SpaTRACE: Spatiotemporal recurrent auto-encoder for reconstructing signaling and regulatory networks from spatiotemporal transcriptomics data

Hongliang Zhou[1,*], Hao Chen[2,3,4,*,†], Zoe Rudnick[1,5], Sarah I. Baalbaki[1], Young Je Lee[1], and Jose Lugo-Martinez[1,†]

[1] Ray and Stephanie Lane Computational Biology Department, Carnegie Mellon University, Pittsburgh, PA 15213, USA
[2] Department of Computer Science, University of Illinois Chicago, Chicago, IL 60607, USA
[3] Center for Bioinformatics and Quantitative Biology, University of Illinois Chicago, Chicago, IL 60607, USA
[4] Richard and Loan Hill Department of Biomedical Engineering, University of Illinois Chicago, Chicago, IL 60607, USA
[5] Department of Electrical and Computer Engineering, Carnegie Mellon University, Pittsburgh, PA 15213, USA

[*]These authors contributed equally.
[†]Correspondence: chenhao@uic.edu, jlugomar@andrew.cmu.edu

Abstract. Cell–cell communication (CCC) drives the coordinated cellular dynamics underlying development, regeneration, and disease. Recent advances in spatiotemporal transcriptomics now enable measurement of gene expression within both spatial context and developmental progression. However, most existing CCC methods depend heavily on curated ligand–receptor (LR) databases and assume steady-state expression, limiting their applicability to understudied species and preventing robust inference of dynamic signaling cascades. Here, we introduce SpaTRACE, a transformer-based Granger-style recurrent autoencoder for *de novo* reconstruction of signaling and gene-regulatory interactions directly from developmental spatiotemporal transcriptomics data. SpaTRACE jointly embeds LR pairs, transcription factors (TFs), and target genes (TGs) into a unified latent space that integrates spatial neighborhoods with temporal causality. Trained to predict future TG expression along pseudotime-sampled trajectories, the model learns biologically meaningful embeddings that capture LR→TG signaling and TF→TG regulatory influence without relying on pathway priors. Across synthetic datasets, SpaTRACE accurately reconstructs LR–TG interactions, TF–TG regulation, and correct LR pairings, outperforming existing CCC tools—especially under pathway-agnostic settings. Applied to mouse midbrain development and axolotl brain regeneration, SpaTRACE recovers canonical signaling modules, identifies stage-specific transitions, and uncovers previously under-characterized interactions, producing CCC signals with high co-expression fractions and bivariate Moran's I comparable to leading spatial methods. Together, SpaTRACE establishes a general, statistically powerful framework for dissecting dynamic intercellular communication and regulatory networks from spatiotemporal transcriptomics. A user-friendly implementation is available at: https://github.com/VariaanZhou/SpaTRACE.git .

1 Introduction

Cell–cell communication (CCC) orchestrates the coordinated behaviors that underlie tissue development, homeostasis, and disease progression by transmitting molecular signals between neighboring cells. These signaling events—primarily mediated through ligand–receptor (LR) interactions—activate downstream transcriptional programs and gene-regulatory networks that shape cellular identity and function. With the rise of single-cell transcriptomics, numerous computational frameworks have emerged to infer intercellular signaling from transcriptomic measurements [1]. CellPhoneDB [2] uses permutation-based statistical testing over curated LR databases to identify significant sender–receiver interactions, whereas CellChat [3] integrates LR signaling with network-propagation models to quantify pathway activity and recover higher-order communication patterns among cell populations. Integrative approaches such as NicheNet [4] link ligands to downstream target genes (TG) via prior knowledge of pathway-based regulatory networks, and Connectome [5] further expands this idea by constructing global ligand–receptor–transcription factor (TF) networks.

In parallel, advances in spatial transcriptomics (ST) have enabled spatially explicit CCC models that incorporate cellular locations and tissue architecture. COMMOT [6] formulates intercellular communication as an optimal transport problem on spatial graphs, SpaTalk [7] integrates curated knowledge with spatial enrichment analysis to uncover signaling among neighboring cells, and CellChat v2.0 [3] leverages spatial proximity to refine LR-based inference. Together, these developments highlight a rapidly expanding methodological landscape for modeling CCC from single-cell and spatial transcriptomic data.

Recent advances in developmental and regenerative biology have introduced spatiotemporal transcriptomic datasets that measure spatially resolved gene expression across sequential developmental stages or timepoints [8,9]. Unlike static ST, these datasets capture both the cellular microenvironment and the temporal progression of cell states, providing a unique opportunity to reconstruct tissue-level trajectories as development unfolds. Despite this potential, most of the current CCC inference tools are not equipped to effectively leverage spatiotemporal information. First, their reliance on curated LR databases restricts applicability to well-characterized organisms and introduces knowledge biases arising from incomplete or inaccurate annotations. Second, most existing approaches assume a steady-state expression landscape, making them ill-suited to capture the dynamic or lineage-dependent communication processes that unfold during development. Third, these methods primarily focus on LR co-expression patterns while overlooking the downstream transcriptional responses that mediate the functional consequences of signaling, thereby preventing the reconstruction of temporally coherent LR–TF–TG signaling cascades. The truly *de novo*, pathway-independent inference of complete signaling and gene-regulatory networks directly from developmental spatial transcriptomic data remains largely unexplored.

Inspired by Velorama [10], which extends nonlinear Granger causality to developmental trajectories by leveraging DAG-structured pseudotime or RNA-velocity lineages to identify TFs whose lagged expression predicts downstream gene activity, and by STEAMBOAT [11], which applies a transformer-based architecture to infer CCC by learning LR interaction patterns directly from ST data, we introduce the Spatio-Temporal Recurrent Autoencoder for CCC inference (SpaTRACE). SpaTRACE is a transformer-based framework that identifies *de novo* gene–gene regulatory interactions by learning contextual embeddings of LR pairs, TFs, and TGs. This unified embedding architecture jointly models microenvironments and temporal dynamics, enabling the reconstruction of LR–TG and TF–TG signaling cascades without reliance on pathway priors, thus providing a flexible and unbiased foundation for uncovering novel communication mechanisms in complex tissues. The model demonstrates robust performance across synthetic datasets and faithfully reconstructs CCC and gene-regulatory networks in both mouse midbrain development and axolotl brain regeneration.

2 Methods

Our pipeline operates on single-cell–based developmental spatial transcriptomics datasets annotated with pseudotime and cell types. The central motivation is that when an ST dataset is sufficiently large and well-sampled across developmental stages, its lineage structure can be faithfully reconstructed and interpreted as the mean developmental trajectory of an underlying stochastic process. Building on this intuition, we developed a cell–cell communication (CCC) inference framework based on a self-supervised autoencoder architecture trained on sampled short-course cell trajectories. The model learns dynamic, context-aware rep-

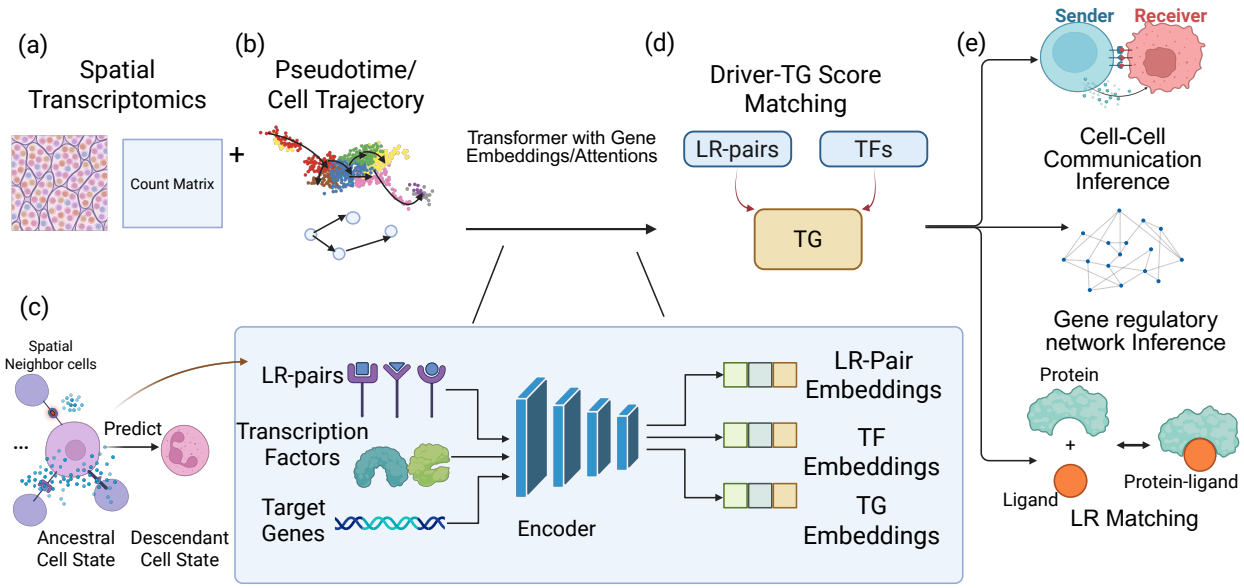


Fig. 1: Overview of the SpaTRACE framework. (a) Input spatial transcriptomics (ST) data annotated with (b) pseudotime information. (c) The model learns latent embeddings for each ligand–receptor (LR) pair, transcription factor (TF), and target gene (TG). (d) The learned entity embeddings are used to reconstruct driver–target relationships between LRs/TFs and TGs, enabling (e) downstream analyses such as cell–cell communication (CCC) inference, gene regulatory network (GRN) reconstruction, and LR pairing.

resentations for each ligand–receptor pair and gene, enabling the reconstruction of key gene–gene regulatory events from the learned embeddings and attention weights.

The SpaTRACE pipeline (Fig. 1) consists of four main stages. (1) Short cellular trajectories are sampled along the pseudotime axis, together with their spatial microenvironments, to capture local developmental dynamics while preserving temporal continuity (Fig. 1a–b, Section 2.2). (2) A recurrent autoencoder learns latent representations (entity embeddings) for driver signals—TFs and LR co-expression—and for TGs (Fig. 1c, Sections 2.3–2.4). (3) These entity embeddings are used to quantify each driver's predictive influence on its targets, generating candidate gene–gene regulatory hypotheses(Fig. 1d, Section 2.6). (4) Finally, the inferred driver–target relationships are integrated with spatial context to reconstruct CCC networks between spatially enriched sender and receiver cell types (Fig. 1e, Section 2.7).

2.1 Problem formulation

Let the dataset consist of a set of cells $\mathcal{C}$, genes $\mathcal{G}$, cell types $\mathcal{T}$, and batches $\mathcal{B}$. A spatial transcriptomics dataset is represented by a gene expression matrix $\mathbf{X} \in \mathbb{R}^{|\mathcal{C}|\times|\mathcal{G}|}$ and a pairwise spatial distance matrix $\mathbf{D} \in \mathbb{R}^{|\mathcal{C}|\times|\mathcal{C}|}$. Each cell $c \in \mathcal{C}$ is annotated with a cell type $\tau(c) \in \mathcal{T}$, a batch label $b(c) \in \mathcal{B}$, and a pseudotime value $\pi(c) \in [0, 1]$, indicating its relative position along a continuous developmental trajectory. The user provides predefined sets of ligands $\mathcal{G}_L$, receptors $\mathcal{G}_R$, and transcription factors $\mathcal{G}_{TF}$; all remaining genes are treated as target genes $\mathcal{G}_{TG}$ by default. Optional sender $\mathcal{T}_A \subseteq \mathcal{T}$ and receiver ($\mathcal{T}_B \subseteq \mathcal{T}$) cell types may also be specified; if omitted, all cell types serve as receivers, and sender types are inferred through spatial enrichment.

Cell Trajectory Representation We model cellular development as a temporal trajectory originating from an initial state $\mathbf{x}_0$ and evolving through intermediate states $(\mathbf{x}_0, \ldots, \mathbf{x}_t, \ldots, \mathbf{x}_T)$ toward a terminal state $\mathbf{x}_T$. For each focal cell c, the expression of its downstream target genes at the next time step, $\mathbf{x}^{TG}_{t+1,c} \in \mathbb{R}^{|\mathcal{G}_{TG}|}$, is assumed to be regulated by upstream molecular inputs within a finite temporal window $i = t - k, \ldots, t$.

