Mixtures of Neural Cellular Automata: A Stochastic Framework for Biological Growth Modelling

Salvatore Milite [1] **Giulio Caravagna** [2] **Andrea Sottoriva** [1]

Abstract

Neural Cellular Automata (NCAs) offer a powerful framework for modeling self-organizing processes with potential applications in biomedicine. However, their deterministic nature limits their ability to represent the stochastic dynamics of real biological systems.

We introduce the Mixture of Neural Cellular Automata (MNCA), a novel extension that incorporates stochasticity and probabilistic rule clustering. By combining intrinsic noise with learned rule assignments, MNCAs can capture heterogeneous local behaviors and emulate the randomness inherent in biological processes.

We assess MNCAs on synthetic tissue simulations and spatial transcriptomics data from mouse intestine. Our results show improved reconstruction of biological growth patterns and interpretable segmentation of local rules, establishing MNCAs as a promising tool for modeling complex biological dynamics.

1. Introduction

Biological systems exhibit complex spatial patterns driven by stochastic interactions between cells and molecules. Traditional models often struggle to capture this complexity due to limited expressiveness and scalability. Cellular Automata (CAs) can simulate self-organization through simple local rules, but classical CAs are difficult to parameterize, and Neural Cellular Automata (NCAs), though more flexible, are typically deterministic, limiting their biological realism.

Stochasticity is central to processes such as gene expression and differentiation, where identical genomes can lead to diverse behaviors due to molecular noise. Capturing this variability is essential for accurate modeling.

We propose the **Mixture of Neural Cellular Automata (MNCA)**, a novel framework that incorporates stochasticity and clustering into NCAs, enabling the simulation of heterogeneous and probabilistic cellular behaviors.

Our contributions are:

- **MNCA Framework:** We develop the MNCA framework that extends NCAs by incorporating stochasticity and clustering, allowing for the simulation of a more diverse set of local behaviors.
- **Biological Applications:** Demonstrates MNCA on simulations of tissue development and on real spatial transcriptomic data from mouse intestine.

2. Background

2.1. Spatial Modelling of Biological Systems

Spatial mathematical modeling plays a crucial role in understanding biological systems by capturing how spatial distributions and interactions among cells, molecules, or organisms influence system-level behavior. This is essential for studying emergent phenomena where spatial context shapes biological function.

In many cases, spatial organization directly determines biological outcomes. Morphogen gradients guide differentiation during development (Rogers & Schier, 2011), spatial cues in the intestine regulate stem cell identity (Beumer & Clevers, 2021), and tumor cell positioning affects cancer progression and treatment response (Seferbekova et al., 2023).

Agent-based models are widely used to represent such spatial phenomena. Among these, stochastic cellular automata offer a simple yet powerful framework, applied to cancer growth (Tari et al., 2022; Lewinsohn et al., 2023; Sottoriva et al., 2010), development (Ermentrout & Edelstein-Keshet, 1993), and ecology (Balzter et al., 1998). Parameters are typically inferred from experiments or via Approximate

[1]Computational Biology Research Centre, Fondazione Human Technople, Milan, Italy [2]Department of Mathematics and Geosciences, University of Trieste, Trieste, Italy. Correspondence to: Salvatore Milite <salvatore.milite@fht.org>, Andrea Sottoriva <andrea.sottoriva@fht.org>.

Proceedings of the Workshop on Generative AI for Biology at the 42nd International Conference on Machine Learning, Vancouver, Canada. PMLR 267, 2025.

Bayesian Computation (ABC) (Noble et al., 2022), with recent efforts exploring ABC–Deep Learning integration (Cess & Finley, 2023).

2.2. Cellular Automata

Cellular Automata (CA) are discrete computational models that simulate the physical dynamics of complex systems through simple local interactions. Historically CA where introduced by Von Neumann as a model of self-reproducing systems (Von Neumann et al., 1966)

In a CA, the state of each cell $s_i \in S$ at discrete time t is updated based on a local update rule:

$$s_i^{t+1} = f\left(s_i^t, \{s_j^t \mid j \in \mathcal{N}(i)\}\right), \tag{1}$$

where $\mathcal{N}(i)$ denotes the neighborhood of cell i, and f is a deterministic function defining the update rule.

While the update function f in a standard cellular automaton is strictly deterministic, probabilistic rules are more fitting for capturing the randomness inherent in processes like biological evolution. A stochastic cellular automaton (SCA) introduces randomness into the update rule:

$$s_i^{t+1} \sim P\big(f\left(s_i^t, \{s_j^t \mid j \in \mathcal{N}(i)\}\right)\big) \tag{2}$$

Specifically, each cell's next state is drawn from a probability distribution P that depends on its current state and the states of its neighbors.

2.3. Neural Cellular Automata

Neural Cellular Automata (NCAs) extend traditional CA by replacing the deterministic update function with a neural network (Mordvintsev et al., 2020; Gilpin, 2019). The update rule becomes:

$$s_i^{t+1} = s_i^t + \phi\left(s_i^t, \{s_j^t \mid j \in \mathcal{N}(i)\}; \theta_k\right), \tag{3}$$

where ϕ is a neural network parameterized by weights θ. The neural network processes the current state of a cell and its neighbors to compute an update, which is added to the existing state. This enables the NCA to learn complex behaviors via backpropagation, avoiding hand-crafted rules or inefficient evolutionary and sampling-based methods.

Recent work has extended NCAs in various directions: attention-based updates (Tesfaldet et al., 2022), generalization to graph domains (Grattarola et al., 2021), latent-space dynamics (Menta et al., 2024), and PDE approximation (Richardson et al., 2024). Multi-rule approaches have also emerged, such as (Hernandez et al., 2021), where an autoencoder maps an image to its generative rule. However, these models remain largely deterministic, limiting their biological realism where stochastic, cell-specific behavior is key (Noble et al., 2022).

Stochastic NCAs have been explored in generative contexts: as VAE decoders (Palm et al., 2022), denoising diffusion models (Kalkhof et al., 2025), or hierarchical processes (Zhang et al., 2024). (Zhang et al., 2021) proposed a stochastic NCA outputting Bernoulli probabilities.

We focus on NCA formulations that preserve classical locality assumptions, crucial for modeling morphogenesis, collective behavior, and development. Accordingly, we compare only with models like the GCA (Zhang et al., 2021). In particular, in this work, we will use a GCA that learns the mean and variance of a Gaussian distribution, as it better aligns with our experimental setup (this is equivalent to equation 2 where P is a Gaussian).

3. Mixture of Neural Cellular Automata

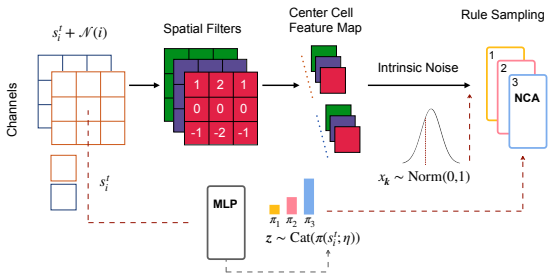

Figure 1. **Graphical description of the MNCA** The model integrates signals from a central cell and its neighbors using spatial filters, while a Rule Selector outputs probabilities for the different NCAs.

