

# Agent-based modeling of cell differentiation in mouse ICM organoids



Dissertation zur Erlangung des  
naturwissenschaftlichen Doktorgrades  
der Julius-Maximilians-Universität Würzburg

vorgelegt von  
**Simon Schardt**

geboren in  
**Germersheim**

Würzburg, 2022



This document is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (CC BY-NC-ND 4.0):  
<http://creativecommons.org/licenses/by-nc-nd/4.0> This CC license does not apply to third party material (attributed to another source) in this publication.

Eingereicht am 08.11.2022

**Mitglieder der Promotionskommission:**

Vorsitzender: Prof. Jörg Schultz

Gutachter: Prof. Sabine C. Fischer

Gutachter: Prof. Philip Kollmannsberger

Tag des Promotionskolloquiums: 13.01.2023

Doktorurkunde ausgehändigt:

## Eidesstattliche Erklärungen nach §7 Abs. 2 Satz 3, 4, 5 der Promotionsordnung der Fakultät für Biologie

### Eidesstattliche Erklärung

Hiermit erkläre ich an Eides statt, die Dissertation: **Agentenbasierte Modellierung der Zelldifferenzierung in Maus ICM Organoiden**", eigenständig, d. h. insbesondere selbständig und ohne Hilfe eines kommerziellen Promotionsberaters, angefertigt und keine anderen, als die von mir angegebenen Quellen und Hilfsmittel verwendet zu haben.

Ich erkläre außerdem, dass die Dissertation weder in gleicher noch in ähnlicher Form bereits in einem anderen Prüfungsverfahren vorgelegen hat.

Weiterhin erkläre ich, dass bei allen Abbildungen und Texten bei denen die Verwertungsrechte (Copyright) nicht bei mir liegen, diese von den Rechtsinhabern eingeholt wurden und die Textstellen bzw. Abbildungen entsprechend den rechtlichen Vorgaben gekennzeichnet sind sowie bei Abbildungen, die dem Internet entnommen wurden, der entsprechende Hypertextlink angegeben wurde.

### Affidavit

I hereby declare that my thesis entitled: "**Agent-based modeling of cell differentiation in mouse ICM organoids**" is the result of my own work. I did not receive any help or support from commercial consultants. All sources and / or materials applied are listed and specified in the thesis.

Furthermore, I verify that the thesis has not been submitted as part of another examination process neither in identical nor in similar form.

Besides, I declare that if I do not hold the copyright for figures and paragraphs, I obtained it from the rights holder and that paragraphs and figures have been marked according to law or for figures taken from the internet the hyperlink has been added accordingly.

Würzburg, den - - - - -

- - - - -  
Signature PhD-student

## Zusammenfassung

Die embryonale Entwicklung von Säugetieren unterliegt komplexen biologischen Zusammenhängen, die es zu verstehen gilt. Bevor jedoch das gesamte Gebilde der Entwicklung zusammengesetzt werden kann, müssen zunächst die einzelnen Bausteine genauer verstanden werden. Einer dieser Bausteine ist die zweite Zellschicksalsentscheidung und beschreibt die Differenzierung von Zellen der inneren Zellmasse des Embryos hin zu Epiblast- und primitiven Endodermzellen. Diese Zellen teilen sich daraufhin räumlich auf und bilden die anschließend die Grundlagen für den Embryo und den Dottersack. In Organoiden der inneren Zellmasse wird ebenfalls beobachtet, wie sich diese zwei Typen von Vorläuferzellen bilden, und sich in gewissem Maße räumlich voneinander trennen. Diesem Phänomenen widmete sich diese Arbeit im Verlaufe der letzten drei Jahre. Über diese Zelldifferenzierung ist bereits bekannt, dass die ersten Anzeichen für Epiblast- und primitive Endodermdifferenzierung jeweils die Expressionslevel der Transkriptionsfaktoren NANOG und GATA6 sind. Dabei nehmen Zellen mit niedriger Expression an GATA6 und hoher Expression an NANOG das Epiblastschicksal an. Sind die Expressionen umgekehrt, so entsteht eine primitive Endodermzelle. Bei der räumlichen Aufteilung der beiden Zelltypen ist noch nicht eindeutig geklärt, welcher Mechanismus dazu führt. Eine gängige Hypothese besagt, dass die Ursache für die räumliche Umlagerung der Zellen in der unterschiedlichen Adhäsion der Zellen liegt. In dieser Arbeit wird jedoch die Möglichkeit einer globalen Zell-Zell-Kommunikation untersucht. Die gewählte Vorgehensweise bei der Untersuchung dieser Phänomene folgt dem Motto "Die Mathematik ist das nächste Mikroskop der Biologie" [1]. Mit Hilfe mathematischer Modellierung wird das zentrale genregulierende Netzwerk im Mittelpunkt dieser Arbeit in ein Gleichungssystem umgewandelt, welches es ermöglicht, die zeitliche Entwicklung von NANOG und GATA6 unter Einfluss eines externen Signals zu beschreiben. Ein besonderes Augenmerk liegt dabei auf der Herleitung neuer Modelle mit Hilfe von Methoden der statistischen Mechanik, sowie dem Vergleich mit bestehenden Modellen. Nach einer ausführlichen Stabilitätsanalyse werden die Vorteile des hergeleiteten Modells dadurch deutlich, dass ein exakter Zusammenhang der Modellparameter und der Formierung von heterogenen Mischungen zweier Zelltypen gefunden wurde. Dadurch lässt sich das Modell einfach kontrollieren und die Proportionen der resultierenden Zelltypen bereits im Voraus abschätzen. Dieses mathematische Modell wird außerdem kombiniert mit einem Mechanismus zur globalen Zell-Zell Kommunikation, sowie einem Modell zum Wachstum eines Organoiden. Dabei wird gezeigt dass die globale Zell-Zell Kommunikation dazu in der Lage ist die Bildung von Schachbrettmustern, sowie auch umrandenden Muster anhand unterschiedlich ausbreitender Signale zu vereinen. Zusätzlich wird der Einfluss der Zellteilung und somit des Organoidwachstums auf die Musterbildung genauestens untersucht. Es wird gezeigt, dass dies zur Bildung von Clustern beiträgt und infolgedessen eine gewisse Zufälligkeit in ansonsten perfekt sortierte Muster einbringt.