To approximate this stochastic dynamical process, we employ a recurrent autoencoder conditioned on the current TG state $\mathbf{x}^{TG}_t$ and two classes of upstream regulatory drivers:

1. Transcription factor (TF) abundances, $\mathbf{x}^{TF}_{t-k:t} \in \mathbb{R}^{|\mathcal{G}_{TF}|}$: TF expression levels across the preceding time steps.
2. Local ligand–receptor (LR) co-expression, $\mathbf{\Lambda}^{LR}_{t-k:t} \in \mathbb{R}^{|\mathcal{G}_L \times \mathcal{G}_R|}$: computed from ligand signals secreted by spatial neighbors $\mathcal{N}_\theta(c) = \{\, c' \in \mathcal{C} : D(c, c') \leq \theta \,\}$, integrated with receptor abundance in the focal cell:

$$\Lambda^{l,r}_{t,c} = I^l_{t,c} \cdot x^r_{t,c}, \qquad I^l_{t,c} = \sum_{c' \in \mathcal{N}_\theta(c)} \frac{x^l_{t,c'}}{\theta^2}, \tag{1}$$

where $I^l_{t,c}$ denotes the local ligand intensity, computed as the aggregated ligand abundance among neighboring cells, normalised by the area θ^2 of the neighborhood radius. Here $x^l_{t,c'}$ is the ligand expression in neighbor c', and $x^r_{t,c}$ is the receptor expression in the focal cell. Conceptually, $\Lambda^{l,r}_{t,c}$ quantifies the signaling strength induced by ligand l available around cell c at time t, modulated by the receptor r's abundance to drive downstream gene regulation.

At each time point t, a cell c is represented by a concatenated multi-modal profile combining LR co-expression, TF expression, and TG expression: $\mathbf{x}^M_{t,c} = [\, \mathbf{\Lambda}^{LR}_{t,c} \,\|\, \mathbf{x}^{TF}_{t,c} \,\|\, \mathbf{x}^{TG}_{t,c} \,] \in \mathbb{R}^{|\mathcal{G}_L|\cdot|\mathcal{G}_R|+|\mathcal{G}_{TF}|+|\mathcal{G}_{TG}|}$. All molecular features are log-transformed and normalised across cells to ensure comparability across modalities and to mitigate batch effects. Each cell trajectory is compactly represented as a sequence of multi-modal profiles: $\mathcal{P} = \{\, \mathbf{x}^M_{t,c} \,\}^T_{t=0}$. For convenience, we use $\mathcal{G}_{LR}$ as shorthand for $\mathcal{G}_L \times \mathcal{G}_R$, and $\mathcal{G}_M$ for $\mathcal{G}_{LR} \cup \mathcal{G}_{TF} \cup \mathcal{G}_{TG}$.

2.2 Sampling cell trajectories from pseudotime

As the first step, cell trajectories are sampled along pseudotime $\pi(c)$ within the latent manifold by performing random walks on the kNN graph constructed in diffusion [12] or UMAP space. Using a fixed window size T, each random walk yields a sequence of distinct cells $(c_0, c_1, \ldots, c_T)$, which represent successive developmental states and are constrained to follow monotonically increasing pseudotime for biological consistency. Each trajectory defines a training sequence $\mathcal{P} = \{\, (\mathbf{x}^M_{t,c} \,\}^T_{t=0}$, where $\mathbf{x}^M_{t,c}$ denotes the normalized multi-modal profile of cell c at step t (Section 2.1). These sampled trajectories serve as discrete approximations of continuous developmental processes and form the basis for modeling spatiotemporal gene regulation.

2.3 Encoder framework

The encoder is designed to project multi-modal, cell-specific profiles into contextualized embeddings that capture both gene-level and cell-level dependencies. For each molecular modality, the model first introduces a global entity token (embedding) and then applies a self-attention layer to generate cell-specific embeddings for these entities. In parallel, the encoder derives a cell-level embedding for each modality, providing a compact summary of the intra- and inter-cellular context that enables effective cross-temporal attention between drivers and the downstream TGs.

Gene Token. Each molecular entity is assigned a global embedding token in a shared latent space $\mathbb{R}^d$. These global tokens represent static, cell-invariant descriptors that encode intrinsic regulatory roles independent of specific cellular contexts. Formally, we define the embedding matrices for LR, TF, and TG entities as $\mathbf{E}^{LR} \in \mathbb{R}^{|\mathcal{G}_{LR}|\times d}$, $\mathbf{E}^{TF} \in \mathbb{R}^{|\mathcal{G}_{TF}|\times d}$, $\mathbf{E}^{TG} \in \mathbb{R}^{|\mathcal{G}_{TG}|\times d}$. We denote the embedding vectors for specific entities by $\mathbf{e}^{l,r}$ for each LR pair, $\mathbf{e}^f$ for each transcription factor, and $\mathbf{e}^g$ for each target gene. Concatenating these matrices yields a unified global gene-token matrix $\mathbf{E} = [\, \mathbf{E}^{LR} \,\|\, \mathbf{E}^{TF} \,\|\, \mathbf{E}^{TG} \,] \in \mathbb{R}^{|\mathcal{G}_M|\times d}$,. This matrix defines the complete set of global gene tokens shared across all cells and temporal states in the model.

Gene Embedding. To capture context-dependent regulatory dynamics, we generate per-cell gene embeddings by modulating global gene tokens with expression-dependent weights derived from each cell's molecular profile. Let $\mathcal{M} = \{\mathrm{LR}, \mathrm{TF}, \mathrm{TG}\}$ denote the modular categories, and $\mathbf{E} \in \mathbb{R}^{|\mathcal{G}_M|\times d}$ the global gene-token matrix. For each category $m \in \mathcal{M}$, the per-cell gene embedding of cell c at time t is defined as

$$\mathbf{Q}^m_{t,c} = \mathrm{SelfAttn}\big(\phi_m(\mathbf{x}^M_{t,c}) \odot \mathbf{E}^m\big)\,, \tag{2}$$

where $\phi_m(\mathbf{x}^M_{t,c}) \in \mathbb{R}^{|\mathcal{G}_m|}$ provides modulation weights derived from the expression profile $\mathbf{x}^M_{t,c}$, and $\odot$ denotes elementwise product. Each modulation function maps the present cell profile to normalised attention weights:

$$\phi_m(\mathbf{x}^M) = \mathrm{softmax}\Big(W^m_{\mathrm{disc}}\sigma\big(W^m_2\sigma(\mathbf{V}^m_{\mathrm{val}} \odot \mathbf{x}^M) + \alpha^m \mathbf{V}^m_{\mathrm{val}} \odot \mathbf{x}^M\big)\Big), \tag{3}$$

where $\mathbf{V}^m_{\mathrm{val}} \in \mathbb{R}^{|\mathcal{G}_M|\times d}$ projects expression features into a lower-dimensional contextual space, serving as internal value embeddings for the entities, $W^m_2 \in \mathbb{R}^{d\times d}$ captures nonlinear interactions, $W^m_{\mathrm{disc}} \in \mathbb{R}^{|\mathcal{G}_M|\times d}$ produces attention logits, and $\alpha^m \in \mathbb{R}^{|\mathcal{G}_M|}$ are bias terms. Here, $\sigma(\cdot)$ denotes the LeakyReLU activation. The resulting weights ϕ_m rescale $\mathbf{E}^m$, after which the model applies a self-attention layer to obtain per-cell, context-specific embeddings $\mathbf{Q}^m_{t,c}$. Individual vectors $\mathbf{q}^{l,r}_{t,c}$, $\mathbf{q}^f_{t,c}$, and $\mathbf{q}^g_{t,c}$ represent the embeddings of $(l,r) \in \mathcal{G}_{LR}$ pairs, $f \in \mathcal{G}_{TF}$, and $g \in \mathcal{G}_{TG}$, respectively.

Cell Embedding. To summarize the global regulatory state of each cell, we define a cell embedding derived from its aggregated LR, TF, and TG expression profiles. Each modality-specific embedding is obtained by projecting the corresponding expression vector into the latent space via trainable linear mappings:

$$\mathbf{h}^{LR}_{t,c} = W_{LR}\,\mathbf{\Lambda}^{LR}_{t,c}, \qquad \mathbf{h}^{TF}_{t,c} = W_{TF}\,\mathbf{x}^{TF}_{t,c}, \qquad \mathbf{h}^{TG}_{t,c} = W_{TG}\,\mathbf{x}^{TG}_{t,c}, \tag{4}$$

where $W_{\mathrm{m}} \in \mathbb{R}^{d\times|\mathcal{G}_{\mathrm{m}}|}$ for each $\mathrm{m} \in \{\mathrm{LR}, \mathrm{TF}, \mathrm{TG}\}$. The resulting vectors $\mathbf{h}^{LR}_{t,c}$, $\mathbf{h}^{TF}_{t,c}$, and $\mathbf{h}^{TG}_{t,c}$ serve as compact latent representations summarizing the overall LR, TF, and TG activities of cell c.

2.4 Decoder framework

The decoder reconstructs the future target gene expression state $\mathbf{x}^{TG}_{t+1,c}$ from the current and ancestral molecular profiles $\mathbf{x}^M_{1:t,c}$ together with their corresponding cell embeddings. Inheriting the formulation of Velorama [10], we adopt an attention-based architecture grounded in the Granger causality principle, modelling the next cellular state as a weighted combination of past regulatory inputs. Unlike Velorama, where lagged contributions are encoded through predefined DAG powers, our model directly uses attention coefficients to quantify the predictive influence of historical signals. Furthermore, a multi-head prediction scheme disentangles LR- and TF-specific contributions under both global and cell-specific contexts, enhancing the interpretability and resolution of the inferred driver–target relationships at distinct context and stages.

Attention Mechanisms Two complementary attention mechanisms are incorporated: (i) temporal attention, which captures dependencies between past and present cell states; and (ii) molecular attention, which quantifies the relevance of driver entities—TFs and LR pairs—to TGs.

Temporal attention scores quantify how the regulatory state of a cell at earlier time steps influences its current state:

$$a^{C,m}_{t-i,t} = \langle \mathbf{h}^m_{t-i,c}, \mathbf{h}^{TG}_{t,c}\rangle, \qquad 0 \le i \le L, \;\; m \in \{\mathrm{TF}, \mathrm{LR}\}, \tag{5}$$

where L denotes the maximum look-back horizon in pseudotime. Molecular attention scores evaluate the regulatory compatibility between drivers and targets. Per-cell and global variants are defined as

$$a^{(m,\mathrm{percell})\to g}_{t-i,t,f} = \langle \mathbf{q}^f_{t-i,c}, \mathbf{q}^g_{t,c}\rangle, \qquad a^{(m,\mathrm{global})\to g}_{t-i,t,f} = \langle \mathbf{e}^f, \mathbf{e}^g\rangle, \tag{6}$$

where $\mathbf{q}$ and $\mathbf{e}$ denote per-cell and global embeddings, respectively; $g \in TG$ is the target gene, $m \in \{\mathrm{TF}, LR\}$ is the modality, and $f \in \mathcal{G}_m$ denotes a specific driver entity. Per-cell attention captures dynamic, cell-specific regulatory interactions across pseudotime, whereas global attention encodes baseline, context-invariant affinities.

Multi-Head Prediction The decoder models next-state TG expression as a weighted combination of past driver activities, where attention weights serve as Granger-style influence coefficients. To enhance interpretability and disentangle LR and TF effects, the model predicts TG expression separately for each modality ($m \in \{\mathrm{TF}, \mathrm{LR}\}$) and scope ($s \in \{\mathrm{global}, \mathrm{percell}\}$), while sharing a common latent representation. Each prediction head estimates the expression of TG g for cell c at time $t+1$ as

$$\hat{y}^{(m,s)\to g}_{t+1,c} = \sum_{i=0}^{L}\sum_{f\in\mathcal{F}_m} \sigma\big(a^{C,m}_{t-i,t}\, a^{(m,s)\to g}_{t-i,t,f}\big)\, x^{(m)}_{t-i,c,f}, \tag{7}$$

where σ applies a softmax over all (i, f) pairs to form a convex combination. Feature sets are defined as $\mathcal{F}_{\text{TF}} = \mathcal{G}_{TF}$ and $\mathcal{F}_{\text{LR}} = \mathcal{G}_{LR}$, with inputs $x^{(\text{TF})}_{t-i,c,f} = x^{f}_{t-i,c}$ and $x^{(\text{LR})}_{t-i,c,(l,r)} = \Lambda^{l,r}_{t-i,c}$.

Global and per-cell attention heads share temporal coefficients $a^{C,m}$, while molecular weights $a^{(m,s)\to tg}$ capture driver-specific effects within each modality and scope. This unified formulation integrates temporal and molecular attention to infer dynamic, context-dependent driver–target relationships along developmental trajectories, while maintaining efficient and interpretable embedding learning.

2.5 Model optimization

The input cells are randomly partitioned into training (S) and validation (V) sets. For each partition, the cell trajectories are sampled according to the procedure described in Section 2.2. Each trajectory $\mathcal{P} \in S$ consists of a sequence of ordered cells along pseudotime, sampled as described in Section 2.2.