The Mixture of Neural Cellular Automata (MNCA) is a framework that extends traditional Cellular Automata (CA) by incorporating multiple sets of local update rules within a single grid-based system. In MNCA, each cell can be governed by one of several distinct automata, allowing for the modeling of heterogeneous systems where different regions exhibit unique local interactions.

3.1. Model Definition

Consider a grid of N cells, where each cell i has a state $s_i^t \in S$ at time t and S is the set of possible states. We define a set of K distinct cellular automata, each characterized by its own neural-network parametrized transition function $\phi_k : S^{|\mathcal{N}(i)|+1} \to S$, with parameters θ_k and where $\mathcal{N}(i)$ denotes the neighborhood of cell i.

The state update for cell i at time $t+1$ is given by:

$$\boldsymbol{z} \sim \mathrm{Cat}(\pi(s_i^t, \eta)) \tag{4a}$$

$$x_k \sim \mathrm{Norm}(0,1) \tag{4b}$$

$$s_i^{t+1} = s_i^t + \prod_{k=1}^{K} \phi_k \left(s_i^t, \{s_j^t \mid j \in \mathcal{N}(i)\}, x_k; \theta_k\right)^{z_k} \tag{4c}$$

where $z_i \in \{0,1\}$ is the automaton assigned to cell i, $\boldsymbol{z}$ is a random variable distributed according to a Categorical distribution. A neural network with parameters η controls the rule assignment probability. The intrinsic Gaussian noise x_k can be used as an internal source of randomness to implement intra-rule stochastic updates. To back-propagate through the Categorical distribution at training time, we use the Gumbel-Softmax trick (Jang et al., 2016).

4. Experiments

Our main motivation for studying Mixtures of Neural Cellular Automata is to model cellular development. To demonstrate its potential, we designed a synthetic experiment of stem-driven tissue growth. We then applied our model to real data from mouse intestinal crypts.

4.1. Synthetic Data of Biological Development

For the setup in Figure 2A, we generated 200 simulations of tissue growth on 35×35 grids. Each grid square represents a cell, with simulations initialized from a central cluster of $[5, 15]$ stem cells and evolved for 35 steps. Stem cells divide into either new stem or intermediate cells, with two intermediate types showing decreasing plasticity. Differentiation yields two terminal types, one requiring interaction between intermediate and type-1 differentiated cells. Full simulation details are in Appendix A.

We trained the three models from Section 3 to reconstruct system states over time using one-hot encoded cell types as input (training details in Appendix D). Reconstruction results (Figure 2B–C) show that MNCAs achieved significantly better fidelity than deterministic NCAs, with a KL divergence drop from 2.045 to 0.018 (Table 1).

MNCAs also generated realistic synthetic tissues, better matching true spatial patterns and producing all cell types—including DIFFERENTIATED 2, which standard NCAs failed to generate.

Rule assignment analysis showed MNCAs learned distinct dynamics for most cell types, allocating one rule to empty space and merging the two intermediates due to their similar behavior. Identifying class-specific behaviors in the system.

Compared to ABC-based ABMs (Appendix F), MNCAs achieve similar simulation quality with minimal supervision.

Table 1. **Comparison of NCA variants.** KL-div, Size-W, and Border-W are respectively the KL divergence for the cell types probability and the Wasserstein distance of the cell composition, tissue size, and border size in the generated and original dataset. For a formal definition of the metrics see Appendix B

MODEL	KL-DIV	SIZE-W	BORDER-W
NCA	2.057 ±0.000	0.547 ±0.000	0.430 ±0.000
GCA	0.112 ±0.003	0.477 ±0.002	0.339 ±0.005
MNCA	**0.018 ±0.001**	**0.061 ±0.008**	**0.184 ±0.010**

While ABC requires prior knowledge and manual tuning, our model learns directly from the data.

4.2. 10X Visium Analysis of Intestinal Crypts Recovery

To evaluate our model in a real biological context, we analyzed a Visium spatial transcriptomics dataset from a full mouse intestine section (Parigi et al., 2022). The dataset includes a tissue image with gene expression measured in fixed-size spatial spots, capturing only a terminal snapshot from the adult mouse, with no temporal dynamics.

To introduce a dynamic element, we manually annotated the stem cell compartment and tested the model's ability to reconstruct the tissue by iteratively adding neighbor spots. The details on this procedure are described in Appendix E.

For each spot, we computed expression values for several core biological pathways, based on curated annotations from (Schubert et al., 2018). These pathways, encompassing both internal and environmental signaling components, approximate the molecular context sensed by each cell.

In this complex setting, MNCAs achieved markedly lower reconstruction error (MSE: 0.055 NCA, 0.049 GCA, 0.021 MNCA), demonstrating superior performance also in reconstructing real data.

5. Conclusion

In this work, we extended Neural Cellular Automata (NCA) by introducing stochasticity, proposing the Mixture of Neural Cellular Automata (MNCA) to bridge modern NCA models with the stochastic nature of biological agent-based simulations.

Through synthetic experiments and spatial transcriptomic data analysis, we showed that MNCAs can robustly and interpretably simulate complex biological systems.

This work opens several future directions: scaling to larger systems, improving rule interpretability, and handling incomplete time-series data. The general MNCA framework can also be adapted to the structural and technical nuances of emerging single-cell spatial biology technologies.

