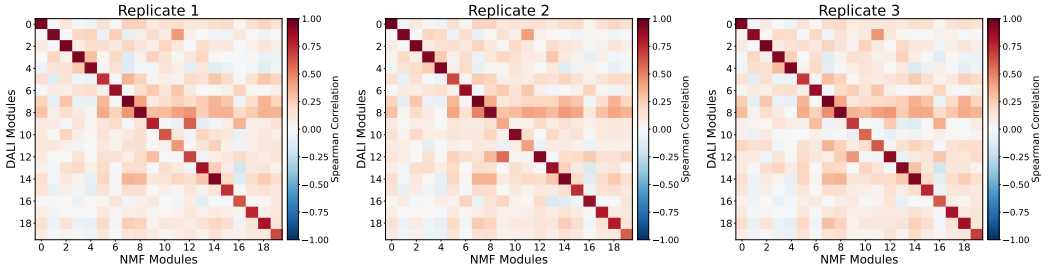

Figure S8: Spearman Correlation of gene rankings in the fit decoder from NMF-initialized DALI vs NMF factors. Replicates correspond to independent random initializations of the PDE parameters in the DALI model.

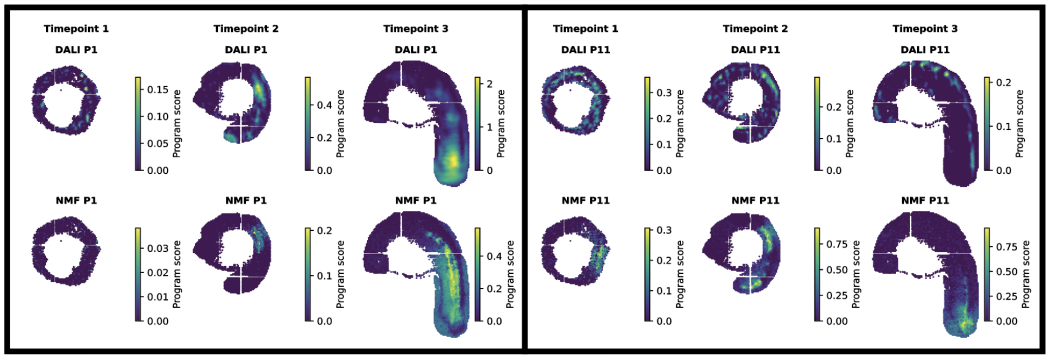

Figure S9: Program 1 (P1, left) and Program 11 (P11, right) obtained from fitting DALI after NMF initialization (upper) vs the same programs at initialization (lower). Timepoint 1 = 12 hpf, Timepoint 2 = 18 hpf, Timepoint 3 = 24 hpf. Program activations are plotted spatially for each timepoint.

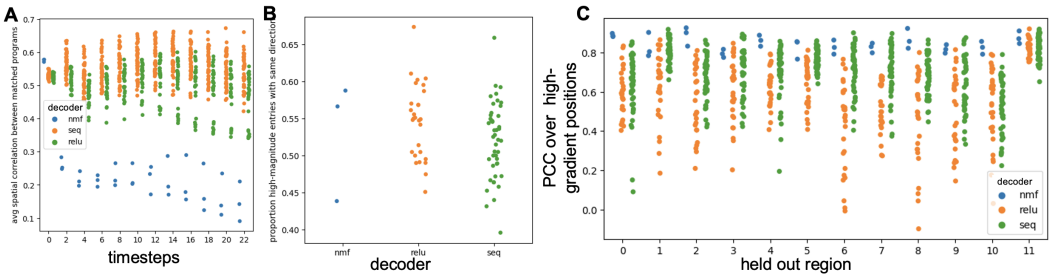

Figure S10: (a) Spatial correlation between matched program activation across initializations with respect to the three decoders. (b) Proportion of high-magnitude entries in the learned reaction weights that agree in sign across initializations. Random is 50%. (c) Spatial correlation across initializations between norm of gradient of average expression of top differentially expressed gene in held-out region at each spatial location in E10.5

.