The primary learning objective is the reconstruction loss of target gene expressions, with each head contribute equally. For each target gene $g \in \mathcal{G}_{\text{TG}}$, we minimize the mean squared error (MSE) between the observed expression $x^{g}_{t+1,c}$ and the model prediction from the four decoder heads is a weighted sum of each:

$$\mathcal{L}_{\text{pred}} = \frac{1}{|S|} \sum_{\mathcal{P}\in S} \frac{1}{|\mathcal{P}|} \sum_{(t,c)\in\mathcal{P}} \sum_{g\in\mathcal{G}_{\text{TG}}} \sum_{\substack{m\in\{\text{TF},\text{LR}\} \\ s\in\{\text{glob},\text{cell}\}}} w^{m}_{s} . \big(\hat{y}^{(m,s)\to g}_{t+1,c} - x^{g}_{t+1,c}\big)^2. \tag{8}$$

where w^{m}_{s} represent the contribution of each head. We further incorporate ℓ_1-norm regularization terms $\mathcal{L}^{\text{glob}}_{\text{att}}$ and $\mathcal{L}^{\text{cell}}_{\text{att}}$ on the global and per-cell attention weights, respectively, to encourage sparse and interpretable regulatory programs. The coefficients $\lambda^{\text{glob}}_{\text{att}}$ and $\lambda^{\text{cell}}_{\text{att}}$ balance their relative contributions. The total training objective, combining predictive accuracy with these regularization terms, is

$$\mathcal{L} = \mathcal{L}_{\text{pred}} + \lambda^{\text{glob}}_{\text{att}} \mathcal{L}^{\text{glob}}_{\text{att}} + \lambda^{\text{cell}}_{\text{att}} \mathcal{L}^{\text{cell}}_{\text{att}}. \tag{9}$$

The Noam learning rate scheduler [18] was employed during model optimization. It linearly increases the learning rate during an initial warm-up phase and subsequently decays it proportionally to the inverse square root of the training step. This scheduling strategy stabilizes early training, mitigates divergence, and enhances convergence in architectures that incorporate layer normalization and attention mechanisms. An EarlyStopping criterion was applied, and the best-performing model was restored based on validation loss.

2.6 Deciphering gene-gene interactions from learned embeddings and attention

After training, we extract interaction scores at two complementary levels: (i) attention-derived scores, which quantify the predictive contribution of LR or TF queries to TG keys across temporal lags; and (ii) embedding-derived scores, computed as cosine similarities between learned global embeddings. These gene-level measures are filtered to construct signaling and regulatory networks, which are subsequently aggregated to the cell-type level to yield interpretable summaries of putative cell–cell communication.

Gene–gene interactions. For each ligand–receptor pair $(l, r) \in \mathcal{G}_{LR}$ and target gene $g \in \mathcal{G}_{TG}$, we define

$$s_{(l,r)\to g} = \underset{t,c,i}{\text{Mean}} \big\langle \mathbf{q}^{l,r}_{t-i,c}, \mathbf{q}^{g}_{t,c} \big\rangle, \qquad \tilde{s}_{(l,r)\to g} = \big\langle \mathbf{e}^{l,r}, \mathbf{e}^{g} \big\rangle, \tag{10}$$

where $\mathbf{q}^{l,r}_{t-i,c}$ and $\mathbf{q}^{g}_{t,c}$ are the embeddings per-cell of the LR pair and the target gene, evaluated at lagged and current pseudotimes $t-i$ and t, respectively; $\mathbf{e}^{l,r}$ and $\mathbf{e}^{g}$ are their corresponding global tokens; and $\text{Mean}_{t,c,i}$ denotes the average over pseudotime, cells, and temporal lags.

To obtain an overall LR interaction strength independent of specific target genes, we take the filtered mean across all TGs:

$$s_{(l,r)} = \underset{g}{\text{Mean}} \big[s_{(l,r)\to g} \, \mathbf{1}\big(s_{(l,r)\to g} > \theta\big) \big] , \tag{11}$$

where $\mathbf{1}$ denotes the indicator function and θ is the filtering threshold default by $\theta = 0.5$ or defined by a percentile. Analogously, for TF–TG interactions, we define

$$s_{f\to g} = \underset{t,c,i}{\text{Mean}} \big\langle \mathbf{q}^{f}_{t-i,c}, \mathbf{q}^{g}_{t,c} \big\rangle, \qquad \tilde{s}_{f\to g} = \big\langle \mathbf{e}^{f}, \mathbf{e}^{tg} \big\rangle. \tag{12}$$

In our experiments, we adopt the mean operator for aggregation such that $s_\cdot$ quantifies the average effect of candidate drivers on downstream TGs. The resulting scores can be thresholded (or ranked) to identify putative regulatory links. In practice, embedding-derived scores ($\tilde{s}_\cdot$) provide a computationally efficient approximation, whereas attention-derived scores ($s_\cdot$) are more expressive, and could be restricted to cells beloning to specific batches $b \in \mathcal{B}$ to capture effects that vary across developmental stages. Unless otherwise specified, we report attention-derived scores throughout because of their greater expressivity.

2.7 Reconstruct cell–cell communication

To summarize these effects at the cell-type level, we assign ligands and receptors to cell types based on their expression profiles and then aggregate gene-level scores across spatially enriched sender–receiver neighborhoods. Let $\mathcal{C}_A = \{\, c \in \mathcal{C} \mid \tau(c) = A \,\}$ and $\mathcal{C}_B = \{\, c \in \mathcal{C} \mid \tau(c) = B \,\}$ denote the sets of cells belonging to sender type $A \in \mathcal{T}_A$ and receiver type $B \in \mathcal{T}_B$, respectively. For each receiver cell $c \in \mathcal{C}_B$, define $\mathcal{N}_A(c) \subseteq \mathcal{C}_A$ as the set of sender-type A cells within its spatial neighborhood. The communication strength from A to B mediated by ligand l and receptor r is given by

$$W_{A\to B}^{(l,r)} = \operatorname*{Mean}_{c\in\mathcal{C}_B}\left[\sum_{c'\in\mathcal{N}_A(c)} x_{c'}^{l}\, x_{c}^{r}\, s_{(l,r)}\right], \tag{13}$$

where $x_{c'}^{l}$ and x_{c}^{r} denote the expression levels of ligand l in the sender cell c' and receptor r in the receiver cell c, respectively. The operator $\mathrm{Mean}_{c\in\mathcal{C}_B}$ averages over all receiver cells.

This procedure yields a cell-type–level communication network, where each directed edge $A \to B$ can be decomposed into contributions from individual ligand–receptor pairs. The resulting network provides an interpretable estimate of the average influence of sender populations on receiver populations through ligand–receptor signaling. In practice, weak interactions and lowly expressed genes are filtered based on expression intensity or interaction strength to highlight the most salient communication pathways.

3 Results

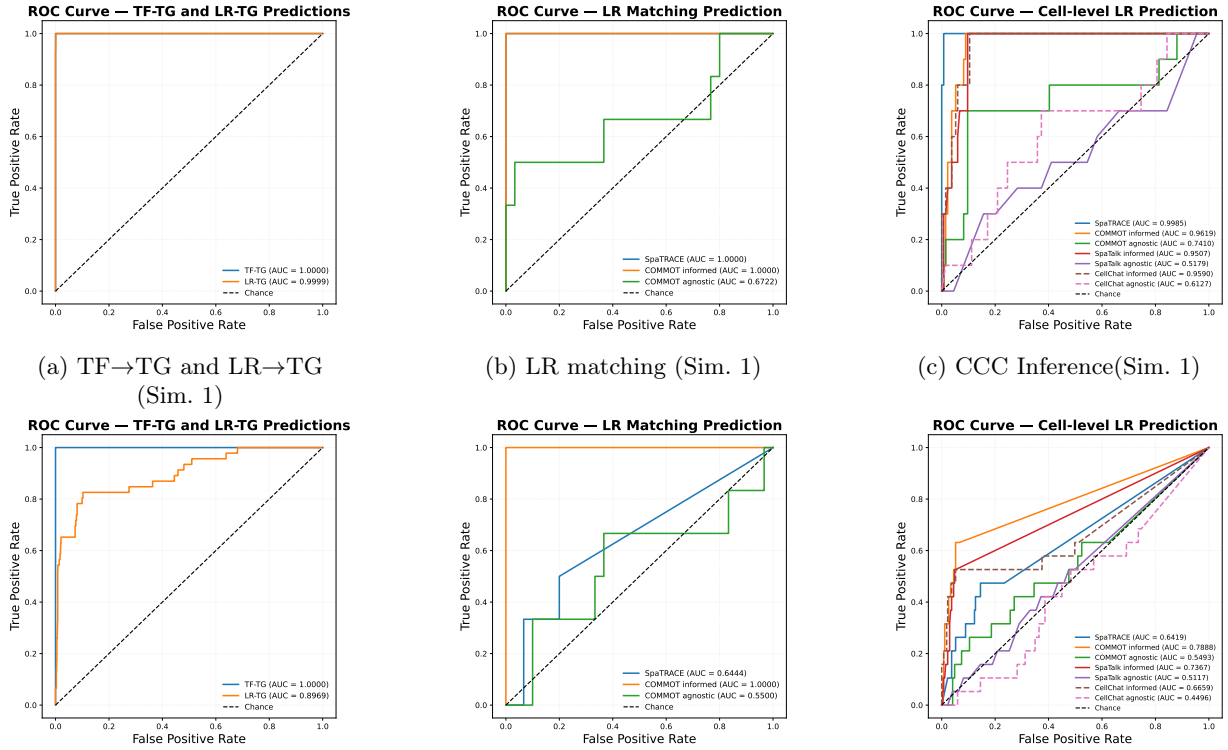
SpaTRACE was first quantitatively evaluated on two synthetic simulation datasets and subsequently applied to two biological systems—progenitor cell differentiation in the embryonic mouse midbrain (E12.5–E16.5) [8] and axolotl brain regeneration (DPI 0–60) [9]—for qualitative assessment. We then benchmarked SpaTRACE against three existing CCC inference tools, COMMOT, SpaTalk, and CellChat, using both simulated and real datasets. Specifically, we evaluated (i) the co-expression fractions (Appendix D.1) of identified ligand–receptor pairs across batches and (ii) the bivariate Moran's I statistic (Appendix D.2), as implemented in LIANA+ [13], to quantify the spatial autocorrelation of inferred ligand–receptor interactions. Additionally, we assessed thematic consistency by comparing enriched GO terms and signaling pathways (Appendix D.3), measuring the overlap between predicted pathways and progenitor differentially expressed genes (DEGs) using the Jaccard index.

3.1 Simulation datasets

To evaluate the model's capability and robustness, we used two synthetic datasets emulating distinct experimental conditions. Simulation 1 represents a low-noise, well-specified setting, whereas Simulation 2 introduces substantially higher biological and technical variability—including stochastic gene-expression noise, perturbed pseudotime trajectories, and weakened spatial coherence—mimicking artifacts commonly observed in real spatial transcriptomics. Together, these simulations enable assessment of model performance across both clean and artifact-rich regimes. Each dataset contained 2,400 cells and 64 genes (1,800 sender and 600 receiver cells), comprising six ligands, six receptors, six transcription factors, and 46 target genes. Details of the simulation design (cell differentiation trajectories, spatial distributions, and pseudotime assignments) are provided in Appendix A.

We evaluated the model across three tasks: LR→TG regulation prediction, LR-matching prediction, and CCC inference; for the CCC task, performance was measured by recovery of the correct sender–ligand–receptor

–receiver relationships. Ground-truth regulatory structures for each dataset appear in Figs. S2 and S3. All baseline methods were assessed under two conditions: (1) a pathway-informed setting using the true ligand–receptor pairs, and (2) a pathway-agnostic setting using a fully paired dummy LR library to remove informative priors (Appendix A.3). SpaTRACE does not rely on curated LR databases and reconstructs all regulatory relationships *de novo*. Because only COMMOT supports explicit LR-pairing, LR matching was compared solely against this method. For CCC inference, we additionally benchmarked against CellChat and SpaTALK.