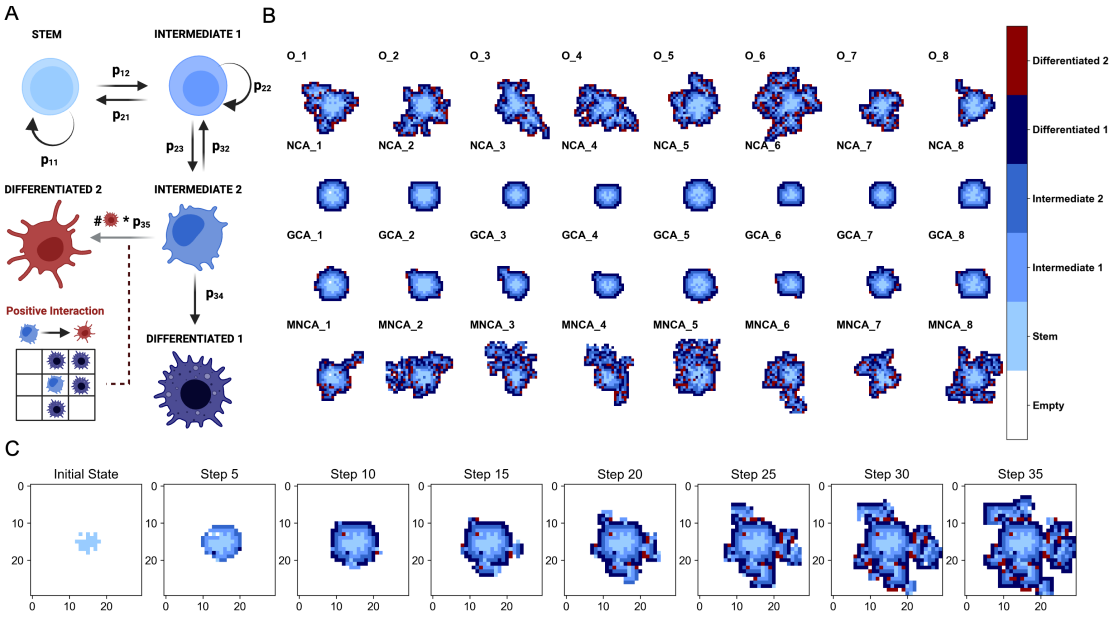

Figure 2. **Visualization of MNCA's stochastic framework applied to synthetic tissue growth.** The figure illustrates the development of cellular patterns over time, starting from an initial configuration of stem cells and evolving into differentiated tissue structures. In panel A there is a brief description of the tissue model used for the simulation. In panel B we show an example of 8 realizations of the process. In the first row we have the original final tissue, in the second row the deterministic NCA, in the third the MNCA and in the last row the MNCA with noise (O= Origina). In Panel C there is an example of full tissue evolution with an MNCA starting from a random initial configuration and evolving the model for 35 steps

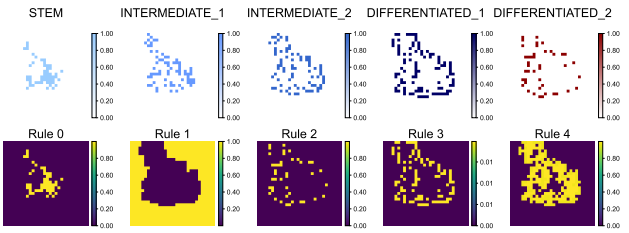

Figure 3. **Visualization of rule assignments in MNCA simulations:** We first split the tissue into separate channels based on cell type to make the visualization easier (shown in the first row). In the second row, we display the assignment probabilities for each rule. The results show that distinct rules are strongly associated with specific cell types, effectively producing a semantically meaningful segmentation of the image.

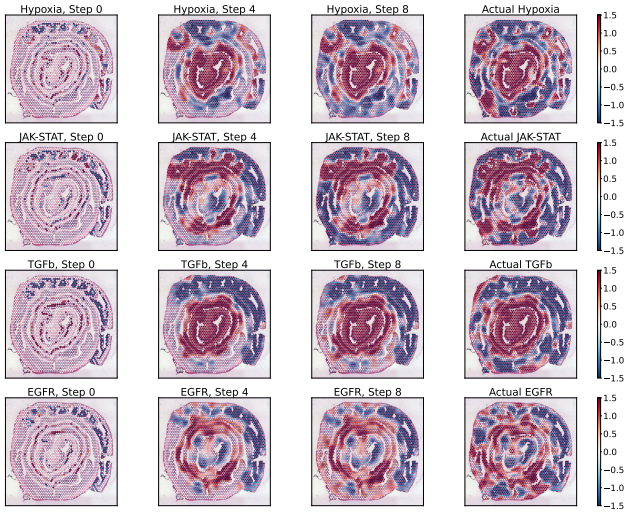

Figure 4. **Analysis of 10X Visium spatial transcriptomics data for mouse intestinal crypts.** The figure highlights the reconstruction process of a set of three ubiquitous cellular pathways starting from annotated stem cell compartments. We let it evolve for 12 steps. We then plot the original expression in the last column for comparison. Here we show the MNCA model with noise as it is the one that achieves the lowest MSE.

References

Badia-i Mompel, P., Vélez Santiago, J., Braunger, J., Geiss, C., Dimitrov, D., Müller-Dott, S., Taus, P., Dugourd, A., Holland, C. H., Ramirez Flores, R. O., et al. decoupler: ensemble of computational methods to infer biological activities from omics data. *Bioinformatics Advances*, 2 (1):vbac016, 2022.

Balzter, H., Braun, P. W., and Köhler, W. Cellular automata models for vegetation dynamics. *Ecological modelling*, 107(2-3):113–125, 1998.

Beumer, J. and Clevers, H. Cell fate specification and differentiation in the adult mammalian intestine. *Nature reviews Molecular cell biology*, 22(1):39–53, 2021.

Cess, C. G. and Finley, S. D. Calibrating agent-based models to tumor images using representation learning. *PLoS Computational Biology*, 19(4):e1011070, 2023.

Ermentrout, G. B. and Edelstein-Keshet, L. Cellular automata approaches to biological modeling. *Journal of theoretical Biology*, 160(1):97–133, 1993.

Gilpin, W. Cellular automata as convolutional neural networks. *Physical Review E*, 100(3):032402, 2019.

Grattarola, D., Livi, L., and Alippi, C. Learning graph cellular automata. *Advances in Neural Information Processing Systems*, 34:20983–20994, 2021.

Hernandez, A., Vilalta, A., and Moreno-Noguer, F. Neural cellular automata manifold. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 10020–10028, 2021.

Jang, E., Gu, S., and Poole, B. Categorical reparameterization with gumbel-softmax. *arXiv preprint arXiv:1611.01144*, 2016.

Kalkhof, J., Kühn, A., Frisch, Y., and Mukhopadhyay, A. Parameter-efficient diffusion with neural cellular automata. *npj Unconventional Computing*, 2(1):1–9, 2025.

Kingma, D. P. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980*, 2014.

Lewinsohn, M. A., Bedford, T., Müller, N. F., and Feder, A. F. State-dependent evolutionary models reveal modes of solid tumour growth. *Nature Ecology & Evolution*, 7 (4):581–596, 2023.

Menta, A., Archetti, A., and Matteucci, M. Latent neural cellular automata for resource-efficient image restoration. *arXiv preprint arXiv:2403.15525*, 2024.

Mordvintsev, A., Randazzo, E., Niklasson, E., and Levin, M. Growing neural cellular automata. *Distill*, 5(2):e23, 2020.

Noble, R., Burri, D., Le Sueur, C., Lemant, J., Viossat, Y., Kather, J. N., and Beerenwinkel, N. Spatial structure governs the mode of tumour evolution. *Nature ecology & evolution*, 6(2):207–217, 2022.

Palm, R. B., González-Duque, M., Sudhakaran, S., and Risi, S. Variational neural cellular automata. *arXiv preprint arXiv:2201.12360*, 2022.

Parigi, S. M., Larsson, L., Das, S., Ramirez Flores, R. O., Frede, A., Tripathi, K. P., Diaz, O. E., Selin, K., Morales, R. A., Luo, X., et al. The spatial transcriptomic landscape of the healing mouse intestine following damage. *Nature communications*, 13(1):828, 2022.

Richardson, A. D., Antal, T., Blythe, R. A., and Schumacher, L. J. Learning spatio-temporal patterns with neural cellular automata. *PLOS Computational Biology*, 20(4): e1011589, 2024.

Rogers, K. W. and Schier, A. F. Morphogen gradients: from generation to interpretation. *Annual review of cell and developmental biology*, 27(1):377–407, 2011.