## Summary

Mammalian embryonic development is subject to complex biological relationships that need to be understood. However, before the whole structure of development can be put together, the individual building blocks must first be understood in more detail. One of these building blocks is the second cell fate decision and describes the differentiation of cells of the inner cell mass of the embryo into epiblast and primitive endoderm cells. These cells then spatially segregate and form the subsequent bases for the embryo and yolk sac, respectively. In organoids of the inner cell mass, these two types of progenitor cells are also observed to form, and to some extent to spatially separate. This work has been devoted to these phenomena over the past three years. Plenty of studies already provide some insights into the basic mechanics of this cell differentiation, such that the first signs of epiblast and primitive endoderm differentiation, are the expression levels of transcription factors NANOG and GATA6. Here, cells with low expression of GATA6 and high expression of NANOG adopt the epiblast fate. If the expressions are reversed, a primitive endoderm cell is formed. Regarding the spatial segregation of the two cell types, it is not yet clear what mechanism leads to this. A common hypothesis suggests the differential adhesion of cell as the cause for the spatial rearrangement of cells. In this thesis however, the possibility of a global cell-cell communication is investigated. The approach chosen to study these phenomena follows the motto "mathematics is biology's next microscope" [1]. Mathematical modeling is used to transform the central gene regulatory network at the heart of this work into a system of equations that allows us to describe the temporal evolution of NANOG and GATA6 under the influence of an external signal. Special attention is paid to the derivation of new models using methods of statistical mechanics, as well as the comparison with existing models. After a detailed stability analysis the advantages of the derived model become clear by the fact that an exact relationship of the model parameters and the formation of heterogeneous mixtures of two cell types was found. Thus, the model can be easily controlled and the proportions of the resulting cell types can be estimated in advance. This mathematical model is also combined with a mechanism for global cell-cell communication, as well as a model for the growth of an organoid. It is shown that the global cell-cell communication is able to unify the formation of checkerboard patterns as well as engulfing patterns based on differently propagating signals. In addition, the influence of cell division and thus organoid growth on pattern formation is studied in detail. It is shown that this is able to contribute to the formation of clusters and, as a consequence, to breathe some randomness into otherwise perfectly sorted patterns.

# Contents

<b>1</b>	<b>Introduction</b>	<b>8</b>
1.1	Mouse embryo development and ICM organoids . . . . .	8
1.2	Mathematical modeling of cell differentiation in mouse ESCs . . . . .	9
1.3	Research questions and proposed solutions . . . . .	9
<b>2</b>	<b>Methods</b>	<b>12</b>
2.1	Cell graph . . . . .	12
2.2	Pair correlation function . . . . .	12
2.3	Moran's $I$ . . . . .	15
2.4	Data pre-processing . . . . .	17
2.5	Computational methods . . . . .	19
<b>3</b>	<b>Models and analyses</b>	<b>22</b>
3.1	Transcriptional regulation . . . . .	22
3.1.1	Generalized phenomenological models . . . . .	22
3.1.2	Phenomenological models . . . . .	29
3.1.3	Thermodynamic models . . . . .	29
3.1.4	Auto-activation dominant model . . . . .	33
3.1.5	Transcriptional regulation summary . . . . .	39
3.2	Cell-cell communication . . . . .	40
3.2.1	Averaged nearest neighbor signaling . . . . .	41
3.2.2	Distance-based global signaling . . . . .	41
3.2.3	Estimation of the stability interval of model ( $M3$ ) . . . . .	42
3.3	Organoid growth . . . . .	43
3.3.1	Cell growth . . . . .	43
3.3.2	Radius dependent cell division . . . . .	44
3.3.3	Cell motion . . . . .	45
3.3.4	Implementation . . . . .	50
3.3.5	Summary . . . . .	51
<b>4</b>	<b>Computational Results</b>	<b>53</b>
4.1	Analysis of experimental data . . . . .	53
4.1.1	24 h-organoids show no clear distribution of cell types . . . . .	53
4.1.2	48 h-organoids exhibit radially distributed cell types . . . . .	53
4.1.3	Experimental data reveals cluster formation . . . . .	53
4.1.4	Conclusion . . . . .	56
4.2	Transcriptional regulation model comparison . . . . .	56
4.2.1	Comparison setup . . . . .	56
4.2.2	Comparison of expression values . . . . .	58
4.2.3	Conclusion . . . . .	58
4.3	Nearest neighbor signaling leads to checkerboard patterns . . . . .	60
4.3.1	Color coding . . . . .	60
4.3.2	Checkerboard pattern . . . . .	60
4.3.3	Cell type proportions . . . . .	61
4.3.4	Conclusion . . . . .	63
4.4	Distance-based global signaling enables a transition from checkerboard to engulfing patterns . . . . .	64
4.4.1	From checkerboard to engulfing patterns . . . . .	64
4.4.2	Cell type proportions . . . . .	67
4.4.3	Patterns match experimental data at long distances . . . . .	67
4.4.4	Conclusion . . . . .	69
4.5	Cell division leads to clustering of equal cell types . . . . .	71
4.5.1	Differentiation time has little influence on the resulting pattern . . . . .	71
4.5.2	Dynamic vs. static – Cell division leads to increased cluster formation . . . . .	74
4.5.3	Number of cell divisions increase cluster sizes . . . . .	74
4.5.4	Conclusion . . . . .	77
4.6	Matching simulations with experimental data . . . . .	79