(a) TF→TG and LR→TG (Sim. 1)

(b) LR matching (Sim. 1)

(c) CCC Inference(Sim. 1)

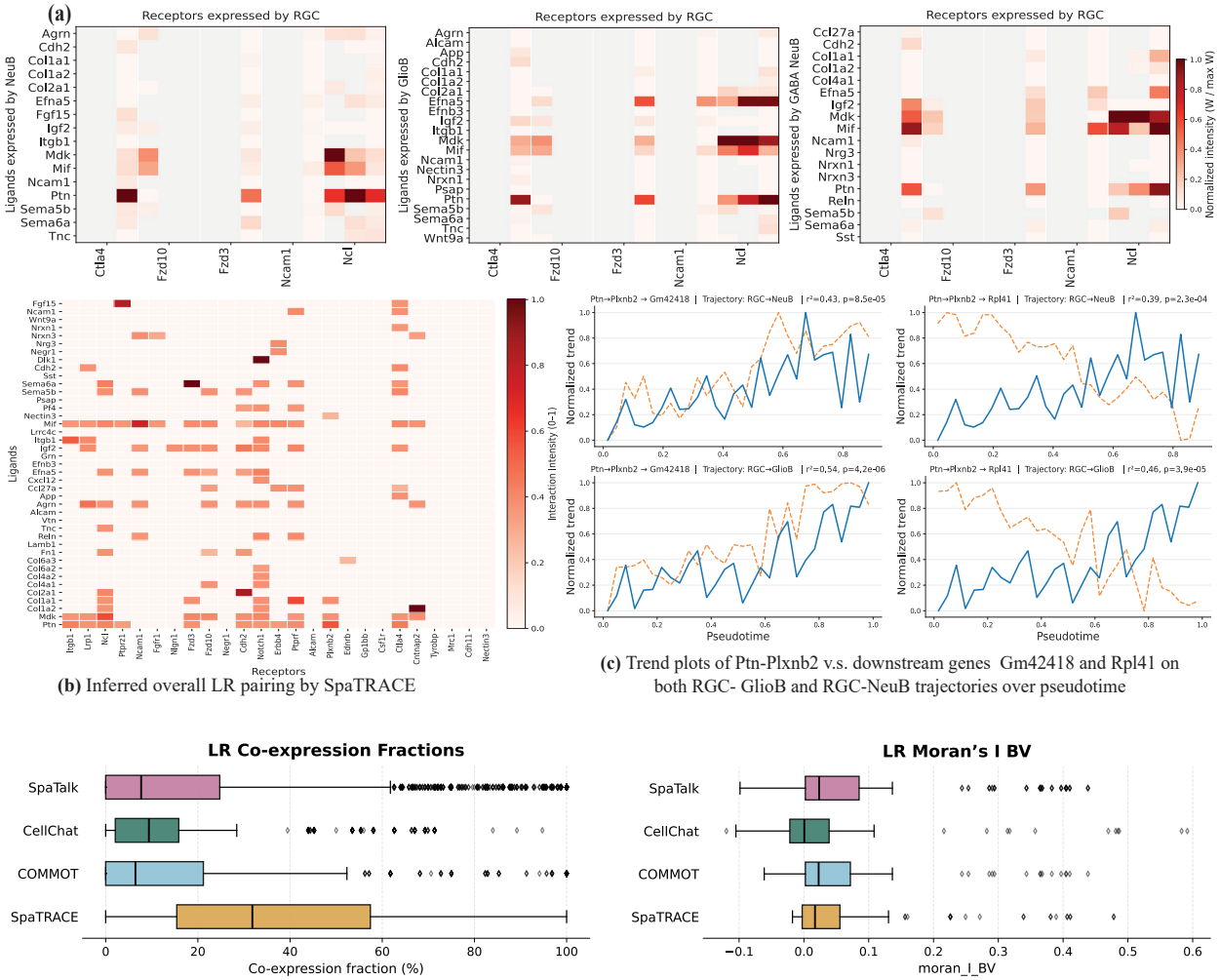
(d) TF→TG and LR→TG (Sim. 2)

(e) LR matching (Sim. 2)

(f) CCC Inference(Sim. 2)

Fig. 2: AUC–ROC curves for SpaTRACE and baseline methods on two simulated datasets. Panels (**a**–**c**) correspond to Simulation 1 and (**d**–**f**) to Simulation 2. (**a**,**d**) LR→TG and TF→TG prediction tasks; (**b**,**e**) LR matching results against COMMOT; and (**c**,**f**) compare SpaTRACE with benchmark approaches.

Figure 2 summarizes the ROC curves and AUC scores for SpaTRACE and the benchmark methods across the two simulated datasets. In the TF→TG and LR→TG tasks (2a and 2d), SpaTRACE accurately recovers regulatory relationships even under high-noise conditions. In the pathway-agnostic setting, both SpaTALK and CellChat failed to detect statistically significant LR interactions at conventional thresholds ($p_{\text{threshold}} = 0.5$). Consequently, their raw interaction scores were used directly as surrogate predictions, yielding near-random CCC performance (AUC ≈ 0.5). This behaviour is expected, as both methods depend substantially on curated LR databases and are not designed to infer LR mappings directly from data. COMMOT remained capable of producing significant interaction scores under the same setting, and these were used for evaluation. Across both simulation scenarios and all evaluation tasks, SpaTRACE consistently achieved the highest AUC score among models operating without pathway priors, and performed only marginally below these methods in the pathway-informed setting. Together, these results demonstrate that SpaTRACE reliably reconstructs multi-scale regulatory interactions and CCC patterns even in the complete absence of ligand–receptor annotations, whereas existing methods experience severe degradation when curated pathway information is removed.



(b) Inferred overall LR pairing by SpaTRACE

(c) Trend plots of Ptn-Plxnb2 v.s. downstream genes Gm42418 and Rpl41 on both RGC- GlioB and RGC-NeuB trajectories over pseudotime

(d) Distributions of co-expression fractions

(e) Distributions of bivariate Moran's I

Fig. 3: Integrated analysis of developmental cell–cell communication. (a) Spatially enriched sender–receiver CCC networks among NeuB, GlioB, and GABAergic NeuB to RGCs. (b) Overall inferred ligand–receptor pairings generated by SpaTRACE. (c) Expression trajectories of downstream *Ptn–Plxnb2* targets *Gm42418* and *Rpl41*. (d) Distributions of LR co-expression fractions for the top 100 inferred pairs across methods. (e) Distributions of bivariate Moran's I scores for the top 100 inferred LR pairs for each method.

3.2 Reconstructing cellular interactions during embryonic mouse midbrain development

We next applied SpaTRACE to the Mouse Organogenesis Spatiotemporal Transcriptomic Atlas (MOSTA) [8] to assess its ability to recover known developmental signaling patterns. Our analysis focused on midbrain progenitor cell differentiation, during which radial glial cells (RGCs) give rise to neuroblasts (NeuBs) and glioblasts (GlioBs) across embryonic stages E12.5–E16.5. Across 26,738 cells and 24,045 genes, we identified 7,167 progenitor cells and their spatially enriched neighbors, including GABAergic and glutamatergic neuroblasts (GABA NeuB and Glu NeuB) (Fig. S4). A detailed description of the data-processing, model training, and extended analyses is provided in Appendix B.1.

Our model recovered both canonical and previously uncharacterized LR interactions with strong developmental relevance. The LR matching inferred from the global tokens (Fig. 3b) indicates that *Mdk* and *Ptn* consistently paired with *Ncl*, *Lrp1*, *Itgb1*, and *Ptprz1*, all of which are well-established receptors or binding partners of the heparin-binding growth factor family [17,19,20,21,22]. As an illustrative example, we examined NeuB autocrine signaling during midbrain maturation (E12.5–E16.5). In the inferred NeuB autocrine

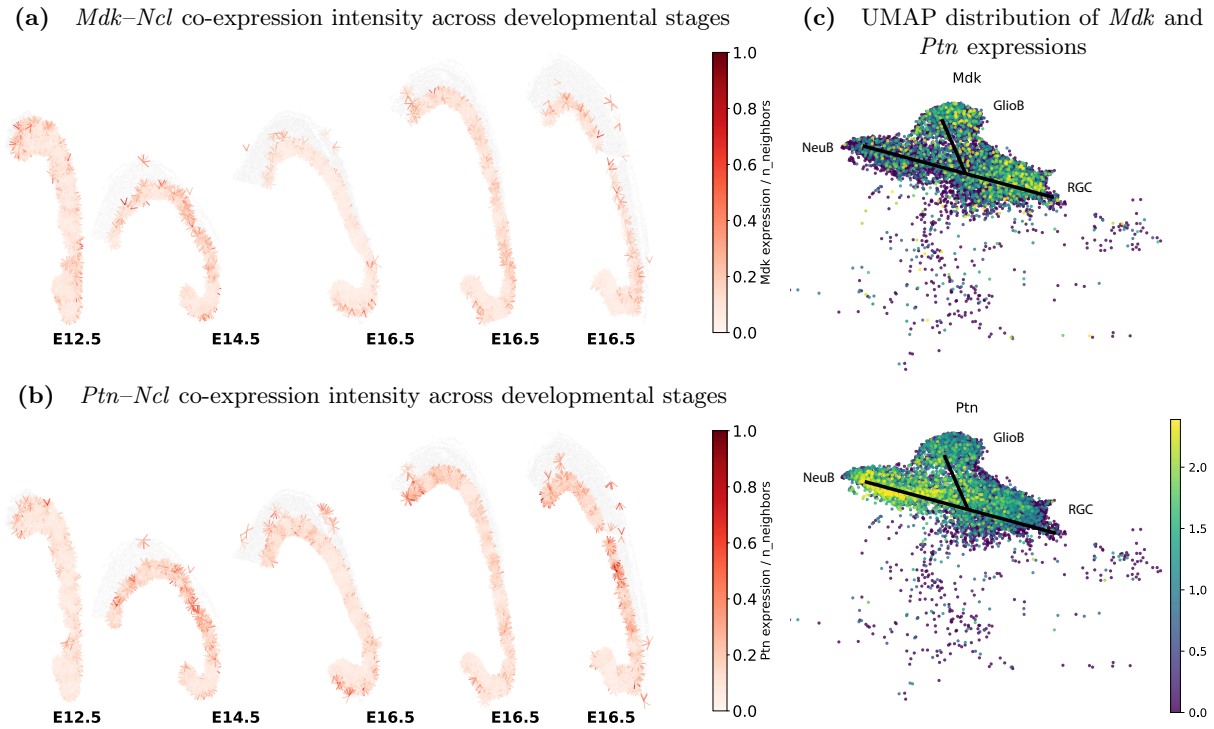


Fig. 4: Ligand–receptor co-expression patterns across mouse midbrain development. Each edge connects a ligand to its corresponding receptor, with edge intensity indicating the normalised co-expression level. (a) *Mdk–Ncl* co-expression across developmental stages (E12.5–E16.5). (b) *Ptn–Ncl* co-expression across developmental stages (E12.5–E16.5). (c) UMAP representation of *Ncl* and *Ptn* expression showing their opposing trends throughout development.

interactions (Fig. S5), *Mdk*-mediated signaling dominated at E12.5 but progressively transitioned toward *Ptn–Ncl* and *Ptn–Ptprz1* interactions by E16.5, indicating a developmental shift from *Mdk*- to *Ptn*-driven pathways. Notably, this *Mdk*-to-*Ptn* shift was not observed in the RGC receiving networks (Fig. 3a), where *Mdk–Ncl* interactions remained consistently elevated from *GlioB-RGC* and *GABA NeuB-RGC*, accompanied by an overall increase in *Ptn–Ncl* signaling. To further investigate this divergence, we visualized the spatial co-expression of *Mdk* and *Ptn* with *Ncl* (Fig. 4a–b), along with their UMAP distributions (Fig. 4c). Within NeuB cells, we observed a temporal increase in *Ptn–Ncl* co-expression concomitant with a decline in *Mdk–Ncl* co-expression. A similar progression—characterized by rising *Ptn* activity and decreasing *Mdk* activity along pseudotime—was also apparent across NeuB trajectories, further corroborating the model's predictions. Together, these findings reveal a temporal progression from proliferative (*Mdk*-driven) to differentiative (*Ptn*-driven) signaling, linking ligand–receptor dynamics to β-catenin–mediated transcriptional programs and illustrating the model's capacity to resolve stage-specific regulatory transitions.

In addition to this developmental shift, the model recovered several established signaling interactions, including *Dlk1–Notch1*, *Col2a1–Cdh2*, and homophilic *Itgb1–Itgb1* binding, underscoring its sensitivity to subtle yet biologically meaningful communication patterns. Beyond canonical pathways, SpaTRACE also identified underexplored interactions such as *Ptn–Plxnb2*, which has been previously reported but remains poorly characterized in the developing brain [21]. We correlated *Ptn–Plxnb2* co-expression with its two predicted downstream genes, *Gm42418* and *Rpl141*, and found consistent pseudotime trajectories and statistically significant correlations for both, further supporting the inferred interaction. Our approach additionally highlighted an *Agrn–Lrp1* association (Fig. 3c), consistent with STRING database predictions [24], as well as a potential *Fgf15–Ptprz1* coupling, which may reflect indirect regulatory relationships.