Schubert, M., Klinger, B., Klünemann, M., Sieber, A., Uhlitz, F., Sauer, S., Garnett, M. J., Blüthgen, N., and Saez-Rodriguez, J. Perturbation-response genes reveal signaling footprints in cancer gene expression. *Nature communications*, 9(1):20, 2018.

Seferbekova, Z., Lomakin, A., Yates, L. R., and Gerstung, M. Spatial biology of cancer evolution. *Nature Reviews Genetics*, 24(5):295–313, 2023.

Sottoriva, A., Verhoeff, J. J., Borovski, T., McWeeney, S. K., Naumov, L., Medema, J. P., Sloot, P. M., and Vermeulen, L. Cancer stem cell tumor model reveals invasive morphology and increased phenotypical heterogeneity. *Cancer research*, 70(1):46–56, 2010.

Tari, H., Kessler, K., Trahearn, N., Werner, B., Vinci, M., Jones, C., and Sottoriva, A. Quantification of spatial subclonal interactions enhancing the invasive phenotype of pediatric glioma. *Cell Reports*, 40(9), 2022.

Tesfaldet, M., Nowrouzezahrai, D., and Pal, C. Attention-based neural cellular automata. *Advances in Neural Information Processing Systems*, 35:8174–8186, 2022.

Von Neumann, J., Burks, A. W., et al. Theory of self-reproducing automata. 1966.

Wolf, F. A., Angerer, P., and Theis, F. J. Scanpy: large-scale single-cell gene expression data analysis. *Genome biology*, 19:1–5, 2018.

Zhang, D., Choi, C., Kim, J., and Kim, Y. M. Learning to generate 3d shapes with generative cellular automata. *arXiv preprint arXiv:2103.04130*, 2021.

Zhang, D., Williams, F., Gojcic, Z., Kreis, K., Fidler, S., Kim, Y. M., and Kar, A. Outdoor scene extrapolation with hierarchical generative cellular automata. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pp. 20145–20154, 2024.

A. Synthetic Simulation of Tissue Growth

This model simulates the development and maintenance of tissue organization from an initial cluster of stem cells. It captures key biological features of stem cell-driven tissue organization, which is particularly relevant for studying systems such as intestinal crypts or clonal hematopoiesis.

The model incorporates several fundamental biological principles:

- **Hierarchical Cell Organization:** The tissue is organized in a hierarchy of cell types, from stem cells through intermediate progenitors to fully differentiated cells, reflecting the organization observed in many epithelial tissues.
- **Local Cell Interactions:** Cell fate decisions are influenced by the local cellular environment, mimicking the role of signaling niches in tissue organization.
- **Differential Division Rates:** Different cell types exhibit distinct proliferation rates, with stem cells and early progenitors showing higher division rates compared to differentiated cells.
- **Cell Type-Specific Survival:** The model implements different death and growth rates for each cell type, reflecting the biological reality where stem cells are more protected while differentiated cells undergo regular turnover.

The model implements a stochastic process where each cell is synchronously updated at each time step based on a set of kinetic parameters. A full description of the procedure is in Algorithm 1

Here we report the parameters we used in our experiment:

The system comprises five distinct cell types $\mathcal{T} = \{\text{STEM}, \text{INT1}, \text{INT2}, \text{DIFF1}, \text{DIFF2}\}$, with dynamics governed by division, death, and survival rates.

The division rates b are defined as: $b_{\text{stem}} = 0.8$, $b_{\text{int1}} = 0.5$, $b_{\text{int2}} = 0.5$, for differentiated cells b is set to 0. Death rates d are specified as: $d_{\text{stem}} = 0$, $d_{\text{int1}} = 0$, $d_{\text{int2}} = 0$, $d_{\text{diff1}} = 0.001$, $d_{\text{diff2}} = 0.001$. Survival rates s follow: $s_{\text{stem}} = 0$, $s_{\text{int1}} = 0$, $s\text{int2} = 0.01$, $s_{\text{diff1}} = 1.0$, $s_{\text{diff2}} = 1.0$.

The base differentiation probabilities are defined by matrix $\mathbf{D} \in \mathbb{R}^{5\times5}$:

$$\mathbf{D} = \begin{pmatrix} 0.3 & 0.8 & 0.0 & 0.0 & 0.0 \\ 0.1 & 0.2 & 0.8 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.2 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 1.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 1.0 \end{pmatrix} \tag{5}$$

where D_{ij} represents the rate of transitioning from type i to type j.

Cell-cell interactions are modeled through the interaction matrix $\mathbf{I} \in \mathbb{R}^{5\times5}$:

$$\mathbf{I} = \begin{pmatrix} 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.3 \\ 0.0 & 0.0 & 0.0 & 0.0 & 0.0 \end{pmatrix} \tag{6}$$

where I_{ij} modifies the differentiation rate based on neighboring cells of type j for cells of type i.

At each time step t, cells undergo stochastic events (division, death, or survival) with probabilities normalized by the total rate. For instance, for the cell type $stem$ the probability of dying would be: $P(\text{event}) = d_{stem}/(b_{stem} + d_{stem} + s_{stem})$. During division events, daughter cells may differentiate according to the probabilities in $\mathbf{D}$, modified by neighboring cells through $\mathbf{I}$. The division is allowed only on empty cells in the Moore neighbor; otherwise, the cell survives.

The model's simplicity and incorporation of key biological principles make it a useful tool for understanding. While it necessarily abstracts many biological details, it captures essential features that drive tissue organization and maintenance.

Algorithm 1 Tissue Growth Simulation

Input: grid size N, initial stem cells n_s
Input: cell rates $\mathbf{R} = \{b_\tau, d_\tau, s_\tau\}$ for each cell type $\tau \in \mathcal{T}$, transition matrices $\{\mathbf{D}, \mathbf{I}\}$
Initialize $G_0 \in \mathbb{Z}^{N \times N}$ with n_s stem cells at random positions
repeat
  for each cell c at position (x, y) in G_t **do**
    if $c \neq$ EMPTY **then**
      $\rho \sim \mathcal{U}(0, 1)$ {Sample uniform random variable}
      $\mathbf{p} = \mathbf{R}(c)/\|\mathbf{R}(c)\|_1$ {Normalize rates to probabilities}
      if $\rho < p_{death}$ **then**
        $G_{t+1}(x, y) \leftarrow$ EMPTY
      else if $\rho < p_{death} + p_{div}$ **then**
        $\mathcal{N} \leftarrow$ EmptyNeighbors(x, y)
        if $\mathcal{N} \neq \emptyset$ **then**
          Sample $(i, j) \sim$ Uniform$(\mathcal{N})$ {Pick a Random Empty Neighbour}
          $\mathbf{k} = \mathbf{B}(c) + \sum_{n \in \mathcal{N}} \mathbf{I}(n)$ {Generate a vector of rates for cell-type division}
          $\mathbf{k} = \mathbf{k}/\|\mathbf{k}\|_1$ {Normalize to probabilities}
          $G_{t+1}(i, j) \sim$ Categorical$(\mathbf{k})$
        end if
      end if
    end if
  end for
  $t \leftarrow t + 1$
until $t = T$

B. Evaluation Metrics for Neural Cellular Automata Models

We evaluate our Neural Cellular Automata (NCA) models using different complementary metrics that capture complementary aspects of the generated cellular patterns. Since our models aim to generate realistic tissue patterns with multiple cell types, we need metrics that assess both the statistical distribution of cell types and their spatial organization.

