

---

# Learning Pairwise Potentials via Differentiable Recurrent Dynamics

---

**Kenji Komiya**

NTT, Inc.  
Atsugi, Kanagawa 2430124  
kenji.komiya@ntt.com

**Andrew Kailiang Jin**

Georgia Institute of Technology  
Atlanta, GA 30332  
ajin40@gatech.edu

**Ryo Nishikimi**

NTT, Inc.  
Atsugi, Kanagawa 2430124  
ryo.nishikimi@ntt.com

**Kunio Kashino**

NTT, Inc.  
Atsugi, Kanagawa 2430124  
kunio.kashino@ntt.com

## Abstract

This paper describes a novel approach to optimizing parameters defining the behavior of pairwise potential models, which can simulate cell migration. Determining appropriate parameters remains a major challenge due to measurement difficulties and high computational cost. To solve this problem, we propose a fully differentiable framework that integrates metrics for evaluating cell positional configurations with recurrent intercellular interaction dynamics. The experimental results showed that our method successfully estimated parameters that reproduce cell positional configurations in both synthetic and biological datasets.

## 1 Introduction

Cell migration governs diverse biological phenomena, from development to diseases [22, 15]. Examples include germ layer separation [14, 6], cardiac tube formation [7, 18], tumor invasion [10, 23], and organoid morphogenesis [16, 5]. These migrations establish characteristic cellular positional configurations that are essential for tissue function. Modeling such dynamics is thus key to understanding, predicting, and controlling morphogenesis.

The pairwise potential model is a common mathematical framework for describing cellular migration [3, 12]. Each cell generates a potential field combining short-range repulsion with mid-range attraction, and migration arises from the sum of field gradients. This class of models has successfully simulated various biological phenomena, including tissue organization [8] and spheroid formation [21], offering a natural yet tractable description of multicellular dynamics.

Despite these successes, determining parameter values of pairwise potential models remains challenging. Direct measurement is extremely difficult because the intercellular forces are at the pico-Newton scale, live tissues are highly dynamic, and available methods (e.g., laser ablation or microdroplet sensors) are invasive and provide only indirect estimates [2, 8, 20]. In practice, available data are usually cellular positions, either as static snapshots or time-series trajectories. A common strategy for parameter optimization is grid search on these data, but it is computationally expensive [13, 11, 9]. A promising direction is to formulate the model as a recurrent system, enabling the use of snapshot and time-series observations as training data, and to make the model differentiable for efficient gradient-based optimization.

In this study, we ask: *Can pairwise potential models be learned in this recurrent and differentiable form?* To answer this, we: (i) propose a learning framework that optimizes potential parameters via a

differentiable discrepancy metric, and (ii) validate it on synthetic datasets and on zebrafish early heart development data to demonstrate applicability.

## 2 Method

### 2.1 Preliminaries: Pairwise Potential Model

We define the set of cell indices as  $\mathcal{C}$  and consider the time interval  $[t_0, T]$  corresponding to a cellular transition. For each cell  $i \in \mathcal{C}$ , its position at time  $t$  ( $t_0 < t < T$ ) is denoted by  $\mathbf{x}_i(t) \in \mathbb{R}^{n_{sd}}$ , where  $n_{sd}$  is the spatial dimension. Given initial positions  $\mathbf{s}(t_0) = \{\mathbf{s}_i(t_0)\}_{i \in \mathcal{C}}$ , we define the simulated position of cell  $i$  at time  $t_n$  ( $t_0 < t_n < T$ ) using a potential model  $P$  parameterized by  $\boldsymbol{\theta}_{ij}$  and Euler-based time integration as

$$\hat{\mathbf{x}}_i(t_n, \boldsymbol{\Theta}) \approx \mathbf{f}_i(\hat{\mathbf{x}}(t_{n-1}), \boldsymbol{\Theta}) \quad (1)$$