<b>5</b>	<b>Discussion and Outlook</b>	<b>81</b>
5.1	Analytical results for the stability of heterogeneous steady states could be extended . . .	81
5.2	Comparison of transcriptional regulation models results in one model overshadowing the rest. . . . .	81
5.3	The effect of cell-cell communication on the resulting differentiation patterns has been largely identified . . . . .	82
5.4	Cell division provides the necessary clustering to describe the patterns in ICM organoids .	82
5.5	Outlook - From ICM organoid towards the embryo . . . . .	83

# 1 Introduction

If an ovum is released and successfully fertilized by a sperm cell, a single cell called zygote is formed. As the zygote starts to divide and grow, it passes different stages of development before it implants into the uterus. It is during these stages, that a large amount of pregnancies already fail due to loss of the fertilized ovum. The exact amount of lost ova is still up to debate and has attracted great interest of researchers over the last decades. Recent studies have estimated around 10-40% [2] lost ova before implantation, which was later narrowed down to 40-50% [3]. This motivates the question of why exactly so many preimplantations fail. For this and probably various other reasons, there exists an astounding research interest in precisely understanding the development of mammalian embryos. In particular, the mouse embryo has been established as a popular model organism in developmental biology. This thesis as well is based on data and experimental observations derived from mouse embryos and mouse embryonic stem cells (ESCs).

## 1.1 Mouse embryo development and ICM organoids

The preimplantation phase of the mouse embryo is shaped by two successive cell fate decisions. During the first decision, cells become either part of the trophectoderm (TE) or the inner cell mass (ICM). The TE consists of a single layer of cells that entirely surrounds the ICM as well as the blastocoel cavity. The ICM will eventually give rise to the fetus. It is in the ICM where the second cell fate decision takes place. ICM cells differentiate into either epiblast (Epi) or primitive endoderm (PrE) cells. In the course of further development, Epi cells give rise to the embryo proper, whereas PrE cells contribute to the yolk sac [4, 5, 6, 7]. In this thesis, only the second cell fate decision is examined in more detail. The transcription factor (TF) NANOG has been identified as the earliest marker for the differentiation from ICM into Epi cells [8]. Likewise, TF GATA6 has been reported to influence the decisions towards PrE cells [9].

In the early blastocyst, i.e. embryonic day 3 (E3.0) after fertilization, both TFs NANOG and GATA6 exhibit high expressions. Subsequently, through a complex network of transcriptional regulations, expressions are regulated to become mutually exclusive (E3.0-E4.5). Therefore, cells with high expressions of NANOG and low expressions of GATA6 are the Epi precursor cells. Conversely, low expressions of NANOG and high expressions of GATA6 lead to PrE precursor cells. Understanding the gene regulatory network (GRN) at the core of transcriptional regulation is vital for any attempt to develop models, methods and hypotheses at this stage of embryonic development. It was demonstrated that overexpression of NANOG leads to a reduction of GATA6. Subsequently, NANOG was identified as an inhibitor of GATA6 [10]. Likewise, induction of GATA6 was shown to down-regulate transcription of several pluripotency associated genes including NANOG [11]. Combined, NANOG and GATA6 mutually inhibit each other. Additionally, both TFs were shown to be positively regulating their own expressions [12, 11]. These interactions account for the mutual exclusivity found in Epi and PrE precursor cells. Yet they do not explain the emergence of a heterogeneous mix of both cell types. It is suggested that an extracellular mechanism like fibroblast growth factor (FGF) signaling is acting as the decision maker in differentiation. Indeed, experiments of inhibiting the FGF/extracellular-signal regulated kinase (ERK) pathway have shown to promote NANOG expression [13]. At the same time, treatment with exogenous FGF4 resulted in increased GATA6 expression [14]. The binding of FGF4 to its receptor FGFR2 was identified leading to the activation of GATA6 and inhibition of NANOG [15, 16]. Furthermore, experiments have demonstrated, that FGF4 synthesis is increased in Epi precursor cells compared to NANOG knockout cells [17]. This suggests an activation of FGF4 by NANOG. Overall, this means that cells with high NANOG expression synthesize more FGF4 that acts on adjacent cells to down-regulate NANOG, in turn leading to higher expressions of GATA6. Thus, heterogeneity of Epi and PrE precursors is made possible.

At the late blastocyst stage, Epi and PrE cells are found spatially segregated with PrE cells forming a monolayer that separates the Epi cells from the blastocoel cavity (Fig. 1). These findings were statistically described as transitions from local to global patterns between embryonic stages of 65-90 to 90+ cells [18]. The exact mechanism that leads to the sorting of the cells is still up to debate. Cell sorting due to differential adhesion could provide one possible explanation [19]. So far, this has not been proven. A study based on the expression levels of E-cadherin, a well known cell adhesion molecule, has found no significant evidence to support this hypothesis, because the difference in E-cadherin occurs only after the segregation has concluded [20]. An alternative to this was presented via Eph/Ephrin ligand receptor pairings that have been proposed to account for differential adhesion in a recent study [21]. At the same time, it seems reasonable to investigate different mechanisms that could induce spatially segregated patterns of two different cell types.