To compare the real and generated tissues, we employ three complementary measures:

- **Kullback-Leibler (KL) Divergence on cell type proportions:** $D_{KL}(P|Q) = \sum_i P(i) \log \frac{P(i)}{Q(i)}$ Where P is the true cell type distribution in the whole cohort and Q is the cell type distribution in the generated sample. This metric is useful for measuring how well our NCA models capture the correct proportions of different cell types.
- **Wasserstein Distance of Tissue Size Distribution:** For each real and generated data tissue we compute its size as simply the sum of all non-empty spots. We then compare the size distribution in real U and generated V using the 1-Wasserstein distance amongst empirical distributions, which for 1D distributions is simply:

$$W_1(U, V) = \int_{-\infty}^{\infty} |F_U(x) - F_V(x)| \, dx \tag{7}$$

  Where F is the empirical Cumulative Distribution Functions $F = \frac{1}{n} \sum_{i=1}^{n} \mathbf{1}(x_i \leq x)$
- **Wasserstein Distance of Tissue Border Size Distribution:** For each real and generated data tissue we compute a border size metric as:

$$B = \sum_{i,j} |(\nabla^2 M)_{i,j}| > \theta \tag{8}$$

  where M is the binary mask of cell occupancy (1 for cells, 0 for empty space), ∇^2 is the discrete Laplacian operator implemented as a 3×3 convolution kernel and $\theta = 0.25$ is a threshold parameter:

$$K = \frac{1}{8} \begin{bmatrix} -1 & -1 & -1 \\ -1 & 8 & -1 \\ -1 & -1 & -1 \end{bmatrix} \tag{9}$$

Table 2. **Model configurations across experiments.** Milestones and Gamma are the epoch and the multiplicative factor for the learning rate schedule, k stands for 10^3

Parameter	Tissue	Visium
Channels	6	14
Hidden dim.	128	256
Rules	5	4
Learning rate	1×10^{-3}	1×10^{-3}
Epochs	800	3000
Residual	No	Yes
Dropout	0	0
Milestones	[500]	[1.5k,2k,2.5k]
Gamma	0.1	0.1
Filters	∇_x,∇_y	∇^2

The combination of these metrics allows us to compare different NCA architectures and evaluate their ability to capture both the statistical and structural properties of real biological tissues. This is essential for developing NCA models that can not only match cell type distributions but also generate spatially coherent and biologically plausible tissue patterns.

We then again compute the 1-Wasserstein distance between the border statistics of real and generated data.

C. Model Architectures and Parameters

We maintained consistent neural network architectures across our two experiments (Visium spatial transcriptomics, and synthetic biological simulations). All models were trained using the Adam optimizer (Kingma, 2014) and with a Multi-Step LR Scheduling, where every time the training reaches a set of epoch milestones, the learning rate gets multiplied by a scaling factor gamma.

To keep the model simple and all rule functions ϕ_k follow a standard architecture with two 1 by 1 convolutional layers:

$$h = \text{ReLU}(\text{Conv}_{1\times1}\text{Cat}(s_i^t, \nabla_x s_i^t, \nabla_y s_i^t)) \tag{10}$$

$$s_i^{t+1} = s_i^t + \text{Conv}_{1\times1}(\text{Cat}(h, x_k)) \tag{11}$$

where s_i^t represents the state at time t, ∇_x and ∇_y are Sobel filters for spatial derivatives, x_k is the gaussian noise, and Cat denotes channel-wise concatenation. This is the implementation of Equation 4c, in case of 3 the implementation is the same but with $\text{Conv}_{1\times1}(\text{Cat}(h))$ instead of $\text{Conv}_{1\times1}(\text{Cat}(h, x_k))$, having now just one of those networks. The tissue in experiment 4.2 is distorted and rotated in the slide; there is no clear up and down, as such, instead of the Sobel filters we use a discrete Laplacian filter ∇^2.

The network π is implemented by a network of the same type with the only difference that the input in this case is just the current cell value x_t and not $\text{Cat}(x_t, \nabla_x x_t, \nabla_y, x_t)$. We used the Mean Squared Error (MSE) against the target as our loss in all experiments.

We report the parameters used in the experiment in Table 2. All experiments were run on a single NVIDIA Tesla V100-SXM2-32GB.

D. Training routines for the Neural Cellular Automata

We use a simple algorithm (described in Algorithm 2) to train our automata on the biological time series data. For each epoch with a specific time-window size, the algorithm samples a random part of the time series and learns to reconstruct it by computing the loss with the original realization every τ steps (which has been set to 1 for our experiments). The training algorithm for Neural Cellular Automata (NCA) introduced in (Mordvintsev et al., 2020) is actually more complicated then the one we used above, mainly because in the task of image morphogenesis, the model has to evolve without supervision for

a long time.

Algorithm 2 NCA Training for Biological Time-Series

Input: target sequences $\{S_1, ..., S_n\}$,sequence length T
Input: window size w, number of cell types K, epochs E,
Input: evolution steps of the automata τ, number of tissue samples M, small stability constant ϵ
Initialize model parameters θ randomly
Initialize optimizer
for epoch = 1 **to** E **do**
  $t_{start} \sim \mathcal{U}(0, T - w)$ {Sample random window}
  for $t = t_{start}$ **to** $t_{start} + w$ **step** τ **do**
    $X_t \leftarrow S_t$ {Encode states}
    $Y_t \leftarrow S_{t+\tau}$ {Future states}
    for $t_{pred} = t$ **to** $t + \tau$ **step** 1 **do**
      if $t_{pred} == t$ **then**
        $\hat{Y}_t \leftarrow f_\theta(X_t)$ {NCA prediction, from input}
      else
        $\hat{Y}_t \leftarrow f_\theta(\hat{Y}_t)$ {NCA prediction, from evolved input}
      end if
    end for
    // Compute loss and update
    $\mathcal{L} \leftarrow \frac{1}{M}\text{MSE}(Y_t, \hat{Y}_t)$
    $\nabla\theta \leftarrow \nabla_\theta \mathcal{L}$
    $\nabla\theta \leftarrow \frac{\nabla\theta}{||\nabla\theta||+\epsilon}$ {Normalize gradients}
    Update θ using optimizer
  end for
end for

To evaluate the impact of rule complexity, we analyze the KL divergence between the simulated and target cell-type distributions as a function of the number of mixture rules. As illustrated in Figure 5, models with an increasing number of rules tend to exhibit lower KL divergence, suggesting that greater rule diversity enhances the model's ability to approximate the target distribution.

However, this improvement is not strictly linear, and we observe diminishing returns beyond a certain number of rules. This saturation effect suggests that while additional mixture components increase flexibility, excessive complexity does not necessarily translate into significant performance gains. Furthermore, we find that introducing internal stochasticity in the model slightly reduces KL divergence for this task. These findings support the idea that an excessively high number of rules can be detrimental, as performance improvements do not sufficiently compensate the increased computational cost.