$$= \sum_{j \in \mathcal{C}, j \neq i} \nabla P(\mathbf{d}_{ij}(\hat{\mathbf{x}}(t_{n-1})), \boldsymbol{\theta}_{ij}) \delta t + \hat{\mathbf{x}}_i(t_{n-1}, \boldsymbol{\Theta}), \quad (2)$$

where  $\boldsymbol{\Theta} = \{\boldsymbol{\theta}_{ij}\}_{i,j \in \mathcal{C}}$  is the set of all parameters,  $\hat{\mathbf{x}}(t) = \{\hat{\mathbf{x}}_i(t)\}_{i \in \mathcal{C}}$  is the simulated cellular positions at time  $t$ ,  $\mathbf{d}_{ij}$  is the relative position vector from  $i$ -th to  $j$ -th cells,  $\delta t$  is the time step, and  $t_n = t_0 + n\delta t$ . By recursively applying  $\mathbf{f}_i$ , the simulated position at step  $n$  can be approximated as

$$\hat{\mathbf{x}}_i(t_n, \boldsymbol{\Theta}) \approx \mathbf{f}_i^n(\mathbf{s}(t_0), \boldsymbol{\Theta}), \quad (3)$$

where  $\mathbf{f}_i^n$  denotes the  $n$ -fold recursive application of the time integration. As the pairwise potential model, we use a Morse-like form defined by

$$P(\mathbf{d}_{ij}, \boldsymbol{\theta}_{ij}) = R_{ij} \exp\left(-\frac{\|\mathbf{d}_{ij}\|}{r_{ij}}\right) - A_{ij} \exp\left(-\frac{\|\mathbf{d}_{ij}\|}{a_{ij}}\right), \quad (4)$$

where the parameter set is given by  $\boldsymbol{\theta}_{ij} = \{R_{ij}, r_{ij}, A_{ij}, a_{ij}\}$ . The first and second terms represent short-range repulsion and mid-range attraction between cells  $i$  and  $j$ .

### 2.2 Our optimization strategy

Our goal is to learn the model parameters  $\boldsymbol{\Theta}$  that can reproduce the given target configurations  $\mathbf{x}^{\text{target}}(t) = \{\mathbf{x}_i^{\text{target}}(t)\}_{i \in \mathcal{C}}$ , where  $\mathbf{x}_i^{\text{target}}(t)$  is the  $i$ -th target cell position at time  $t$ , by minimizing a discrepancy metric between the simulated configurations  $\hat{\mathbf{x}}(t, \boldsymbol{\Theta}) = \{\hat{\mathbf{x}}_i(t, \boldsymbol{\Theta})\}_{i \in \mathcal{C}}$  and the target configurations. We use two discrepancy metrics: the Euclidean distance and the Sinkhorn divergence.

The Euclidean distance is given by

$$\mathcal{L}_{\text{euclidean}} = \sum_{t \in \mathcal{T}} \sum_{i \in \mathcal{C}} |\mathbf{x}_i^{\text{target}}(t) - \hat{\mathbf{x}}_i(t, \boldsymbol{\Theta})|, \quad (5)$$

where  $\mathcal{T}$  denotes the set of time points at which target configurations are given. This metric assumes that each target cell has a corresponding cell simulated by the migration of the initial cells and, therefore, requires cell-level positional matching over time. In this study, the initial cell corresponding to each target cell is obtained by solving the linear assignment problem as follows:

$$\min_{W \in \{0,1\}^{|\mathcal{C}| \times |\mathcal{C}|}} \sum_{i,j \in \mathcal{C}} |\mathbf{x}_i^{\text{target}}(t) - \mathbf{s}_j(t_0)| w_{ij}, \quad (6)$$

subject to

$$\sum_{j \in \mathcal{C}} w_{ij} = 1, \quad \sum_{i \in \mathcal{C}} w_{ij} = 1, \quad (7)$$

where  $|\mathcal{C}|$  indicates the number of the cells.  $W = (w_{ij})$  is a matrix where each element is either 0 or 1, and  $w_{ij} = 1$  means that the  $i$ -th target cell  $\mathbf{x}_i^{\text{target}}(t)$  are assigned to the  $j$ -th initial cell  $\mathbf{s}_j(t_0)$ .