Besides the mouse embryo, in vitro model systems have already been investigated. Mouse ESCs from the Epi portion of the ICM have been engineered to replicate the differentiation process during preimplantation development in vitro [22]. These ESCs express high levels of both NANOG and GATA6 and regained their ability to differentiate into PrE- and Epi-like cells. Based on these cells, ICM organoids were grown in vitro to provide a different perspective on the full picture of Epi and PrE segregation [23]. Due to their large cell number and the symmetry provided by their spherical geometry, they are well suited for statistical analyses. Organoids that aged 24 h and 48 h have already been shown to resemble the total cell type proportions as well as the neighborhood proportions of the mouse blastocyst to some extent. Furthermore, microscopy images of 48 h-organoids show signs of spatial segregation by PrE-like cells engulfing Epi-like cells illustrated in figure 1. Further statistical analyses on these data are still an ongoing effort and will partly be discussed throughout this thesis.

## 1.2 Mathematical modeling of cell differentiation in mouse ESCs

Mathematically, cell differentiation is described by the temporal evolution of transcription factors via a system of ordinary differential equations (ODEs) that is intended to represent the various interactions in the GRN. These models have been investigated on different levels. At the single cell level, models of two mutually inhibiting genes were described in great detail [24, 25]. The resulting ODE systems are characterized by two different steady state solutions which are associated with a corresponding cell fate. These states are characterized by high expressions of one protein and low expressions of the other, and vice-versa. Auto-activation has shown the capability to include a third steady state into the system, that resembles a state of simultaneous high expression levels. However, the resulting cell fate decisions are solely dependent on the initial conditions of the system and do not incorporate any extracellular influences. At the multi-cellular level, agent-based models are used to describe every cell as a single agent. This way, interactions between the cells can be realized. Adjacency-based interaction models, i.e. models in which cells are capable to communicate with adjacent cells, have been thoroughly investigated in the current research landscape. One prominent example, which was also a personal starting point for this complete journey, is the Delta-Notch signaling pathway in *Drosophila* [26]. There, the reconstruction of patterns of alternating cell types has been successfully realized. For the mouse embryo as well, adjacency-based interaction models were used to recreate the cell-cell communication via FGF signaling [27, 28]. The averaged nearest neighbor (NN) signals used throughout these studies lead to a very characteristic pattern that is best described as an irregular checkerboard pattern, i.e. a pattern in which cells of one type try to avoid being adjacent to cells of equal type. The successive combination of cell division, cell sorting and apoptosis, it has demonstrated that this leads to the desired spatial segregation of cells [29, 19, 30]. However, the successive nature of these simulations do not accurately present a realistic scenario, where all of these processes run in parallel. Another potential mechanism that might explain the spatial segregation of cell types is global signaling. This idea has already been explored to some extent [31]. Furthermore, in recent findings on mouse ESCs, the ability of FGF4 to migrate over distances beyond a cell’s nearest neighbors has been demonstrated [32].

## 1.3 Research questions and proposed solutions

This thesis focuses on the development and analysis of a computational model for the in silico generation of mouse ICM organoids. The central hypothesis motivating this research is that global cell-cell communication is the cause of the spatial segregation of Epi and PrE cells in ICM organoids. In order to confirm or refute this hypothesis, simulated organoids have to reflect the experimental data as accurately as possible. To this end, several questions must first be answered.

The first question that arises is how to quantify the experimental data to capture both local and global characteristics of a pattern. In [23], a cell neighborhood analysis was used to reveal a local clustering of PrE and Epi precursor cells. To express this by a single number, Moran’s  $I$  is used here as a measure of spatial auto-correlation [33]. This also leads to the finding of local clustering in both 24 h- and 48 h-organoids. In order to quantify the global characteristics of an engulfing pattern, pair correlation function (PCFs) are used [34, 35]. As a result, it turns out that the 24 h-organoids are largely unclassifiable. Contrary to this, the radial distribution of cells with low NANOG and high GATA6 expression (N-G+) in 48 h-organoids is verified. Thus, the goal of the computational model is to generate organoids that match the characteristics of the experimental data.

The next question in line concerns the modeling of the cell fate decisions. Therefore, the mathematical basis for transcriptional regulation is established. Contrary to existing transcription models, an

investigation of different modeling approaches is taken into consideration. In a fully generalized model with two cells, it is possible to analytically derive conditions that guarantee heterogeneity similar to [26]. For more specific models, the current state of the art is given by phenomenological models that describe any inhibition and activation of TFs by an individual Hill equation [36]. This is compared to a thermodynamic model that is derived from the methods introduced in [37, 38, 39]. A third and final model is given in the form of a special case of the thermodynamic model. By assuming the auto-activation in the GRN to be so dominant that transcription becomes synonymous with a TF binding in the proximity of its own promoter, a simplified model arises. For this model a thorough stability analysis reveals key parameters that ensure heterogeneity of the simulated organoids. This feature alone assures that this model is far superior to the others in terms of application.

The growth of an ICM organoid is a crucial part of its development. Therefore, the question arises, how the growth of an ICM organoid can be modeled. To this end, an organoid growth model that is partly inspired by [40, 41] is applied. In this model, cells are described by the position of their centroid in two and three dimensions, as well as their size given by their radii. The radius is used to decide if a cell will divide or not based on a fixed probability distribution. This leads to an estimation of the cell growth rates and organoid growth rates that correspond to the experimental data of the ICM organoids. The Morse potential [42] is used in an overdamped system of equations of motion to describe the adhesion and repulsion of the cells. In combination this leads to the formation of circular (2D) and spherical (3D) tissues that closely resemble the ICM organoids.