E. Analysis of Spatial Transcriptomics Data

We evaluated our models on Visium spatial transcriptomics data from mouse intestinal tissue. The intestine is an excellent system for studying tissue evolution due to its rapid and continuous cellular turnover, driven by stem cells located in the crypts. These stem cells divide and differentiate into various cell types as they migrate upward to form the villi, creating a dynamic, self-renewing structure. This property means that the spatial dimensions correspond also to a differentiation pseudotime, that we can use to generate time-series "like" data, by considering the cells at the bottom of the crypt as the ones starting at time zero and then slowly going forward in time by adding cells towards the villum. The differentiation process is tightly regulated by a complex network of signaling interactions involving different cell types and compartments, which collectively determine cell fate and maintain tissue homeostasis. This intricate coordination makes the intestine a powerful model for exploring how spatial and temporal signals influence tissue development and evolution. The main idea is summarized in a cartoon in Figure 6. Figure 7, on the other end, shows the time series in the actual slide.

Visium spatial transcriptomics is a spot-based technology that maps gene expression directly onto tissue sections by combining high-resolution imaging with RNA sequencing. Each spot on the array has a diameter of approximately 55 micrometers, capturing RNA from a defined area of the tissue that typically includes multiple cells, allowing spatially

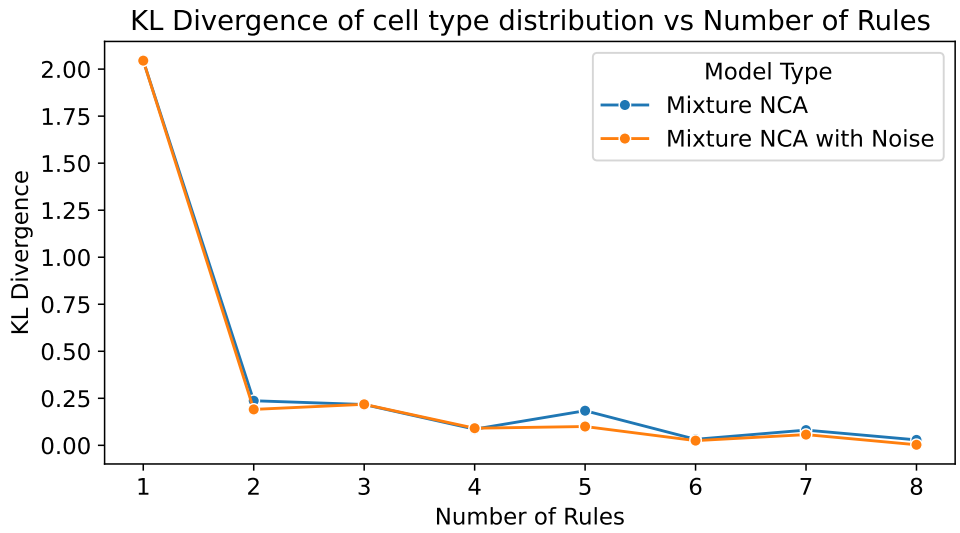

Figure 5. **KL divergence of the cell-type distribution as a function of the number of rules used in the Mixture NCA model.** As the number of rules increases, the divergence between the simulated and target distributions decreases, indicating improved alignment with the expected cell-type dynamics.

resolved transcriptomic profiling of a high number of genes. The dataset consists of approximately 5,000 spots per sample, each containing expression measurements for over 20,000 genes.

Data processing followed standard single-cell analysis protocols using the Scanpy package (Wolf et al., 2018):

1. Normalization of counts to 10,000 reads per spot followed by log1p transformation
2. Feature selection using the top 3000 genes in terms of variance
3. PCA dimensionality reduction to 50 principal components
4. Neighborhood graph construction using 10 nearest neighbors
5. Leiden clustering for initial cell type annotation

Cell types were annotated using established marker genes for intestinal tissue:

- Epithelial cells: Epcam, Krt8, Krt18
- Goblet cells: Muc2, Tff3, Spink4
- Enteroendocrine cells: Chga, Chgb, Sst
- Stem cells: Lgr5, Ascl2, Olfm4
- Immune cells: Cd3e, Cd19
- Stromal cells: Col1a1, Vim, Acta2

To get a subset of putative stem spots (spots with a putative high content of stem cells) we looked at the stem marker expression in the different clusters and took the ones with some signal. As the proportion of stem cells in the intestine is quite low, we expect the signal to be relatively small, and indeed we find just two clusters with a detectable signal of stem markers (Figure 8).

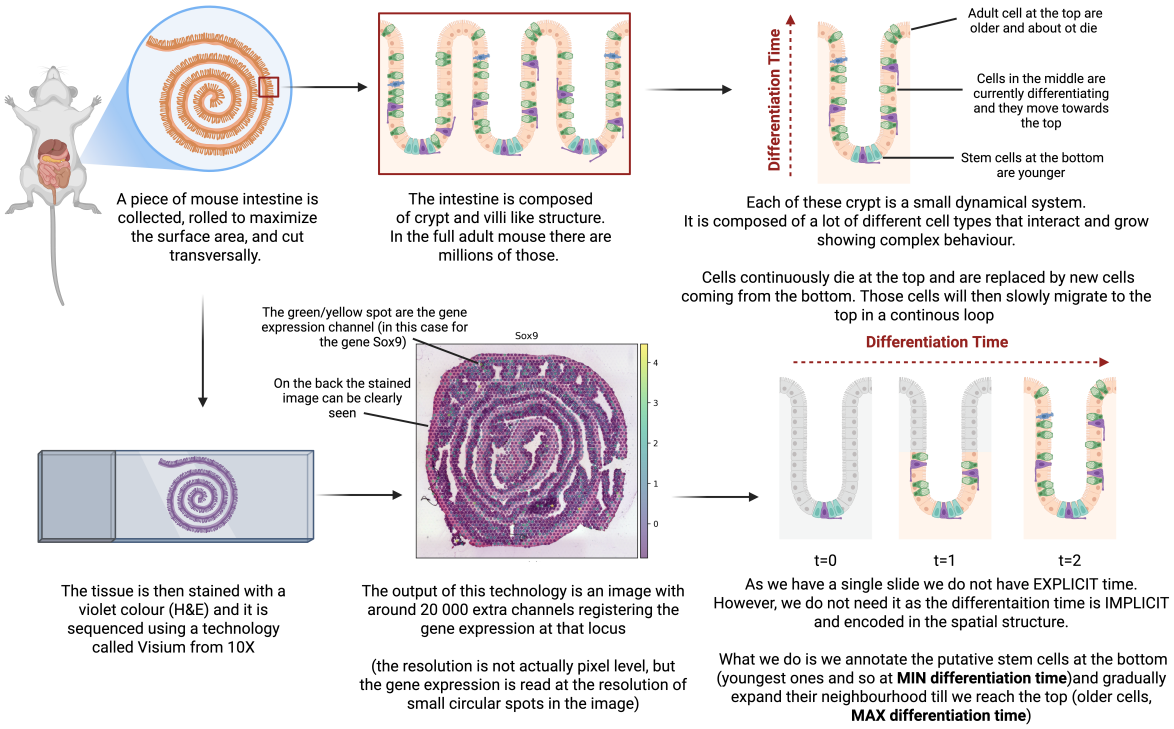

Figure 6. **Overview of spatial transcriptomics on mouse intestine using 10X Visium.** The rolled tissue is sectioned, stained, and sequenced to capture spatial gene expression. Each crypt encodes a differentiation trajectory from stem cells at the base to mature cells at the top, allowing inference of pseudo-temporal dynamics from spatial structure.