The Sinkhorn divergence enables a smooth, differentiable, and noise-robust comparison of configurations and is well-suited for empirical point clouds [19]. Unlike the Euclidean distance, it does not require pointwise correspondence between target and simulated cells. The corresponding loss function is defined as

$$\mathcal{L}_{\text{sinkhorn}} = \sum_{t \in \mathcal{T}} \mathcal{S}_\varepsilon(\mathbf{x}^{\text{target}}(t), \hat{\mathbf{x}}(t, \boldsymbol{\Theta})), \quad (8)$$

where  $\mathcal{S}_\varepsilon(\cdot, \cdot)$  denotes the entropy-regularized Sinkhorn divergence with regularization parameter  $\varepsilon$ . When combined with Euler-based time integration and the pairwise Morse potential, both metrics are differentiable with respect to  $\Theta$ . Therefore, the parameters  $\Theta$  can be learned using gradient-based methods via automatic differentiation, like recurrent machine learning models.

### 2.3 Datasets and Experimental Settings

We used two types of data: synthetic data and biological data based on cell tracking. We synthesized three characteristic positional configurations: a ring, a square, and a triangle. These shapes are loosely inspired by biological morphologies, such as the hollow spherical structures. In the ring configuration, cells are placed within the annular region between two concentric circles with radii of 0.9 and 1.0. In the square configuration, cells are located within the area bounded by two concentric equilateral squares with side lengths of 1.5 and 1.6. In the triangle configuration, cells are positioned in the region between two concentric equilateral triangles with side lengths of 1.65 and 1.75.  $|\mathcal{C}| = 2000$  cells are randomly distributed within the respective target regions. We validated two migration scenarios: (i) single target: cells initially placed randomly migrate to a single configuration (ring, square, or triangle), and (ii) sequential target: cells migrate from random to ring and then to triangle. We used the inside ratio, which is the proportion of the number of cells within the target region to the total number of cells, as an evaluation metric. We also applied the proposed framework to the cell tracking positional configuration from zebrafish heart development. We focus on the specific development stage, which occurs from 20 to 21.5 hours post-fertilization. During this stage, cells transition from a ring to a cone-shaped configuration. The model learned cellular trajectories every 10 minutes from the initial cellular position. Target cell positions at each time point were obtained by extracting nuclear coordinates from imaging data published in a previous study [7].

For synthetic data, we tested the Euclidean loss and Sinkhorn divergence loss, while the Euclidean loss was used for the biological data. We compared two parameterization schemes: an individual scheme, where each cell has unique parameters ( $|\mathcal{C}| \times 4$  parameters), and a mutual scheme, where each cell pair has unique parameters ( $|\mathcal{C}|^2 \times 4$  parameters). The individual scheme represents a fully isotropic interaction, whereas the mutual scheme can capture pseudo-anisotropic interactions by allowing the potential to vary across different cell pairs. Synthetic targets used  $\delta t = 0.2$  hour with five steps, while biological targets used  $\delta t = 1/6$  hour with nine steps. All experiments used Adam (lr=0.01 for synthetic and lr=0.1 for biological targets) for 1000/3000 epochs.

## 3 Result and Discussion

Figure 1 illustrates the learning results for the triangle target. The loss monotonically decreases as learning progresses. The learned smooth potential distribution drives cellular configurations, and these migrations change the potential distribution. These recursive updates realize the target triangle configurations. Table 1 shows the evaluation results with different learning conditions. For (i) the single target, ten of twelve tested scenarios achieved high fidelity (inside ratio  $> 0.8$ ). The other two scenarios with the triangle target failed their learning. For (ii) the sequential target, only the Euclidean-based optimization with mutual parameters succeeded in reproducing both targets sequentially. If we can set a proper metric and number of parameters for each scenario, our proposed framework can learn efficiently its potential through differential-based optimizations.