Cell fate decisions are controlled by the transcriptional regulation. Yet the source of heterogeneity, i.e. a mixture of two different cell types, still has to be introduced. This is handled by the cell-cell communication, which leads to the question how it influences the formation of different cell type patterns. And more specifically, if it is capable of generating engulfing patterns and thus mimic the patterns found in the ICM organoids. To this end, two different types of signal are investigated. A nearest neighbor signal only capable of reaching adjacent cells and a distance based signal that disperses throughout the whole organoid with decreasing intensity based on the traveled distance. For the distance-based signaling, the signal dispersion is described by a single parameter. Even though engulfing patterns can be reproduced with the distance based signal, the hypothesis that global signaling is responsible for the pattern formation in ICM organoids was simultaneously disproven. Still it is demonstrated that the pattern formation could be influenced by the way that the intercellular signal disperses throughout the organoid. Thus, the formation of checkerboard as well as engulfing pattern is unified under the notion of differently dispersing signals. Furthermore, the simulation results are used to classify the experimental data. For this purpose, simulations with static and dynamic geometries were performed and compared, highlighting the extent of cell division on pattern formation. Cell division is already believed to promote the formation of clusters of equal cell types [43]. However, existing models focus either on cell division or transcriptional regulation rather than both simultaneously. In this thesis, these components are combined into a multiscale model. In the end, the ICM organoids are successfully classified through organoids that are generated by the computational model under the influence of cell division and a global cell-cell communication with low to medium signal dispersions.

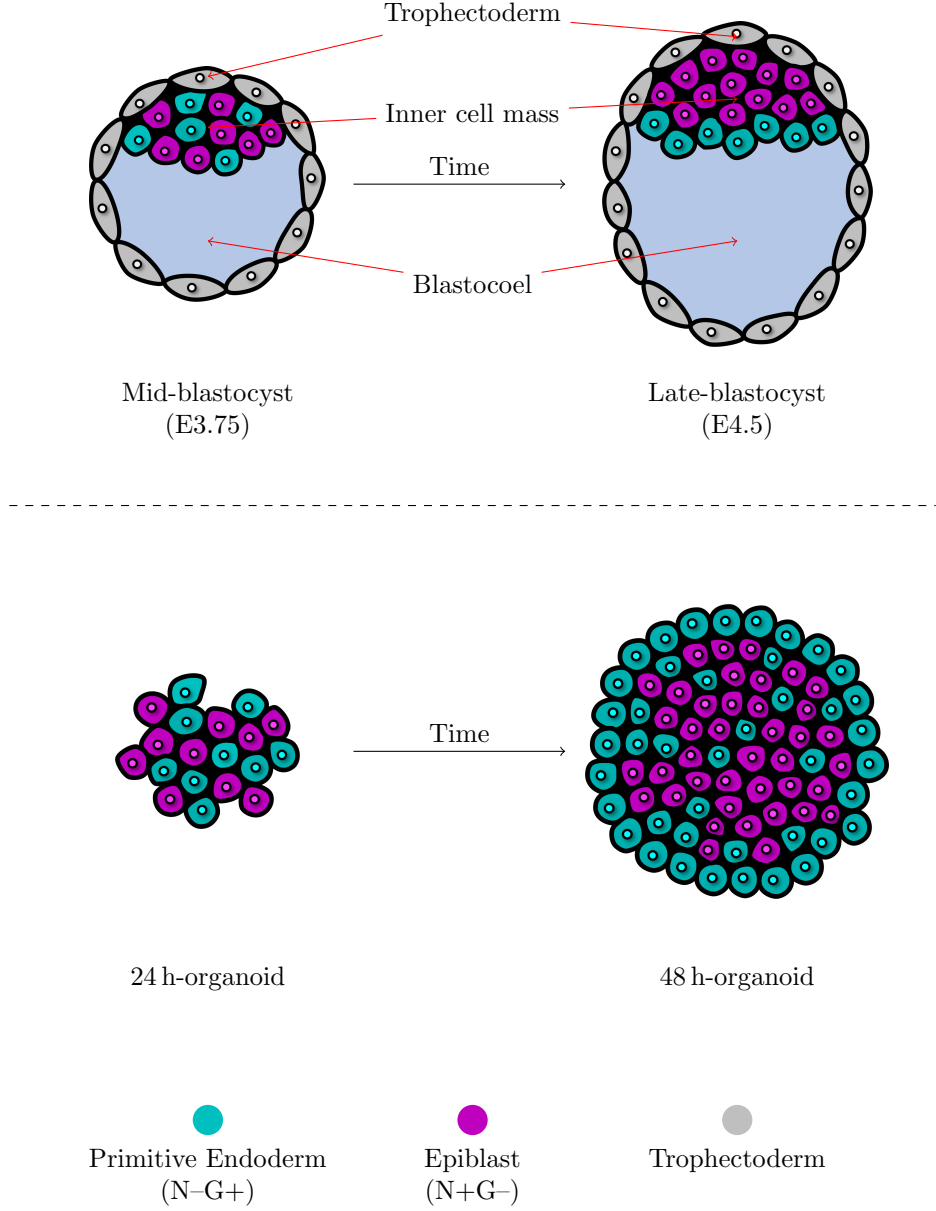


Figure 1: Illustrative comparison of the mouse embryo during embryonic day 3.75 and 4.5 with their corresponding ICM organoid counterparts at 24 h and 48 h. PrE precursor cells, i.e. cells with low NANOG expression are colored in cyan. Conversely, Epi precursor cells are colored in magenta. In the embryo, the ICM together with the blastocoel cavity (blue) is surrounded by the trophectoderm (grey).

## 2 Methods

In this section, it will be shown how an organoid, in vitro or in silico, can be represented by a graph. This graph forms the basis for the cell-cell interactions in the simulations but also for the chosen quantification methods, PCF and Moran's  $I$ . Furthermore, it provides the means to compare simulation results with experimental data. For this comparison, a pre-processing pipeline will be established. Finally, an overview of basic computational methods will be given.