Quantitatively, the top 100 inferred CCC events inferred by SpaTRACE achieved bivariate Moran's I scores comparable to those of existing spatial inference tools, including COMMOT and SpaTalk, while

yielding a higher fraction of co-expressed ligand–receptor pairs (Figs. 3a and 3b). Together, these metrics demonstrate the model's ability to generate biologically meaningful and spatially coherent signaling hypotheses. To further assess the biological relevance of our predictions relative to known findings, we evaluated the thematic overlap (Appendix D.3) between the DEGs of each batch and the inferred LR→TG and TF→TG pairs at the corresponding developmental stage (Figs. S6a and S6b), using randomly selected pathways as a baseline. We additionally visualized the top 20 of the union of Gene Ontology (GO) enrichment terms derived from the inferred LR→TG and TF→TG pathways (Fig. S6c). Most of these enriched GO terms are related to neural development and cell differentiation, including "nervous system development" (GO:0007399) and synapse formation related activities (GO:0099068, GO:0097119 and GO:0032230), highlighting the thematic consistency between our inferred regulatory networks and established biological processes.

3.3 Reconstructing cellular interactions during axolotl brain regeneration

We applied a similar analysis pipeline to the axolotl brain regeneration dataset [9]. This dataset, generated using Stereo-seq technology, provides a high-resolution spatial atlas of axolotl brain regeneration across multiple time points—2, 5, 10, 15, 20, 30, and 60 days post-injury (DPI)—alongside an uninjured control group. According to the original study, the regenerative program remains active between 2 and 20 DPI.

We first performed spatial enrichment analysis to identify cell types preferentially localized within the neighborhoods of progenitor cells. To contrast regenerative versus homeostatic states, we selected, for each time point between 2 and 20 DPI, the union of all DE ligands, receptors, and TFs relative to the control group. In total, we retained 114 ligands, 120 receptors, and 117 TFs, along with 2,000 TGs of interest.

Because the axolotl is a relatively under-studied species lacking a well-curated ligand–receptor (LR) interaction database, benchmark methods could not be applied directly. We therefore focused on a qualitative evaluation of the inferred pathways. As an illustrative example, we examined communication between *IMN* and *nptxEX*, an intermediate cell state along the inferred differentiation trajectory, tracking changes in CCC patterns from 10 to 20 DPI. Chord plots of the top 30 interaction pairs (Fig. S8) revealed strong signals dominated by *VIM*-associated pathways at 10–15 DPI and by *PSAP*-associated pathways at 15–20 DPI. Our model also detected *GPI–ENO1* interactions, supported by functional evidence from the STRING database [24], as well as the pairing between *VIM* and *ENO1*, which has been reported to promote *VIM* binding to *RPSA* to form the *RPSA/VIM* complex in the human brain [23]. In addition to these known interactions, our analysis identified several putative signaling relationships not currently annotated in mouse or human LR databases. Although direct experimental evidence is not yet available, multiple studies report co-localization or co-expression of these gene pairs, suggesting that they may represent previously uncharacterized but biologically relevant interactions. Finally, co-expression and bivariate Moran's I analyses (Fig. S9) confirmed that our approach consistently produces a high co-expression fraction across the inferred pairs, further supporting the robustness and biological plausibility of the generated hypotheses.

4 Discussion

SpaTRACE is a spatiotemporal framework for inferring cell–cell communication (CCC) that integrates ligand activity with spatial context and temporal continuity. Using a Granger-style recurrent autoencoder with a shallow transformer layer, it captures key spatial–temporal dependencies while remaining interpretable. SpaTRACE identifies *de novo* signaling programs and extends beyond CCC to infer LR–TG and TF–TG regulatory relationships. Its pathway-free design also enables mechanistic hypothesis generation in species and tissues with limited annotation. However, several limitations remain. Representing ligand–receptor activity as the product of their expression implicitly assumes that ligand and receptor levels change synchronously. This ignores potential temporal lags between them and may conflate simple co-expression with true biochemical interaction. In addition, the absence of explicit receptor–TF coupling limits the model's ability to infer the entire LR-TF-TG cascade signaling pathway. A fixed spatial radius for neighborhood aggregation may also fail to capture variation in ligand diffusion range or local microenvironmental structure. Future extensions could also incorporate generative trajectory models, such as NicheFlow [25] or MOSCOT [26], to produce smoother microenvironmental transitions along sampled developmental paths, thereby yielding more biologically coherent dynamics and enhancing the robustness of CCC inference.

References

1. Dimitrov, Daniel, et al. "Comparison of methods and resources for cell-cell communication inference from single-cell RNA-Seq data." Nature communications 13.1 (2022): 3224.
2. Efremova, Mirjana, et al. "CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes." Nature protocols 15.4 (2020): 1484-1506.
3. Jin, Suoqin, Maksim V. Plikus, and Qing Nie. "CellChat for systematic analysis of cell–cell communication from single-cell transcriptomics." Nature protocols 20.1 (2025): 180-219.
4. Browaeys, Robin, Wouter Saelens, and Yvan Saeys. "NicheNet: modeling intercellular communication by linking ligands to target genes." Nature methods 17.2 (2020): 159-162.
5. Raredon, Micha Sam Brickman, et al. "Computation and visualization of cell–cell signaling topologies in single-cell systems data using Connectome." Scientific Reports 12.1 (2022): 4187.
6. Cang, Zixuan, et al. "Screening cell–cell communication in spatial transcriptomics via collective optimal transport." Nature methods 20.2 (2023): 218-228.
7. Shao, Xin, et al. "Knowledge-graph-based cell-cell communication inference for spatially resolved transcriptomic data with SpaTalk." Nature communications 13.1 (2022): 4429.
8. Chen, Ao, et al. "Spatiotemporal transcriptomic atlas of mouse organogenesis using DNA nanoball-patterned arrays." Cell 185.10 (2022): 1777-1792.
9. Wei, Xiaoyu et al. "Single-cell Stereo-seq reveals induced progenitor cells involved in axolotl brain regeneration." Science (New York, N.Y.) vol. 377,6610 (2022): eabp9444. doi:10.1126/science.abp9444
10. Singh, Rohit, et al. "Causal gene regulatory analysis with RNA velocity reveals an interplay between slow and fast transcription factors." Cell systems 15.5 (2024): 462-474.
11. Liang, Shaoheng, et al. "Steamboat: Attention-based multiscale delineation of cellular interactions in tissues." International Conference on Research in Computational Molecular Biology. Cham: Springer Nature Switzerland, 2025.
12. Coifman, Ronald R., et al. "Geometric diffusions as a tool for harmonic analysis and structure definition of data: Diffusion maps." Proceedings of the national academy of sciences 102.21 (2005): 7426-7431.
13. Dimitrov, Daniel, et al. "LIANA+ provides an all-in-one framework for cell–cell communication inference." Nature Cell Biology 26.9 (2024): 1613-1622.
14. Liang, Yuchen, et al. "PROST: quantitative identification of spatially variable genes and domain detection in spatial transcriptomics." Nature Communications 15.1 (2024): 600.
15. Trapnell, Cole, et al. "The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells." Nature biotechnology 32.4 (2014): 381-386.
16. Dibaeinia, Payam, and Saurabh Sinha. "SERGIO: a single-cell expression simulator guided by gene regulatory networks." Cell systems 11.3 (2020): 252-271.
17. Muramatsu, Takashi. "Midkine and pleiotrophin: two related proteins involved in development, survival, inflammation and tumorigenesis." The Journal of Biochemistry 132.3 (2002): 359-371.
18. Vaswani, Ashish, et al. "Attention is all you need." Advances in neural information processing systems 30 (2017).
19. Scott, Daniel D., and Marlene Oeffinger. "Nucleolin and nucleophosmin: nucleolar proteins with multiple functions in DNA repair." Biochemistry and cell biology 94.5 (2016): 419-432.
20. Maeda, Nobuaki, et al. "A receptor-like protein-tyrosine phosphatase PTPζ/RPTPβ binds a heparin-binding growth factor midkine: involvement of arginine 78 of midkine in the high affinity binding to PTPζ." *Journal of Biological Chemistry*, vol. 274, no. 18, 1999, pp. 12474–12479.
21. Papadimitriou, Evangelia, et al. "On the role of pleiotrophin and its receptors in development and angiogenesis." The International Journal of Developmental Biology 66.1-2-3 (2021): 115-124.
22. Kadomatsu, Kenji, Satoshi Kishida, and Shoma Tsubota. "The heparin-binding growth factor midkine: the biological activities and candidate receptors." The journal of biochemistry 153.6 (2013): 511-521.
23. Jiang, Hexiang, et al. "Enolase of Streptococcus suis serotype 2 promotes biomolecular condensation of ribosomal protein SA for HBMECs apoptosis." BMC biology 22.1 (2024): 33.
24. Szklarczyk, Damian, et al. "The STRING database in 2025: protein networks with directionality of regulation." Nucleic Acids Research 53.D1 (2025): D730-D737.
25. Sakalyan, Kristiyan, et al. "Modeling microenvironment trajectories on spatial transcriptomics with nicheflow." arXiv preprint arXiv:2511.00977 (2025).
26. Klein, Dominik, et al. "Mapping cells through time and space with moscot." Nature 638.8052 (2025): 1065-1075.

A Appendix A: Simulation Datasets

A.1 Synthetic Data Construction

To quantitatively evaluate the model's performance, we constructed simulation datasets. In particular, scRNA data corresponding to a predefined GRN are generated by SERGIO[16] with both receptors and transcription factors acting as causal regulators of the target genes (TGs). To simulate cell-cell communication in space, we assign spatial coordinates such that cells of the same type are spatially clustered. In particular, a square grid of $n_{\text{spots}} \times n_{\text{spot}}$ positions is first generated, and a Gaussian Process (GP) prior is sampled over this grid to yield smooth latent energies $\phi_{i,c}$ for each spot i and cell type c. These are transformed into probabilities via a softmax with temperature T to regulate the centrality of the resulting distribution:

$$\pi_{i,c} = \frac{\exp(\phi_{i,c}/T)}{\sum_{c'} \exp(\phi_{i,c'}/T)}. \tag{14}$$

Each spot then draws a cell type according to $\pi_{i,c}$, and a actual receptor cell of that type is assigned to the corresponding spatial coordinate. This procedure produces receptor cells that are spatially clustered and correlated across the grid, rather than being scattered independently.

Cell–Cell Communication Simulation To simulate ligand-receptor interactions in tissue, we further generate sender cells, and we treat the original receptor expression values in the receiver cells as the hidden ligand-receptor interaction values Λ. For every receptor of type r, the algorithm places n sender cells of type s within a radius δ disk centered at the receptor's location, whenever $(s \to r)$ is present in the SR set. A sender's coordinate (x_s, y_s) is sampled as

$$(x_s, y_s) = (x_r, y_r) + \rho(\cos\theta,\ \sin\theta), \quad \theta \sim \mathcal{U}(0, 2\pi), \ \ \rho \sim \delta\sqrt{U}, \ \ U \sim \mathcal{U}(0, 1). \tag{15}$$

Ligand and receptor expression are equalized by setting both to a shared activity level proportional to the receiver's LRA signal, with small added independent Gaussian noises:

$$R_{i,k} \ \leftarrow \ \sqrt{R_{i,k}} \ + \ \epsilon_r, \qquad L_{j,k} \ \leftarrow \ \sqrt{R_{i,k}} \ + \ \epsilon_l, \tag{16}$$