Figure 2 shows the results of applying our framework to the zebrafish heart development dataset. Our framework stably learned parameters in this noisy observed dataset. However, there were differences depending on the size of the parameter space. In the individual parameter setting, the Euclidean loss increased monotonically over time steps (Figure 2(a), red), and the simulated trajectories deviated from the ground truth (Figure 2(b)). In contrast, in the mutual parameter setting, the Euclidean loss remained low over time (Figure 2(a), blue), and the simulated trajectories closely matched the ground truth (Figure 2(c)). Some previous research has reported anisotropic cellular behavior in heart development [1, 17, 4], and, as mentioned in subsection 2.3, the mutual scheme has the capacity for pseudo-anisotropic interaction. This difference thus highlights both the anisotropic nature of cellular interactions and the importance of designing an appropriate parameter space for capturing such target phenomena.

Figure 3 shows the pair plot of the learned parameters in the zebrafish heart development model. The histograms of the parameters  $r$  and  $A$  exhibit two distinct peaks, indicating the presence of two predominant parameter regimes. Consistently, their joint distribution displays a clear bimodal structure, suggesting that these parameters capture two characteristic modes of cell-cell interaction.

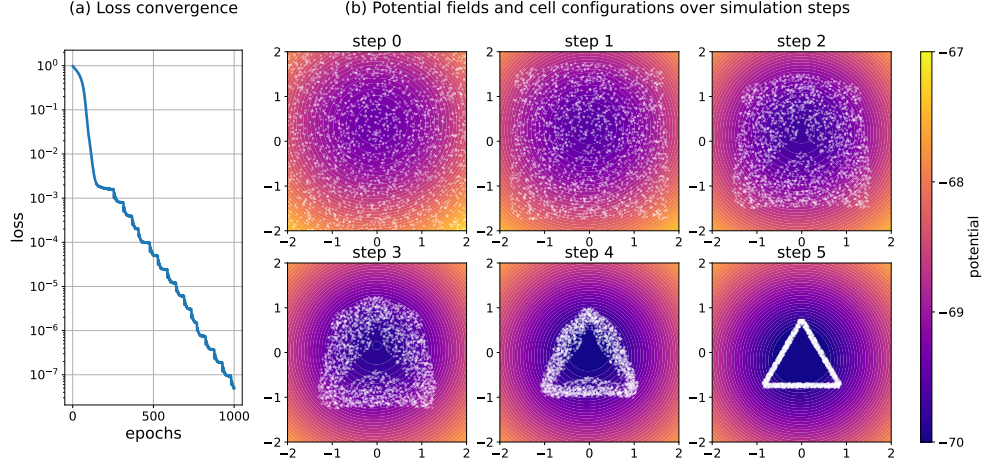


Figure 1: An example result of learning under mutual parameters with Euclidean metric

Table 1: Unified results for single-target and sequential ring→triangle optimizations under Euclidean and Sinkhorn metrics. “Inside ratio (2nd)” is only applicable to the sequential target; for single targets, it is shown as “-”.

Metric	Target	Parameter	Best loss ↓	Inside ratio (1st) ↑	Inside ratio (2nd) ↑
Euclidean	ring	individual	$6.27 \times 10^{-2}$	0.97	—
Euclidean	ring	mutual	$2.53 \times 10^{-8}$	1.00	—
Euclidean	square	individual	$3.53 \times 10^{-2}$	0.84	—
Euclidean	square	mutual	$5.24 \times 10^{-8}$	1.00	—
Euclidean	triangle	individual	$1.46 \times 10^{-1}$	0.54	—
Euclidean	triangle	mutual	$4.80 \times 10^{-8}$	0.97	—
Sinkhorn	ring	individual	$2.68 \times 10^{-2}$	1.00	—
Sinkhorn	ring	mutual	$2.58 \times 10^{-2}$	1.00	—
Sinkhorn	square	individual	$2.68 \times 10^{-2}$	1.00	—
Sinkhorn	square	mutual	$2.60 \times 10^{-2}$	0.97	—
Sinkhorn	triangle	individual	$2.98 \times 10^{-2}$	0.99	—
Sinkhorn	triangle	mutual	$2.72 \times 10^{-2}$	0.51	—
Euclidean	ring→triangle	individual	$2.03 \times 10^{-1}$	0.26	0.48
Euclidean	ring→triangle	mutual	$1.95 \times 10^{-2}$	0.91	0.95
Sinkhorn	ring→triangle	individual	$1.07 \times 10^{-1}$	0.24	0.23
Sinkhorn	ring→triangle	mutual	$5.33 \times 10^{-2}$	1.00	0.70