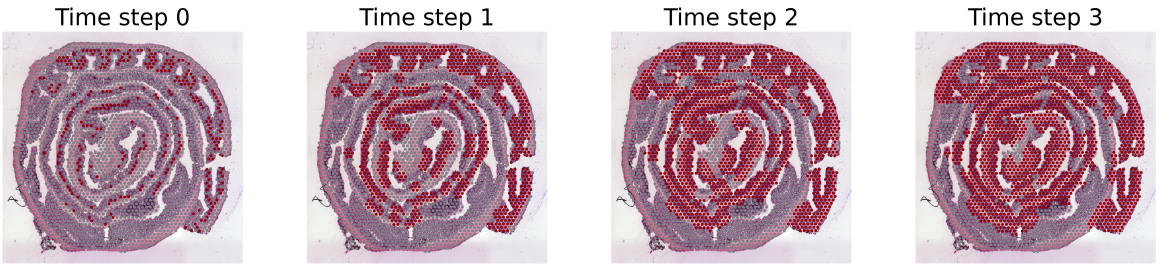

Figure 7. **Time series extraction from static tissue.** Only the first 4 steps are shown here, but the process continues until all the spots in the slide have been covered. Grey dots are empty spots, while red spots are alive cells at a given time.

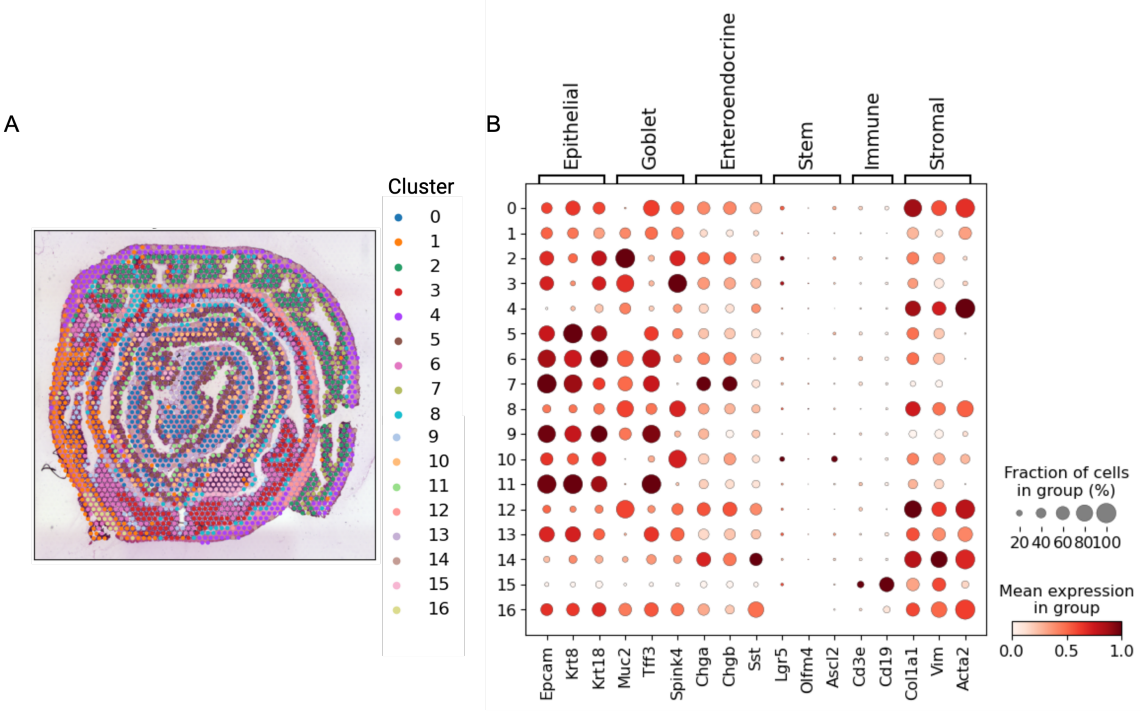

Figure 8. **Leiden clustering and gene marker expression.** We performed Leiden clustering using the first 50 components of the PCA, we then manually looked at Stem Cell Marker expression to select putatively enriched stem spots.

To incorporate a dynamical system into our MNCA framework, we drew inspiration from the continuous renewal process of intestinal villi, which is sustained by an underlying crypt of proliferating stem cells. In this context, we manually annotated the region corresponding to the putative stem cell compartment within the tissue and used it as the initial condition for our simulations. The model was then evaluated on its capacity to accurately reconstruct the spatial organization and differentiation trajectories of the intestinal epithelium.

To simulate a biologically plausible developmental trajectory, we adopted an iterative procedure: beginning from the annotated stem cell compartment, we progressively expanded the simulated domain by sequentially incorporating neighboring cells at each time step. This approach effectively generated a synthetic time series that mimics live imaging or in vivo sampling of tissue development, enabling us to assess the MNCA's ability to capture both structural and temporal aspects of intestinal regeneration.

We did not use the full gene expression but we focused on a set of 14 curated and fundamental pathways from (Schubert et al., 2018). This resource provides curated pathways and their associated target genes, with interaction weights derived from perturbation experiments; for this example, we use the mouse weights and the top 200 most responsive genes ranked by p-value. The pathways include Androgen (male reproductive organ development), EGFR (cell growth, survival, and migration), Estrogen (female reproductive organ growth), Hypoxia (angiogenesis under low oxygen), JAK-STAT (immune response and cell division), MAPK (cell growth and proliferation), NFkB (immune response and cytokine production), p53 (tumor suppression and DNA repair), PI3K (growth and proliferation), TGFb (tissue development and repair), TNFa (immune surveillance and infection protection), Trail (apoptosis induction), VEGF (angiogenesis and vascular permeability), and WNT (morphogenesis and tissue repair).

To get the pathway activity for all 14 pathways we perform a multivariate linear regression as implemented in (Badia-i Mompel et al., 2022). Namely, given a matrix of weights $W \in \mathbb{R}^{G \times P}$ for each gene and pathway, where P is the number

of pathways and G the number of genes and the expression matrix $Y \in \mathbb{R}^{N \times G}$, where N is the number of spots, the final activity score is the $X \in \mathbb{R}^{P \times S}$ matrix of coefficients from a simple multivariate linear model $Y = WX$.