### 2.1 Cell graph

Mathematically, two properties for each single cell are enough to describe the spatial composition of an organoid. These are the position of the cells as well as the information which cells are adjacent to each other, i.e. the cell neighborhoods. Over time, the position of any cell relative to other cells changes, resulting in individual compositions of cell positions and neighborhoods. The goal of this section is to provide a suitable means of representing cell positions and neighborhoods through the use of the cell graph  $G$ . A graph  $G = (V, E)$  consists of ordered sets of vertices  $V$  and edges  $E$ .

**Vertices:** Let  $V = \{v_1, \dots, v_n\}$ , then the vertex  $v_i$  is a representation for the cell with index  $i$ . Each vertex  $v_i$  is assigned a two- or three-dimensional coordinate  $\mathbf{x}_i$ . For the experimental data, the centroids of the cell nuclei have been calculated in [23] based on the segmentation described in [44]. The cell centroid is a fitting measure to describe the position of a cell in space, therefore the terms "centroid" and "position" will be used interchangeably throughout this work. Thus, the computational model that will be introduced in this thesis, as well as its representation as a graph are both based on the cell centroids.

**Edges:** By definition, vertices  $v_i$  and  $v_j$  are connected via an edge  $e_{ij}$  in  $G$ , if and only if  $e_{ij} \in E$ . In addition to the position, cells also have a spatial extension from the cell centroid to the membrane. Based on this, an edge is to be generated in each case, provided that the membranes of two cells are in direct contact. The way this is approached is by propagating spheres around the cell until they meet the spheres of other cells. Thus, the Voronoi tessellation is obtained. The spheres are used to mathematically describe the extension from cell centroid to membrane. Therefore, the Voronoi tessellation approximates the volume of the cells. The dual graph of the Voronoi tessellation is the Delaunay triangulation. Here, for each two tiles that are separated by a common edge, a separate edge connects both centroids, fulfilling our initial goal of creating a graph representing the tissue cell by cell. However, the tiles created by the cells that are part of the convex hull of  $\{\mathbf{x}_1, \dots, \mathbf{x}_n\}$  extend infinitely. Thus, the Delaunay triangulation yields connections of cells that realistically cannot be in contact due to their finite cell size. To prevent this, a cutoff value is introduced, such that edge  $e_{ij}$  is removed from  $G$ , if

$$|\mathbf{x}_i - \mathbf{x}_j| > \text{cutoff}. \quad (1)$$

Here,  $|\cdot|$  denotes the euclidean norm of an  $n$ -dimensional vector. This approach has already proven to create suitable cell graph representations for breast carcinoma spheroids [44] as well as mouse embryos and mouse ICM organoids [23]. Ideally, the cutoff is given as the sum of the radii  $r_i$  and  $r_j$  of the two cells, i.e.

$$|\mathbf{x}_i - \mathbf{x}_j| > r_i + r_j. \quad (2)$$

The radius of the cells is part of the simulations, therefore this approach is feasible. However, for the experimental data there is no information on the radius of the cells. The cutoff was chosen as the mean distance to neighboring cells of the corresponding Delaunay graph of a single organoid plus two times its standard deviation. In a normal distribution, this would approximately amount to the 2.28% longest edges. The distributions are slightly skewed to the right, which means that the percentage of removed edges is slightly larger.

**Distances:** For any two cells  $i$  and  $j$ , their euclidean distance towards each other has already been introduced in (1). Another relevant type of distance is the graph distance  $d_{ij}$ . This will be defined as the length of the shortest path connecting the cells  $i$  and  $j$ . For an unweighted and undirected graph like this, the most efficient way to calculate all of these distances is given by the breadth-first search algorithm (BFS) [45].

### 2.2 Pair correlation function

In order to compare the cell differentiation patterns of in vitro and in silico organoids not only qualitatively but also quantitatively, a measure of the spatial distribution of the different cell types is needed. To this

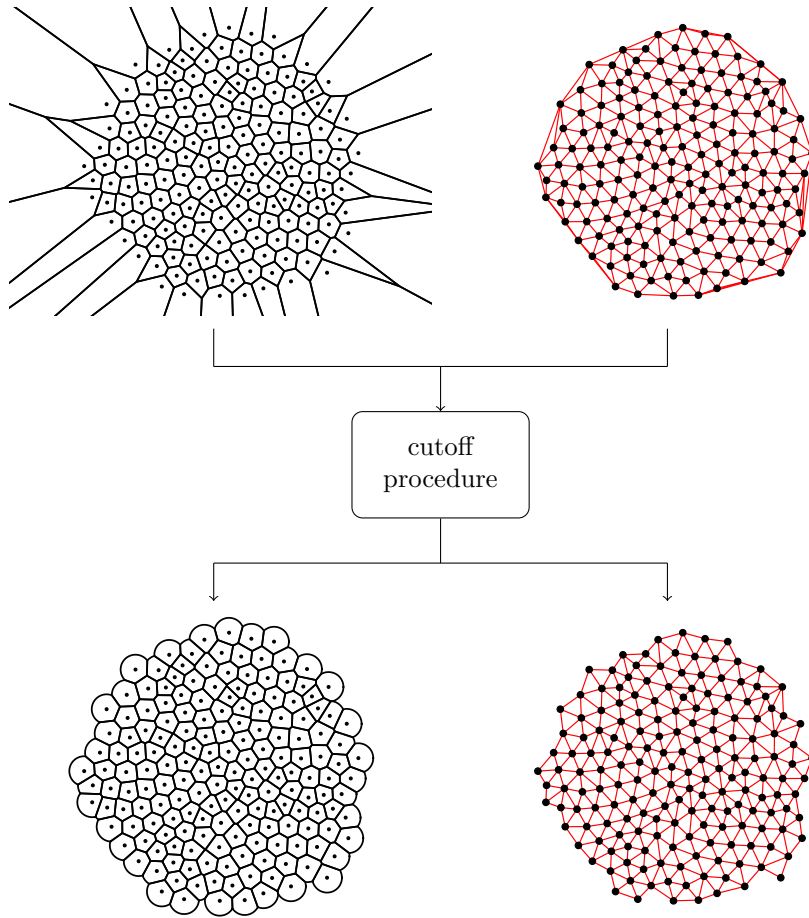


Figure 2: Voronoi tessellation (left) and Delaunay triangulation (right) before and after performing the cutoff procedure for an exemplary 2D organoid. Black dots represent the centroids of each cell. The images on the left depict the size of any individual cell via black edges denoting the bordering cell membranes. On the right, red lines connect vertices of cells in contact with each other. The image in the lower right represents the cell graph in this example.