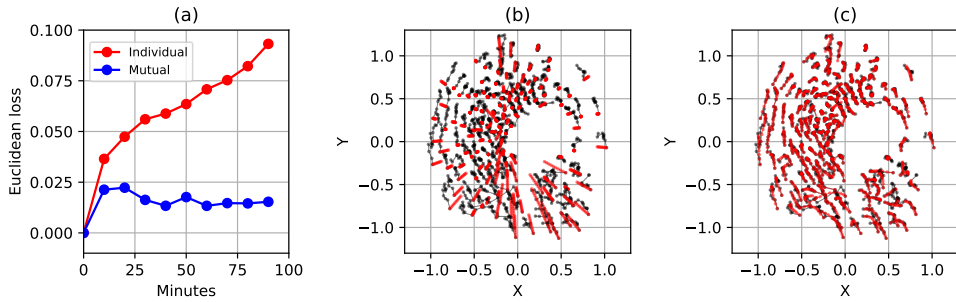


Figure 2: Results of the zebrafish heart development simulation from 20 to 21.5 hours post-fertilization. (a) Time-dependent Euclidean loss with ground truth for individual and mutual parameters. (b) Ground truth (black) and simulated trajectories with individual parameters (red). (c) Ground truth (black) and simulated trajectories with mutual parameters (red).

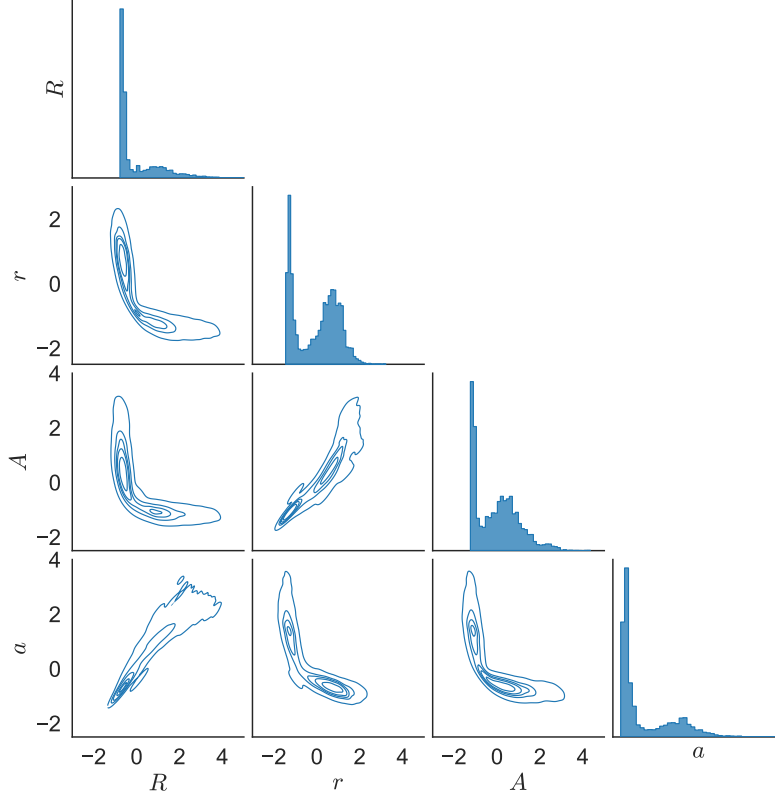


Figure 3: Pair plot of the learned parameters in the zebrafish heart development model. The histograms along the diagonal show the distributions of individual parameters  $R$ ,  $r$ ,  $A$ , and  $a$  while the off-diagonal panels display the estimated kernel density contours for each parameter pairwise combination. All parameters were standardized prior to visualization.