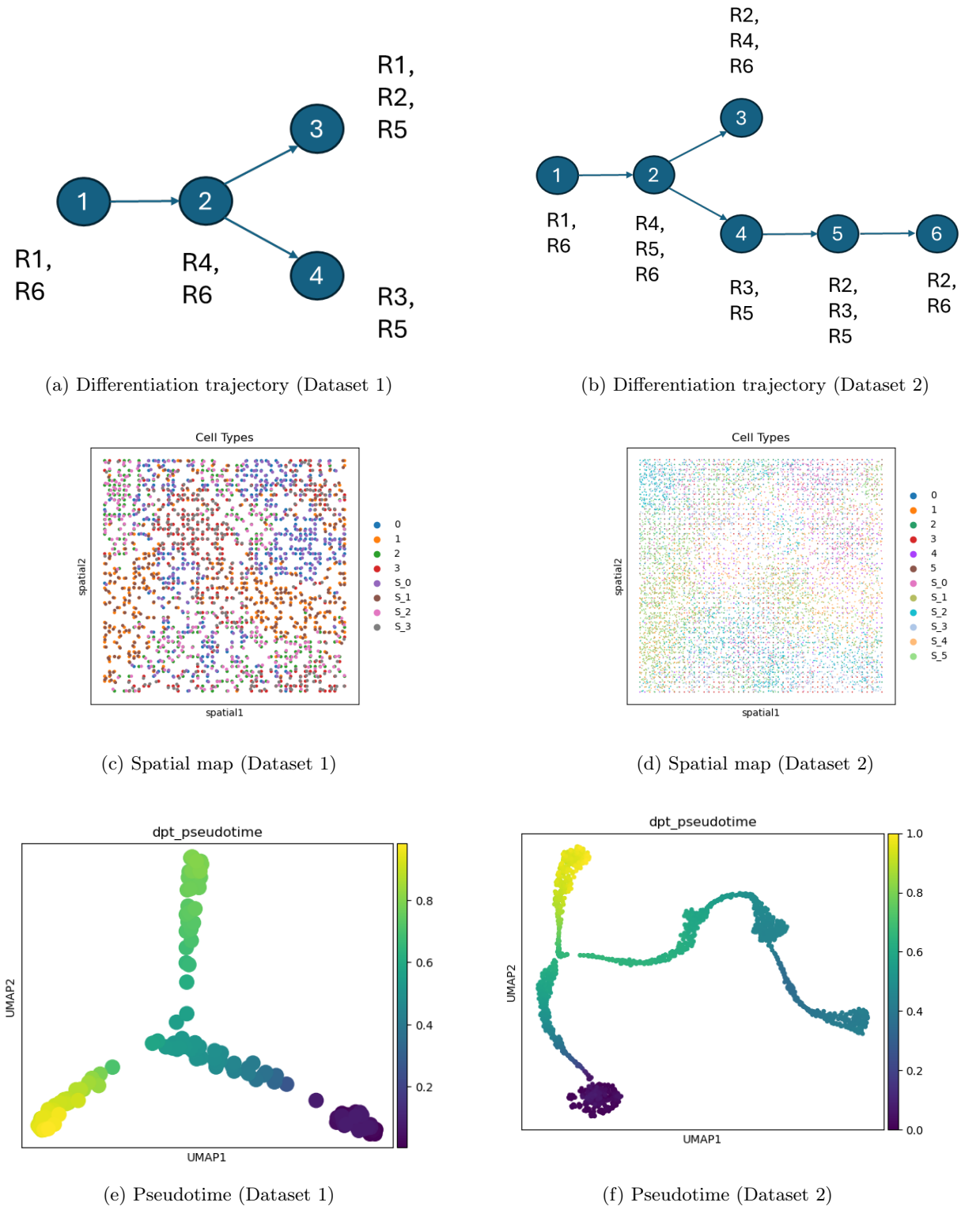
where $\epsilon \sim \mathcal{N}(0,\ \sigma^2)$ ensures variability. This equalization propagates to all sender neighbors of the receptor i within distance δ. To prevent the expression values from being negative, the receptor and ligand values are low-bounded by 0. We generated two simulated datasets with SERGIO, following the simulation scheme as described in section A.1. The ligand–receptor binding and GRN relationships are provided in the supplementary files. The differentiation trajectories, spatial maps, and pseudotime of the two datasets are presented below.

A.2 Simulation Data Description

In Simulation 1, sender cells were positioned within a spatial distance threshold of $\theta = 3$ from receiver cells on a 920×920 grid, with low independent Gaussian noise added to ligand and receptor expression profiles ($\epsilon = 0.1$). The inferred pseudotime progression faithfully reflected the developmental trajectory of cell states.

In Simulation 2, we introduced greater stochasticity to evaluate model robustness, including increased spatial dispersion of sender cells ($\theta = 4$) around receivers on a 138×138 grid, higher independent noise in ligand and receptor expression ($\epsilon = 0.5$), and a disrupted pseudotime ordering achieved by inverting the relative progression of cell types 4, 5, and 6 (see Fig. S1). These perturbations collectively emulate realistic experimental noise and biological heterogeneity.

The matching ground-truth LR-TG, TF-TG, LR and sender-receiver relationships are presented in Figs. S2 and S3.



(a) Differentiation trajectory (Dataset 1)

(b) Differentiation trajectory (Dataset 2)

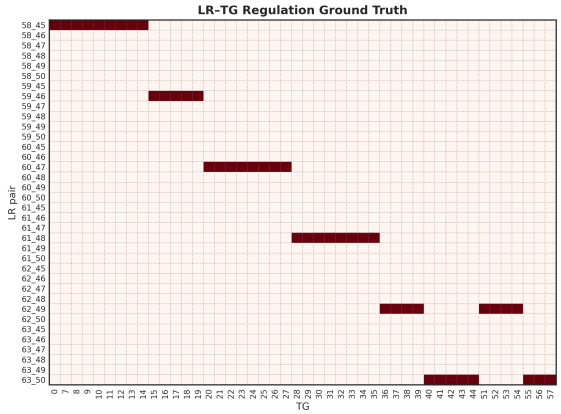
(c) Spatial map (Dataset 1)

(d) Spatial map (Dataset 2)

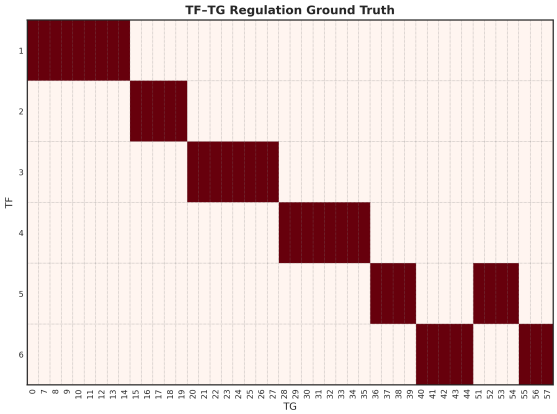
(e) Pseudotime (Dataset 1)

(f) Pseudotime (Dataset 2)

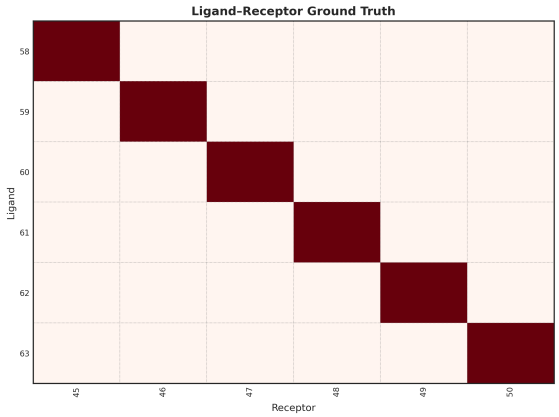
Fig. S1: Overview of simulated datasets generated with SERGIO with sender cells and receptor values added post-hoc. Shown are (a,b) differentiation trajectories, (c,d) spatial maps, and (e,f) pseudotime for Dataset 1 and Dataset 2, respectively.



(a) LR–TG regulatory map

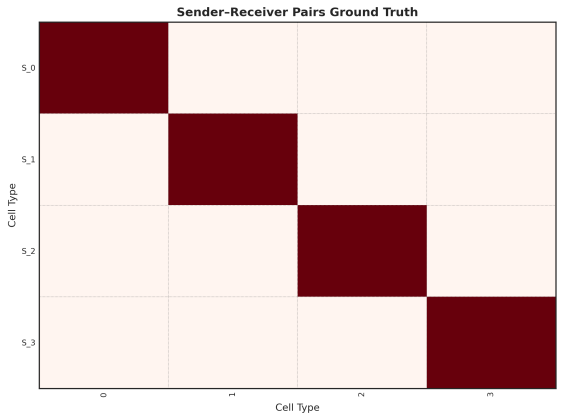


(b) TF–TG regulatory map

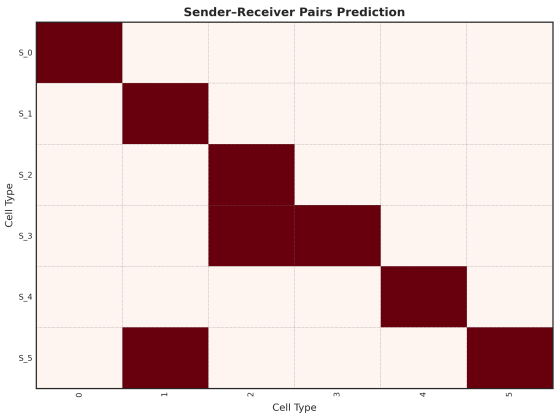


(c) Ligand–receptor pairing map

Fig. S2: Ground truth gene relationships for both simulation data used for model evaluation. Each panel depicts a different type of known biological relationship: (a) ligand–receptor (LR) to target gene (TG) regulation, (b) transcription factor (TF) to TG regulation, (c) LR pairing networks.



(a) Sender–receiver interaction pairs



(b) Sender–receiver interaction pairs

Fig. S3: Ground truth sender-receiver relationships for both simulation data used for model evaluation. (a) sender–receiver relationship for simulation data 1, (b) sender–receiver relationship for simulation data 2

A.3 Dummy ligand–receptor pathways for benchmarking

Benchmarking methods such as COMMOT, CellChat, and SpaTALK require a predefined ligand–receptor (LR) interaction library as input. To emulate a pathway-agnostic setting in which no prior knowledge of LR binding specificity is assumed and any ligand–receptor pair may potentially interact, we constructed a dummy LR network by fully pairing all ligands $l \in L$ with all receptors $r \in R$. For SpaTALK, which additionally relies on a receptor–TF regulatory graph, we introduced a dummy self-regulatory edge $r \to r$ for each receptor to ensure model compatibility. Similarly, for the TF–TG module, we added a dummy edge $r \to g$ linking each receptor to its corresponding ground-truth target gene.

B MOSTA Processing

B.1 Data Processing

The MOSTA dataset [8] was obtained from the original publication (`https://www.cell.com/cell/fulltext/S0092-86742200399-3`). It comprises 26,738 cells and 24,045 genes, including 7,167 annotated progenitor cells—radial glia cells (RGCs), neuroblasts (NeuBs), and glioblasts (GlioBs). Our analysis focused on midbrain progenitor development, tracing the differentiation of RGCs into NeuBs and GlioBs across embryonic stages E12.5, E14.5, and E16.5, where the spatial distribution and pseudotime relationships are illustrated in Fig. S7.

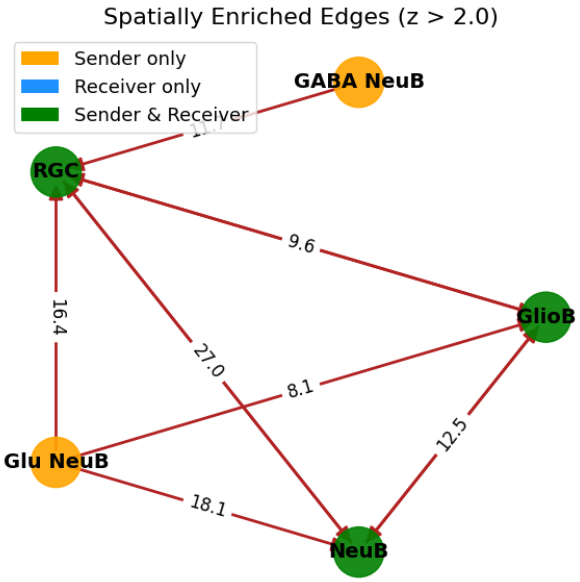
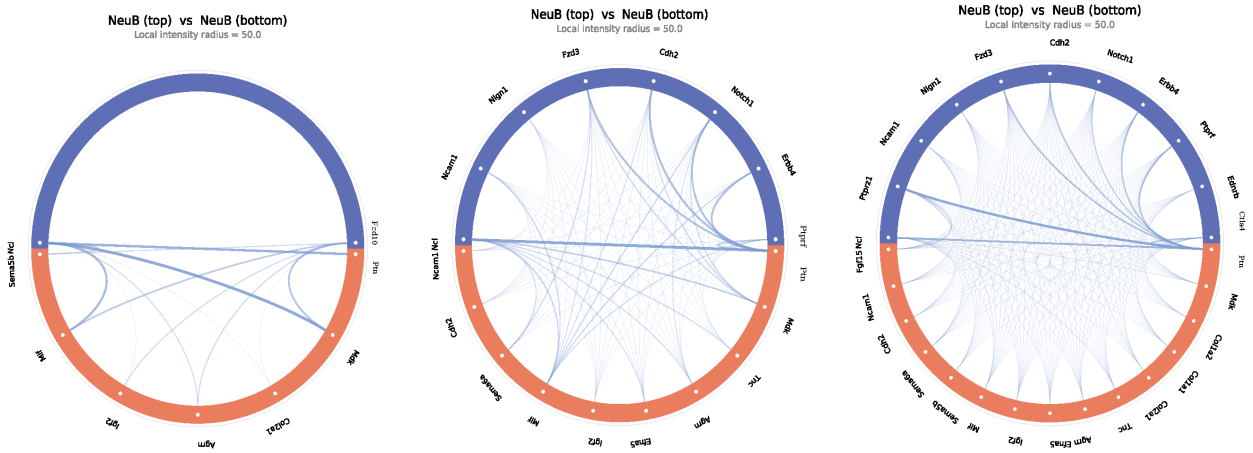


Fig. S4: Spatial Enrichment networks between the senders and receivers in the MOSTA data.