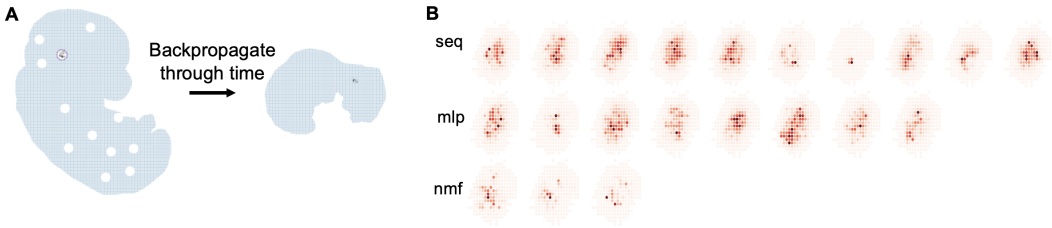

Figure S11: (a) Schematic illustrating taking gradients with respect to spatial locations at E10.5 using DALI. (b) Visualization of gradient norm for top differentially expressed gene in region 5 with respect to initializations (columns) ad models obtained with different decoders (rows)

C.4 Dynamics of Learned Spatiotemporal Trajectories (see Section B.4)

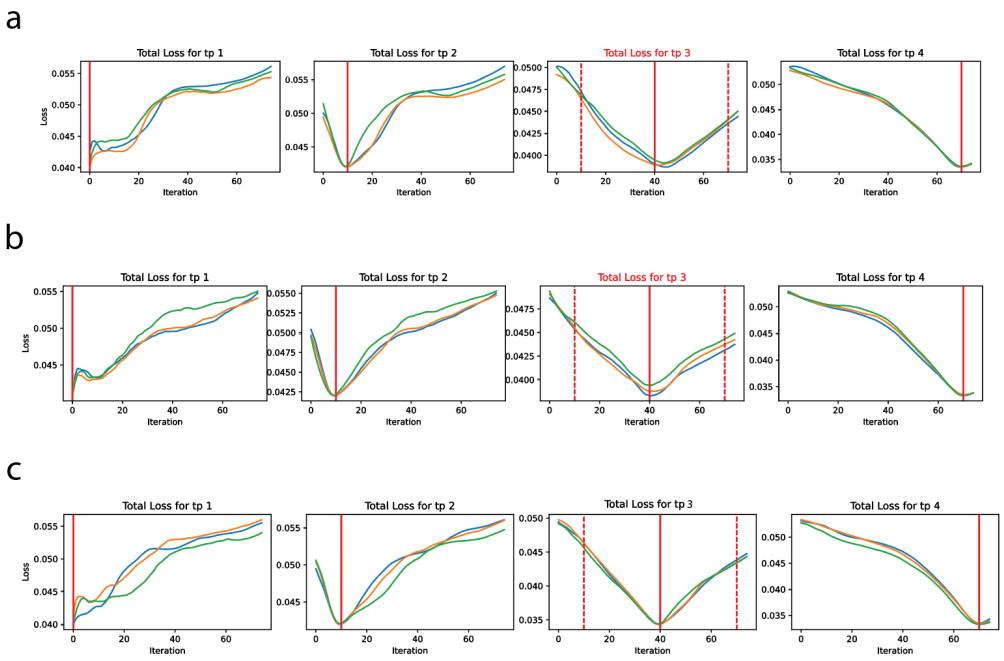

Figure S12: Loss with respect to each timepoint is plotted as a function of number of timesteps ("iterations") of the DALI model for three random initializations of the model (blue, orange, green lines). Vertical red line indicates at which iteration, the loss for that timepoint for observed data is backpropagated. Dashed red lines indicate the corresponding iteration for the previous and subsequent timepoint. a) $y > x$ mask mode. b) $y < x$ mask mode. c) No data held-out