This implies that the learned potentials represent two groups of interactions: a repulsion-dominated group and an adhesion-dominated group. These two regimes may correspond to the two cellular populations reported in previous experimental studies, which differ in their intercellular rearrangements and cell shape dynamics [7]. Taken together, these findings suggest that our framework provides a cell-dynamics-informed approach for identifying and classifying distinct cell subpopulations.

## 4 Conclusion

We introduced a differentiable learning framework for pairwise potential models in multicellular dynamics and demonstrated its applicability using both synthetic and biological datasets. The results confirmed that pairwise potential models can indeed be learned in a recurrent and differentiable form. Moreover, the learned parameters provide interpretable insights, enabling the classification of cells based on their dynamic behaviors. Several issues remain to be addressed in this study. First, our experiments were limited to models with a fixed number of parameters and specific potential functions. A theoretical understanding of the limit of this framework’s applicability—such as the minimum number of parameters required to reproduce a given spatial configuration—remains an open problem. Clarifying this relationship could lead to more efficient approaches for designing desired cell arrangements in real biological systems. Second, while we employed an isotropic potential model to approximate an anisotropic biological process, further research is needed to incorporate inherently anisotropic interactions into a differentiable recurrent form. Developing such models could expand the biological realism of our framework. Finally, extending validation to more complex biological datasets and experimental systems will be essential for confirming the robustness of the proposed approach. We believe that the proposed framework represents a step toward developing interpretable, learnable, and generalizable models for multicellular systems, thereby contributing to the broader goals of bioengineering, including tissue morphogenesis and organoid development.

## Acknowledgment

We thank Professor Melissa Kemp, Andrew’s academic supervisor, for supporting his participation in this research.