end, pair correlation functions (PCFs) will be utilized which are typically used as a tool to characterize heterogeneities in the density of fluids [46]. As such they can be used to identify distances of regions of higher particle densities by basically counting and normalizing all particle pairs found at a specific distance towards each other. For the characterization of patterns of two different cell types, a specific discrete version of the PCF is needed. In [34], a PCF based on a pixelated grid has already been used to quantify patterns in a biological context. In another study, these PCFs were generalized to different types of regular and more importantly to irregular grids [35]. This was realized by using a graph as the foundation of defining the connectivity and thus a discrete distance between two particles. The particles in the context of this study will be cells of two different types. For now, these will be termed  $u^+v^-$  or  $u^-v^+$  cells for cells that exhibit high expression levels of a protein  $u$  and low expression levels of  $v$  and vice-versa. Mathematically, for  $n$  cells, the PCF is defined using the set of all cell pairs at distance  $k$  ( $S_k$ ), the sets of all cell pairs of equal cell type at distance  $k$  ( $S_k^u$  and  $S_k^v$ ) and the sets of all cells of a certain cell type ( $T^u$  and  $T^v$ ):

$$S_k = \{(i, j) \in \mathbb{N}^2 : d_{ij} = k, 1 \leq i, j \leq n\}, \quad (3)$$

$$S_k^u = \{(i, j) \in S_k : i, j \text{ are } u^+v^- \text{ cells}\}, \quad (4)$$

$$S_k^v = \{(i, j) \in S_k : i, j \text{ are } u^-v^+ \text{ cells}\}, \quad (5)$$

$$T^u = \{i \in \mathbb{N} : i \text{ is } u^+v^- \text{ cell } 1 \leq i \leq n\}, \quad (6)$$

$$T^v = \{i \in \mathbb{N} : i \text{ is } u^-v^+ \text{ cell } 1 \leq i \leq n\}. \quad (7)$$

The next step towards the PCF is the formation of the ratios

$$r_{uu} = \frac{|S_k^u|}{|S_k|}, \quad r_{vv} = \frac{|S_k^v|}{|S_k|}. \quad (8)$$

These provide the information about the spatial distribution of the cells. However, these need to be normalized by the probabilities of randomly picking two equal cell types

$$p_{uu} = \frac{|T^u|(|T^u| - 1)}{n(n - 1)}, \quad p_{vv} = \frac{|T^v|(|T^v| - 1)}{n(n - 1)}. \quad (9)$$

By combining (8) with (9), the PCFs are defined via

$$\rho_u(k) = \frac{r_{uu}}{p_{uu}} = \frac{|S_k^u|n(n - 1)}{|S_k||T^u|(|T^u| - 1)}, \quad (10)$$

$$\rho_v(k) = \frac{r_{vv}}{p_{vv}} = \frac{|S_k^v|n(n - 1)}{|S_k||T^v|(|T^v| - 1)}. \quad (11)$$

Now that the PCFs have been introduced, it remains to show how these have to be interpreted given different patterns. To this end, the cells in a simulated 2D organoid were assigned cell fates in different ways. The organoid was cut vertically  $n_x$  times at an equidistant spacing. The same was repeated with  $n_y$  vertical cuts. The mesh generated this way enables a systematic assignment of cell fates in different regions of the organoid. All the cells whose centroid are in one of the rectangular regions of the mesh were assigned to one fate, whereas the adjacent rectangles were filled with the opposite fate. This procedure was repeated for  $n_r$  radial cuts yielding radially separated cell regions. Examples for different patterns were created using  $n_x, n_y, n_r \in \{0, 1, 3, 5\}$  (Fig. 3). The first row of organoids were only separated in one dimension. When the two species are completely spatially separated in one dimension, i.e. two large adjacent cluster, the PCFs have large values in the low distance regions. This happens because the only pairs of different cells with low distance are placed directly at the dividing line, which simply represent fewer cells in relation to the rest. On the other hand, for larger distances you will find fewer pairs of identical cell types, as these are mostly the cell pairs close to the boundary but on the opposite side of the organoid. This leads to an overall decreasing trend in the PCF. Increasing the number of separations leads to the formation of additional peaks and valleys in the PCF. The alternating regions of specific cell fates lead to this phenomenon. Overall, the trend for  $\rho_u$  and  $\rho_v$  remains nearly identical in all three cases. The second row of organoids was divided in both dimensions. Again, the increase of spatial divisions lead to more peaks and valleys in the PCFs and the additional regularity of the patterns leads again to a mostly identical trend for  $\rho_u$  and  $\rho_v$ . Outliers for long distances are quite common in this way of pattern quantification. This is attributed to the low amount of cell pairs found at maximum distances in an organoid (Fig. 4). When observing a single cell, the cells with graph distance  $k$  towards it form a ring around the observed cell with a radius that increases with  $k$ . Thus the number of cell pairs  $|S_k|$  first increases with  $k$ . However, if  $k$  is larger than the distance towards the boundary, the cells with distance