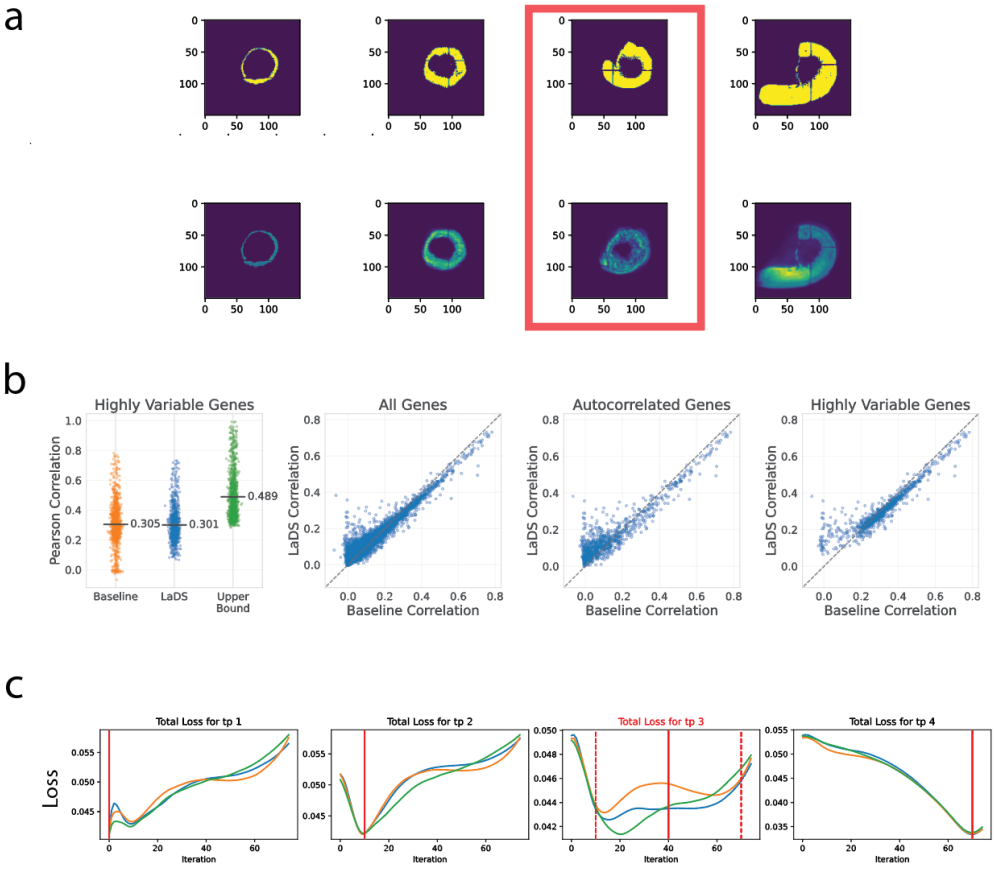

Figure S13: Whole Embryo interpolation where entire 18hpf sample is held out (corresponding to masking scheme 2, Methods). a) Qualitative reconstruction of all 4 samples. Held out sample denoted by red box. b) Quantitative evaluation of reconstruction of held out sample, compared to superposition baseline and autoencoder upper bound). Mean PCC on all genes for Baseline = 0.118, DALI = 0.126, Upper Bound = 0.218. Performance on top 1000 autocorrelated genes and top 1000 highly variable genes are shown. c) Reconstruction dynamics across three random initializations (blue, orange, green). Held out sample denoted in red. Loss with respect to each timepoint is plotted as a function of timesteps ("iterations") of the DALI model. Vertical red line indicates at which iteration, the loss for that timepoint is backpropagated (except for tp 3 = 18 hpf, which is held out). Dashed red lines indicate the corresponding iteration for the previous and subsequent timepoint.