Preprocessing For quality control, we removed genes expressed in fewer than 10 cells and cells expressing fewer than 10 genes. Following the approach described in [8], we performed pseudotime analysis using Monocle3 to reconstruct developmental trajectories. The spatial distributions of the three progenitor populations across developmental stages, together with their corresponding pseudotime maps, are shown in Fig. S7.

Gene Selection To identify spatially variable genes (SVGs), we applied the PROST framework, which quantifies spatial expression structure for each gene using the Pattern Index (PI). The PI measures the

degree to which a gene exhibits coherent spatial structure rather than random spatial variation. Since the PI is a conservative metric and typically yields low absolute values, we retained all ligands, receptors, and transcription factors with PI > 0 to capture even weak but potentially meaningful spatial patterns.



(a) CCC from IMN to nptxEx at 10 DPI (b) CCC from IMN to nptxEx at 15 DPI (c) CCC from IMN to nptxEx at 20 DPI

Fig. S5: Inferred NeuB autocrine cell-cell signaling of E12.5, E14.5, and E16.5, respectively.

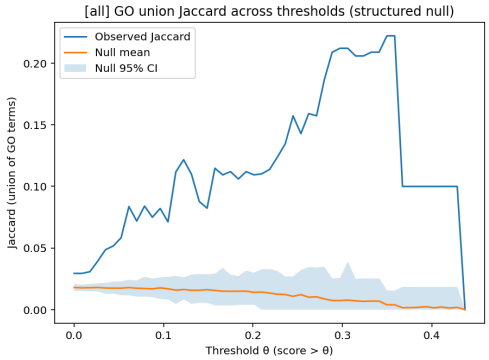
B.2 Training

The five progenitor-associated cell types—NeuB, GlioB, RGC, GABA NeuB, and Glu NeuB—were analyzed as potential signal–sending populations in both autocrine and paracrine contexts, with NeuB, GlioB, and RGC designated as receivers. To focus on informative spatial structure, we retained spatially variable genes identified by PROST [14] that were expressed in at least 10% of cells within any cell type. This filtering resulted in 41 active ligands, 25 receptors, 20 transcription factors (TFs), and 1,501 candidate target genes. Pseudotime trajectories were reconstructed using Monocle 3 [15], and spatial proximity constraints were incorporated to capture short-range communication among progenitor populations.

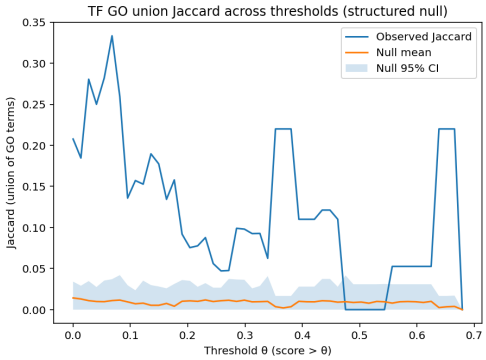
SpaTRACE was trained on the resulting pseudotime-sampled trajectories using default hyperparameters ($d_{\text{hidden}} = 256$, $w^{\text{lr}}_{\text{global}} = 0.3$, $w^{\text{tf}}_{\text{global}} = 0.3$, $w^{\text{lr}}_{\text{percell}} = 0.5$, $w^{\text{tf}}_{\text{percell}} = 0.5$, $\lambda^{\text{global}}_{\text{att}} = 0.3$, $\lambda^{\text{percell}}_{\text{att}} = 0.5$) for 100 epochs with EarlyStopping. Training converged in approximately two hours on an H100-v8 node with eight GPUs. After training, we applied a low-pass filter to retain only the top 10% of LR–TG and TF–TG interaction scores. The resulting ligand–receptor interaction network (Fig. 3c) revealed coherent signaling hierarchies across developmental stages. Thresholding interaction strengths at > 0.4 further highlighted stage-specific and context-dependent LR–TF–TG connectivity patterns, delineating dynamic communication pathways underlying midbrain maturation. SpaTRACE achieved bivariate Moran's I scores comparable to those of established spatial-inference methods such as COMMOT and SpaTalk, while simultaneously attaining a higher co-expression fraction (Figs. 3a and 3b).

B.3 Thematic Overlap on the MOSTA

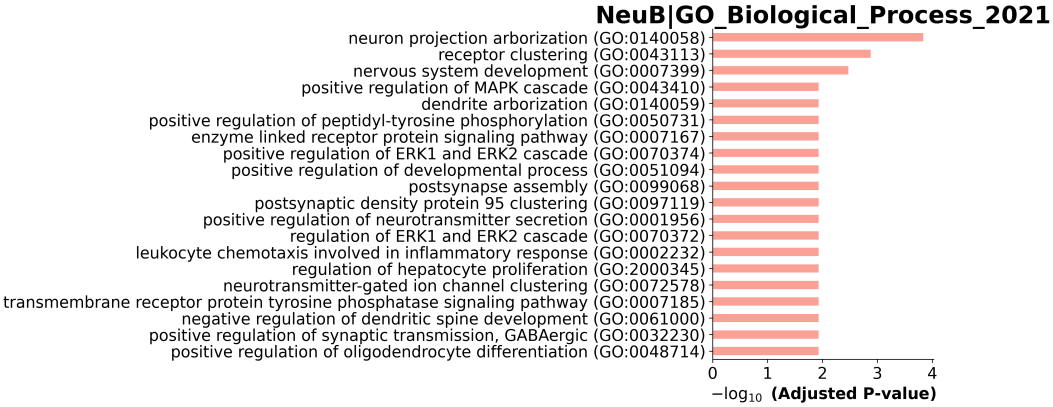
We present here the results of the Thematic Overlap analysis on the MOSTA dataset, along with the top 20 enriched GO terms identified in the recovered LR–TG and TF–TG pathways. A substantial proportion of the enriched terms are consistent with processes related to brain development and glioblast cell differentiation.



(a) Thematic overlap between inferred LR–TG pathways and DE genes. The curve reflects the increasing Jaccard Index with GO terms associated with DE genes.



(b) Thematic overlap between inferred TF–TG pathways and DE genes. The curve reflects the increasing Jaccard Index with GO terms associated with DE genes.



(c) Top 20 GSEA enrichment results on the Union of the inferred LR-TG and TF-TG pathways.

Fig. S6: Comparison of thematic overlap and enrichment analyses. (a–b) Jaccard Index–based thematic overlap between inferred signaling (LR→TG) and regulatory (TF→TG) pathways with DE-gene GSEA results across selection thresholds. The amber line indicates the null expectation with corresponding confidence intervals. (c) GSEA enrichment plots showing strong concordance between inferred LR→TG and TF→TG interactions and DE-gene functional categories, exceeding the null baseline.

C Axolotl Regenerative Telencephalon Interpretation via Spatiotemporal Transcriptomic Atlas

The ARTISTA (Axolotl Regenerative Telencephalon Interpretation via Spatiotemporal Transcriptomic Atlas) dataset [9] provides developmental spatial transcriptomic (ST) data capturing the regenerative process of the axolotl brain. In this study, we focused on the dynamic evolution of progenitor cells during regeneration. The dataset spans 2–60 days post-injury (DPIs). Since axolotl telencephalon regeneration primarily occurs between Days 2–20, we restricted our analysis to this period.

The original study identified EGC cells as the primary progenitor population, which transition through intermediate states—rIPC1 and IMN—before reaching the terminal neuronal subtype, nptxEX. Similar to our processing of the MOSTA dataset, we performed rigorous quality control by filtering out genes expressed in fewer than 10% of cells in any given cell type or stage. To capture the key transcriptional changes across regeneration, we extracted all genes significantly differentially expressed at each stage compared to the control group (log-fold change threshold = 2.5). This analysis yielded 114 ligands, 120 receptors, and 117 transcription factors (TFs). For target genes (TGs), we retained all genes expressed in more than 10% of any cell type, resulting in 3,977 such genes. Following the instruction by [9], we applied Monocle3 for

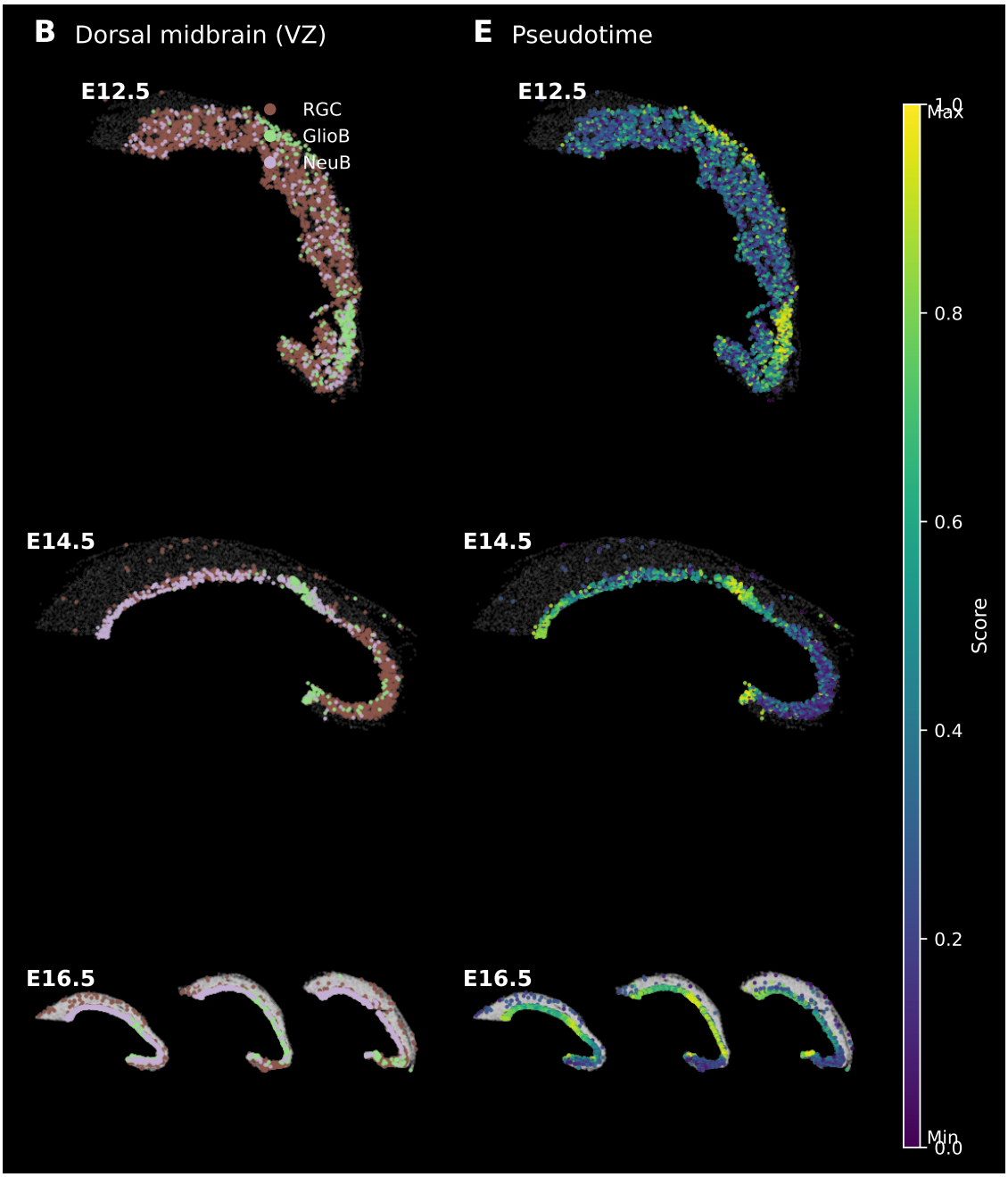
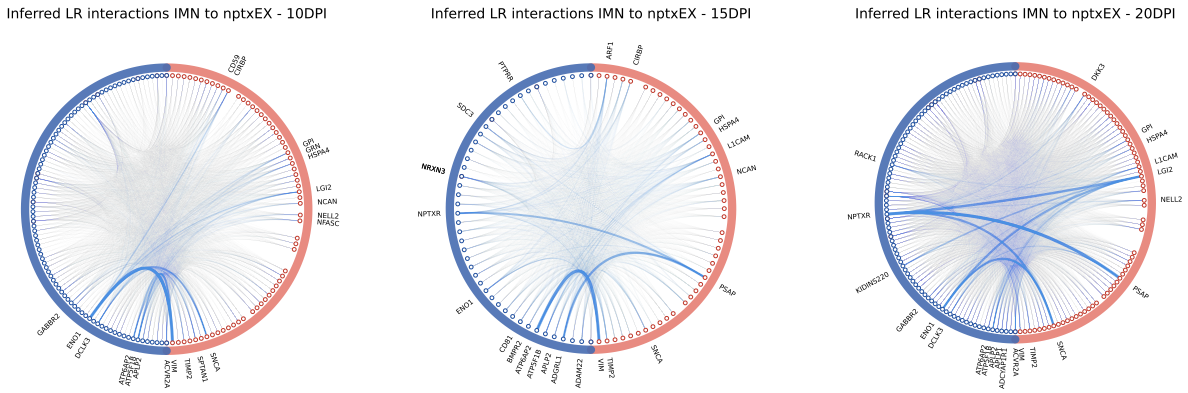