F. Comparison with Approximate Bayesian Computation

We implement a standard Approximate Bayesian Computation (ABC) approach to infer the parameters of our agent-based tissue growth model and compare it with the MNCA results. The parameter space Θ is the same as the original model in Appendix A, here we report the sampling distribution:

- division rates ($\vec{b} \in \mathbb{R}^5 \sim \text{Gamma}(1, 0.1)$)
- death rates ($\vec{d} \in \mathbb{R}^5 \sim \text{Gamma}(1, 0.01)$)
- survival rates ($\vec{s} \in \mathbb{R}^5 \sim \text{Gamma}(1, 0.1)$)
- differentiation rates ($D \in \mathbb{R}^{5\times5} \sim \text{Gamma}(1, 0.1)$)
- cell-cell interaction strengths ($I \in \mathbb{R}^{5\times5} \sim \mathcal{N}(0, 1)$)

Algorithm 3 Parallel ABC for Tissue Growth Model

Require: Observed data $\mathcal{D}$, number of particles N, acceptance threshold ϵ, summary statistic type S
Ensure: Estimated parameters θ^*
Initialize empty sets $\mathcal{A}_\theta, \mathcal{A}_\delta$ for accepted parameters and distances
for $i = 1$ to N **do**
  Sample $\theta_i \sim p(\theta)$ from prior distributions
  Simulate tissue growth $x_i \sim f(\cdot|\theta_i)$ for T steps
  Compute summary statistic $s_i = S(x_i)$
  Calculate distance $\delta_i = Dist(s_i, S(\mathcal{D}))$
  if $d_i < \epsilon$ **then**
    $\mathcal{A}_\theta \leftarrow \mathcal{A}_\theta \cup \{\theta_i\}$
    $\mathcal{A}_\delta \leftarrow \mathcal{A}_\delta \cup \{\delta_i\}$
  end if
end for
Compute weights $w_i = 1/\delta_i$, normalized
Return $\theta^* = \sum_i w_i \theta_i$ for $\theta_i \in \mathcal{A}_\theta$

We trained three different models each with specific summary statistics to compare simulated and observed data:

1. **Cell-type Distribution**: Captures the global proportion of each cell type, including empty spaces. The distance between distributions is computed using the Wasserstein metric W_1. This statistic provides a high-level view of tissue composition but does not capture spatial organization.
2. **Neighborhood Composition**: Computes the average composition of 3×3 neighborhoods around each position, including empty spaces. This metric captures local spatial patterns and cell-type clustering, using the Wasserstein distance for comparison.
3. **Cell-type Correlation Matrix**: Quantifies pairwise correlations between spatial distributions of cell types. Each entry R_{ij} represents the Pearson correlation coefficient between the binary masks of types i and j, with positive values indicating co-occurrence and negative values suggesting spatial segregation. The distance between correlation matrices is computed using the normalized Frobenius norm $\|R_i - R_\mathcal{D}\|_F/\sqrt{2}$.

Parameters are accepted if their distance is below the threshold ϵ, and final estimates are computed as weighted averages of accepted particles, with weights inversely proportional to their distances. The ϵ in this case have been chosen to accept

Table 3. **Comparison of ABC inference on agent-based models of the simulation**

MODEL	KL-DIV	χ^2	SIZE	BORDER-W
ABM MODEL PROPORTION	**0.152 ±0.002**	**0.132 ±0.001**	**0.241 ±0.010**	**0.054 ±0.006**
ABM MODEL NEIGHBORHOOD	0.857 ±0.010	0.488 ±0.004	0.489 ±0.025	0.284 ±0.010
ABM MODEL CORRELATION	0.386 ±0.010	0.285 ±0.005	**0.241 ±0.012**	0.055 ±0.009

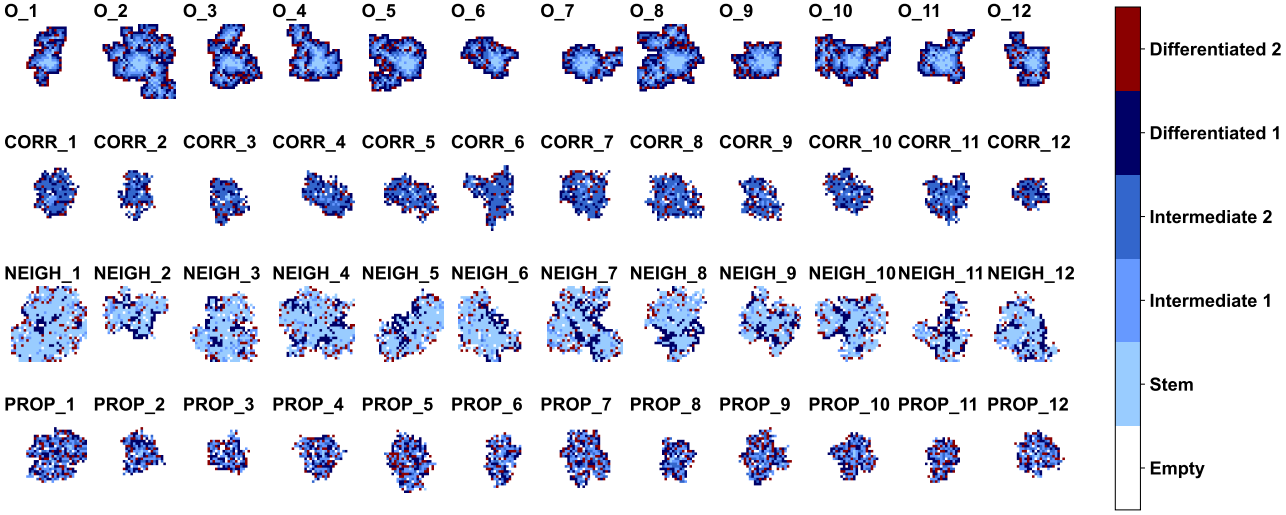

Figure 9. **Simulated tissues by different agent-based models trained with ABC on the simulated dataset.** Top row is 10 tissues from the training data. The second row are tissues generated by the ABC schema with type correlation as summary statistics. The third row are tissues generated by the ABC schema with neighborhood composition as summary statistics. The last row has cell-type distribution as summary statistics.

approximately 10% of the samples (respectively [0.04, 0.4, 0.52]). For the first model, we generated 5000 samples as this is the fastest statistic to compute, while for the others we drew 1000 samples.

We present the results in Table 3, highlighting how performance is significantly influenced by the choice of statistics used for parameter inference. Interestingly, simple cell-type proportions yield the best results, not only in terms of accuracy but also in maintaining consistency across tissue borders and overall size. These results are comparable to our MNCA approach, though in this case, we had to fully specify the model and fine-tune both the statistics and acceptance threshold. It is also crucial to note that our evaluation is based on the KL divergence of cell proportions, which is directly tied to the statistics used for ABC—specifically, the Wasserstein distance between cell-type proportions.

However, the limitations of this approach are well illustrated in Figure 9. A naive selection of statistics may produce models that generate seemingly accurate summary statistics, yet fail to capture the intricate spatial characteristics of real tissue. As a result, while the summary metrics appear realistic, the simulated tissue structure diverges significantly from the actual one. Conversely, models that better preserve spatial coherence tend to exhibit substantial distortions in cell-type proportions.