## References

- [1] Francesco Boselli, Emily Steed, Jonathan B Freund, and Julien Vermot. Anisotropic shear stress patterns predict the orientation of convergent tissue movements in the embryonic heart. *Development*, 144(23):4322–4327, 2017.
- [2] Otger Campas. A toolbox to explore the mechanics of living embryonic tissues. In *Seminars in cell & developmental biology*, volume 55, pages 119–130. Elsevier, 2016.
- [3] Maria R D’Orsogna, Yao-Li Chuang, Andrea L Bertozzi, and Lincoln S Chayes. Self-propelled particles with soft-core interactions: patterns, stability, and collapse. *Physical review letters*, 96(10):104302, 2006.
- [4] Hisao Honda. Left-handed cardiac looping by cell chirality is mediated by position-specific convergent extensions. *Biophysical Journal*, 120(23):5371–5383, 2021.
- [5] Franz P Hutterer, Benedikt Buchmann, Lisa K Engelbrecht, and Andreas R Bausch. Collective cell migration during human mammary gland organoid morphogenesis. *Biophysics Reviews*, 3(4), 2022.
- [6] Mizuho Kanamori, Kenta Oikawa, Kentaro Tanemura, and Kenshiro Hara. Mammalian germ cell migration during development, growth, and homeostasis. *Reproductive Medicine and Biology*, 18(3):247–255, 2019.
- [7] Hinako Kidokoro, Yukio Saijoh, and Gary C Schoenwolf. Nodal signaling regulates asymmetric cellular behaviors, driving clockwise rotation of the heart tube in zebrafish. *Communications Biology*, 5(1):996, 2022.
- [8] Hiroshi Koyama, Hisashi Okumura, Tetsuhisa Otani, Atsushi M Ito, Kazuyuki Nakamura, Kagayaki Kato, and Toshihiko Fujimori. Effective mechanical potential of cell–cell interaction in tissues harboring cavity and in cell sheet toward morphogenesis. *Frontiers in Cell and Developmental Biology*, 12:1414601, 2024.
- [9] Ju-Sung Lee, Tatiana Filatova, Arika Ligmann-Zielinska, Behrooz Hassani-Mahmooei, Forrest Stonedahl, Iris Lorscheid, Alexey Voinov, J Gareth Polhill, Zhanli Sun, and Dawn C Parker. The complexities of agent-based modeling output analysis. *Journal of Artificial Societies and Social Simulation*, 18(4), 2015.
- [10] Kaoru Miyazaki, Jun Oyanagi, Daisuke Hoshino, Shinsaku Togo, Hiromichi Kumagai, and Yohei Miyagi. Cancer cell migration on elongate protrusions of fibroblasts in collagen matrix. *Scientific reports*, 9(1):292, 2019.
- [11] Corrado Monti, Marco Pangallo, Gianmarco De Francisci Morales, and Francesco Bonchi. On learning agent-based models from data. *Scientific Reports*, 13(1):9268, 2023.
- [12] Kyle C Nguyen, Carter D Jameson, Scott A Baldwin, John T Nardini, Ralph C Smith, Jason M Haugh, and Kevin B Flores. Quantifying collective motion patterns in mesenchymal cell populations using topological data analysis and agent-based modeling. *Mathematical biosciences*, 370:109158, 2024.
- [13] Jonas Pleyer and Christian Fleck. Agent-based models in cellular systems. *Frontiers in Physics*, 10:968409, 2023.
- [14] Yeh-Chuin Poh, Junwei Chen, Ying Hong, Haiying Yi, Shuang Zhang, Junjian Chen, Douglas C Wu, Lili Wang, Qiong Jia, Rishi Singh, et al. Generation of organized germ layers from a single mouse embryonic stem cell. *Nature communications*, 5(1):4000, 2014.
- [15] Mona Pourjafar and Vijay K Tiwari. Plasticity in cell migration modes across development, physiology, and disease. *Frontiers in Cell and Developmental Biology*, 12:1363361, 2024.

- [16] Timothy Recaladin, Linda Steinacher, Bruno Gjeta, Marius F Harter, Lukas Adam, Kristina Kromer, Marisa Pimentel Mendes, Marina Bellavista, Mikhail Nikolaev, Giacomo Lazzaroni, et al. Human organoids with an autologous tissue-resident immune compartment. *Nature*, 633(8028):165–173, 2024.
- [17] Nabid Salehin, Tanveer Teranikar, Juhyun Lee, and Cheng-Jen Chuong. Ventricular anisotropic deformation and contractile function of the developing heart of zebrafish in vivo. *Developmental Dynamics*, 252(2):247–262, 2023.
- [18] Corinna Singleman and Nathalia G Holtzman. Analysis of postembryonic heart development and maturation in the zebrafish, danio rerio. *Developmental Dynamics*, 241(12):1993–2004, 2012.
- [19] Richard Sinkhorn. A relationship between arbitrary positive matrices and doubly stochastic matrices. *The annals of mathematical statistics*, 35(2):876–879, 1964.
- [20] Michael Smutny, Martin Behrndt, Pedro Campinho, Verena Ruprecht, and Carl-Philipp Heisenberg. Uv laser ablation to measure cell and tissue-generated forces in the zebrafish embryo in vivo and ex vivo. In *Tissue morphogenesis: methods and protocols*, pages 219–235. Springer, 2014.
- [21] Simon Tanaka. Simulation frameworks for morphogenetic problems. *Computation*, 3(2):197–221, 2015.
- [22] Xavier Trepas, Zaozao Chen, and Ken Jacobson. Cell migration. *Comprehensive Physiology*, 2(4):2369, 2012.
- [23] Hideki Yamaguchi, Jeffrey Wyckoff, and John Condeelis. Cell migration in tumors. *Current opinion in cell biology*, 17(5):559–564, 2005.