Fig. S7: Spatial Distribution of progenitor cell types and dpt across batches E12.5, E14.5, and E16.5.

pseudotime derivation [15]. We randomly sampled 10 developmental trajectories of length 3 for each cell, yielding approximately 30,000 training and validation instances. The model was then trained on these samples using the default hyperparameter settings ($d_{\text{hidden}} = 256$, $w^{\text{lr}}\text{global} = 0.3$, $w^{\text{tf}}\text{global} = 0.3$, $w^{\text{lr}}\text{percell} = 0.5$, $w^{\text{tf}}\text{percell} = 0.5$, $\lambda^{\text{global}}\text{att} = 0.3$, $\lambda^{\text{percell}}\text{att} = 0.5$). Training was performed for 100 epochs with EarlyStopping. The full training run required approximately 15 hours to converge on an H100-v8 node equipped with 8 GPUs. For gene-gene inference, we applied a low-pass filtering to keep only the top 10% LR-TG and TF-TG signals. Then, we reconstructed the LR relationship as in Eq. 11 for each individual batch. We then selected spatially enriched signalling patterns of the progenitor cells and derived the CCC networks in each batch. For each batch, we filtered out ligands that are expressed less than 10% in the sender

cells, and so are the receivers in the receiver cells. For illustration purpose, the top 20 strongest signals of each sender-receiver pair was kept in the visualization (Fig. S8).



(a) CCC from IMN to nptxEX at 10 DPI (b) CCC from IMN to nptxEX at 15 DPI (c) CCC from IMN to nptxEX at 20 DPI

Fig. S8: Comparison of the inferred CCC signaling patterns from IMN to nptxEX from 10 - 20 DPIs. The top 20 signals of each batch is highlighted.

C.1 Quantitative Metrics for Axolotl Brain Regeneration Data

We evaluated the co-expression levels of 174 significantly inferred cell–cell communication (CCC) ligand–receptor (LR) pairs with interaction scores above a threshold of 0.4. These were compared against 400 randomly sampled background LR pairs drawn from the batch-specific background distribution.

As shown in Fig. S9a, the inferred active LR pairs exhibit substantially higher co-expression fractions than the random background. The background LR pairs have a mean co-expression fraction of 0.180, whereas the inferred LR pairs achieve a markedly higher mean of 0.456 (Welch's t-test, $p = 3.74 \times 10^{-41}$). Similarly, the inferred LR pairs show an elevated average bivariate Moran's I value of 0.1213, significantly above that of the randomly sampled background pairs (Welch's t-test, $p = 7.884 \times 10^{-15}$).

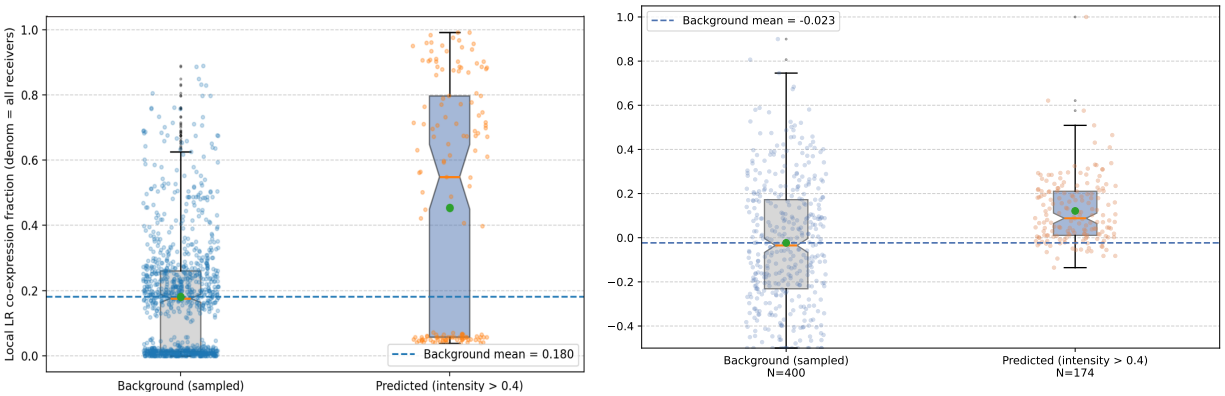
D Quantitative Metrics

D.1 Co-expression fractions

The definition of co-expression fraction follows SpaTALK [7], adapted to our notation. For sender cell type $A \in \mathcal{T}_A$, receiver cell type $B \in \mathcal{T}_B$, ligand $l \in \mathcal{G}_L$, receptor $r \in \mathcal{G}_R$, and distance threshold θ, the co-expression fraction is defined as

$$\mathrm{CF}_{A\to B}(l,r) = \frac{\left|\left\{(c,c') :\ c \in \mathcal{C}_A,\ c' \in \mathcal{C}_B,\ D(c,c') \le \theta,\ x_c^l > 0,\ x_{c'}^r > 0\right\}\right|}{\left|\left\{(c,c') :\ c \in \mathcal{C}_A,\ c' \in \mathcal{C}_B,\ D(c,c') \le \theta\right\}\right|}.$$

Here, x_c^l and $x_{c'}^r$ denote the expression levels of ligand l in sender cell c and receptor r in receiver cell c', respectively. The numerator counts spatially proximal sender–receiver cell pairs that co-express the ligand–receptor pair (l, r), while the denominator counts all sender–receiver pairs within distance θ. Thus, $\mathrm{CF}_{A\to B}(l,r)$ quantifies the fraction of spatially adjacent cell pairs supporting co-expression of ligand l and receptor r.



(a) Average co-expression fraction for the axolotl brain regeneration dataset. Welch's t-test: $p = 3.74 \times 10^{-41}$.

(b) Distribution of bivariate Moran's I for ligand–receptor pairs with intensity > 0.4 in the axolotl brain regeneration dataset. Welch's t-test: $p = 7.884 \times 10^{-15}$

Fig. S9: Comparison between background and predicted (intensity > 0.4) ligand–receptor pairs in the axolotl brain regeneration data. (a) Co-expression fraction. (b) Bivariate Moran's I distribution, measuring spatial autocorrelation between ligand and receptor expression across spots.

D.2 Bivariate Moran's I index

The Bivariate Moran's I index quantifies the spatial correlation between two gene expression patterns by measuring how the expression of a ligand in sender cells covaries with the expression of its corresponding receptor in nearby receiver cells. Let $\mathbf{x} = (x_1^l, \ldots, x_N^l)$ denote ligand expression over all cells, $\mathbf{y} = (x_1^r, \ldots, x_N^r)$ denote receptor expression over cells, and $W = (w_{ij})$ be a spatial weight matrix defined over the N cells. The Bivariate Moran's index is given by

$$I_{\text{biv}}(\mathbf{x}, \mathbf{y}) = \frac{N}{S_0} \cdot \frac{\sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij} \, (x_i - \bar{x}) \, (y_j - \bar{y})}{\sqrt{\sum_{i=1}^{N} (x_i - \bar{x})^2} \sqrt{\sum_{j=1}^{N} (y_j - \bar{y})^2}},$$

where $\bar{x}$ and $\bar{y}$ denote the mean ligand and receptor expression levels, respectively, and $S_0 = \sum_{i=1}^{N} \sum_{j=1}^{N} w_{ij}$ is the total spatial weight. A positive value of I_{biv} indicates that high ligand expression tends to occur near cells with high receptor expression (and similarly for low expression), whereas negative values indicate spatial discordance between the two expression patterns.

D.3 Thematic overlap between inferred programs and DE-driven biology

We assessed thematic consistency between inferred signaling programs and DE-driven biology using the Jaccard Index between their enriched GO terms. Let A denote the set of GO terms enriched from differentially expressed (DE) genes, and $B(\theta)$ the set of GO terms enriched from inferred ligand–receptor to target gene or transcription factor–to–target gene relationships at score threshold θ.

For each of the LR→TG score matrix and the TF→TG matrix, it is first normalized to the interval $[0, 1]$. For selected cell types, DE analysis was carried out using the Wilcoxon rank–sum test (with thresholds of $|\log_2 \text{FC}| \geq 0.25$, adjusted $P \leq 0.05$). The union of DE genes intersected with measured genes formed the anchor set A_{genes}. Over-representation analysis (GO Biological Process, term size $[5, 2000]$, FDR ≤ 0.05, top 50 terms) yielded A. For each ligand–receptor pair and threshold $\theta \in [0, 1]$, active target genes and transcription factors were identified, merged into a single set, and subjected to over-representation analysis. The union of significant terms across all pairs defined $B(\theta)$.

For each threshold θ (51 evenly spaced points) we computed the Jaccard index, $J(\theta) = |A \cap B(\theta)|/|A \cup B(\theta)|$. To assess significance, we compared the observed curves with two null models: a *simple* null preserving only the union size, and a *structured* null preserving the per–pair enrichment yield before union. Each null

distribution was estimated using 200 repetitions with a fixed random seed, from which we derived null means, 95% confidence intervals, and right-tailed P-values. To correct for multiple testing across thresholds, we further compared the observed maxima $\max_\theta J(\theta)$ against the null distributions of maxima (Westfall–Young style correction).

Over-representation analysis was performed using Enrichr GO Biological Process, with the gene universe set to all measured genes. Default settings were: FDR = 0.05, top 50 terms, term size $[5, 2000]$, a 51-point threshold grid, and 200 null repetitions. It is important to note, however, that the thematic overlap only evaluates whether the inferred gene–gene relationships align with current biological knowledge; this metric may be deflated because some GO terms enriched in the differential expression analysis can represent false positives.

E Benchmark Methods

E.1 COMMOT Analysis

For COMMOT, the default *CellChat* mouse database was employed as the reference. Cellular neighborhoods were defined using a distance–threshold approach with a radius of $50\,\mu m$, consistent with our experimental setting. For the MOSTA dataset, analyses were carried out separately for each time point, and only genes expressed in more than 10% of cells within any cell type were retained, following the filtering implemented in custom preprocessing functions.

In cases where a dataset contained multiple batches per time point, the retained ligands and receptors were obtained by combining them across batches, and average communication intensity scores were computed. Ligand–receptor (LR) interaction strengths were quantified using `ct.tl.spatial_communication`, and directional summaries were generated with `ct.tl.communication_direction`. For the axolotl dataset, since no curated LR databases are available, the mouse database was used as a substitute, with the same gene naming. Top-ranked LR pairs were identified by aggregating sender and receiver scores, and visualisations were produced with `ct.pl.plot_cell_communication`. All LR interaction results are summarised as heatmaps of LR interaction intensities.

E.2 CellChat Analysis

For CellChat[3], we used version 1.0 with the default curated mouse database as reference. Each batch was analyzed independently, and we computed the log probability of ligand–receptor (LR) interactions from sender to receiver cell types. A minimum cell count threshold of three and a significance cutoff of 0.05 were applied. However, no LR pairs passed significance across any developmental stage. To enable comparison with other methods, we therefore selected the top 100 LR pairs ranked by interaction weight as CellChat's predicted set.

E.3 SpaTalk Analysis

For SpaTalk[7], we applied the default pipeline using the standard parameter settings provided by the package. The raw gene expression matrices and spatial coordinates from the MOSTA dataset were used as input without additional preprocessing beyond the shared filtering steps described previously. SpaTalk infers cell–cell communication by integrating spatial proximity with ligand–receptor (LR) co-expression patterns through its probabilistic model. Under the default configuration, the method employs pre-trained reference gene sets to estimate spatial communication likelihoods and constructs interaction graphs linking putative sender and receiver cell types. As with the other frameworks, the analysis was performed separately for each developmental time point. The resulting LR interaction probabilities were aggregated across replicates, and the top-ranked interactions were selected based on SpaTalk's internal scoring metric. Interaction networks were visualised using the package's built-in plotting functions, with edge weights representing inferred communication strengths.