$k$  are no longer able to form a complete ring, thus reducing the number of possible cell pairings. If  $k$  exceeds the maximum maximum of  $d_{ij}$ , then  $|S_k| = 0$ . Despite low numbers of cell pairs, the PCF can be reliable for long distances, provided the pairings are mostly occupied by one cell type. Patterns, where the long distance regions are fully occupied by a single cell type, consistently lead to large values in the PCF (Fig. 3 last row). In addition to that,  $\rho_u$  and  $\rho_v$  finally show a different trend. While  $\rho_v$  starts high and decreases with distance  $k$ ,  $\rho_u$  increases at some point again. The increase in  $n_r$  leads again to more peaks and valleys while preserving the separation at long distances. The change in the maximum values of the PCF comes from the change in cell type proportions. If one proportion increases, its normalization  $p_{uu}$  or  $p_{vv}$  decreases. Given the same pairs found at a certain distance, e.g. here at maximum distance, the PCF increases.

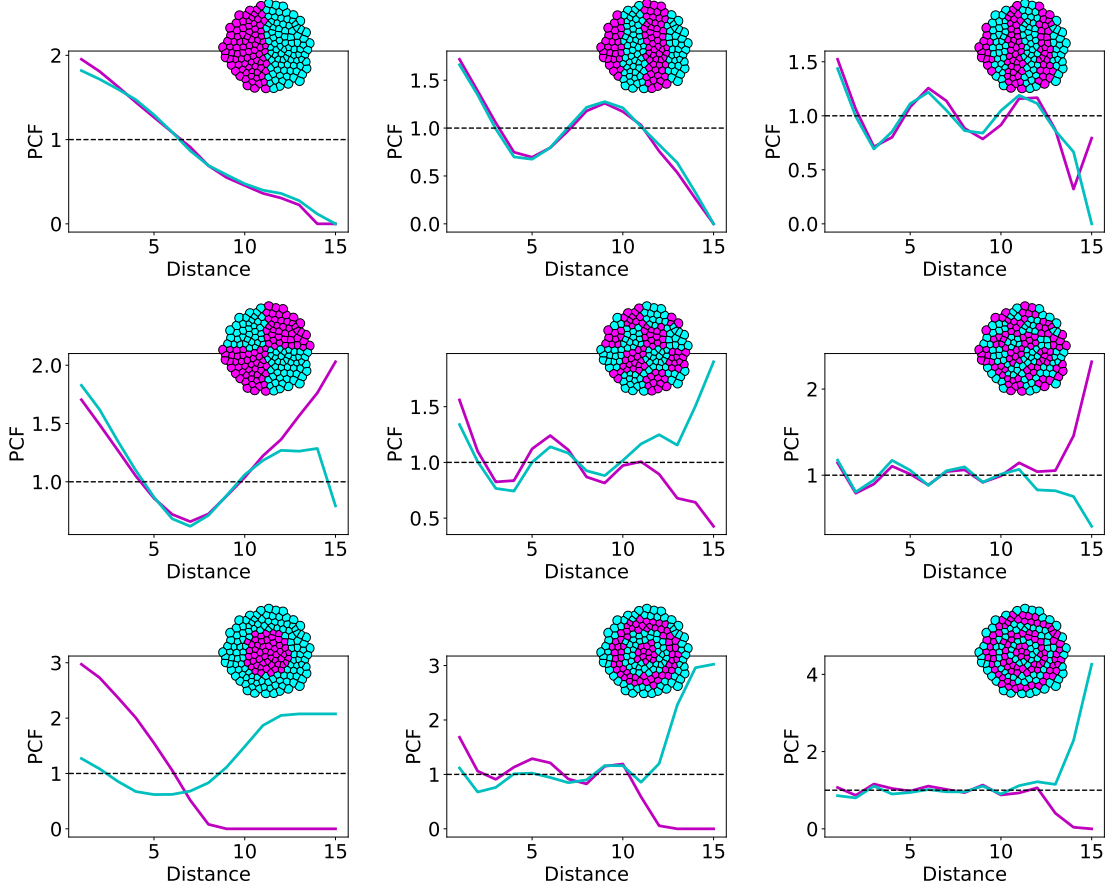


Figure 3: Exemplary PCFs for various patterns. In the top row, the organoid was divided only vertically, whereas in the center row it was also divided horizontally. The bottom row highlights a radial partitioning of the cells in the organoid. From left to right the number of partitionings  $n_x$ ,  $n_y$  and  $n_r$  increase from 1 over 3 to 5. The magenta curves correspond to the pairings of magenta cells. Analogously for cyan. The patterns that lead to the PCFs are included in the top right of every graph.

### 2.3 Moran's $I$

PCFs allow the quantification of the global cell differentiation pattern of an organoid. At the local level, i.e. direct cell neighborhoods, the PCF values for  $k = 1$  already yield some information about the neighborhood. Values above 1 are a sign that equal cells are clustering together, while values below 1 indicate the opposite behavior. However, individual PCF values like this are not normalized making it difficult to compare different organoids. Much like the PCFs, Moran's  $I$  is an alternative measure to quantify spatial auto-correlations. It was first introduced in [33] describing it as a test for the spatial correlation between neighbors. As such, the aim of this section is to establish Moran's  $I$  as a quantitative measure to describe the clustering nature of the cells in a tissue. For a variable of interest  $\mathbf{w} \in \mathbb{R}^n$ ,  $I$  is defined via

$$I := \frac{n}{\sum_{i,j} a_{ij}} \frac{(\mathbf{w} - \bar{\mathbf{w}})^T A (\mathbf{w} - \bar{\mathbf{w}})}{(\mathbf{w} - \bar{\mathbf{w}})^T (\mathbf{w} - \bar{\mathbf{w}})}. \quad (12)$$

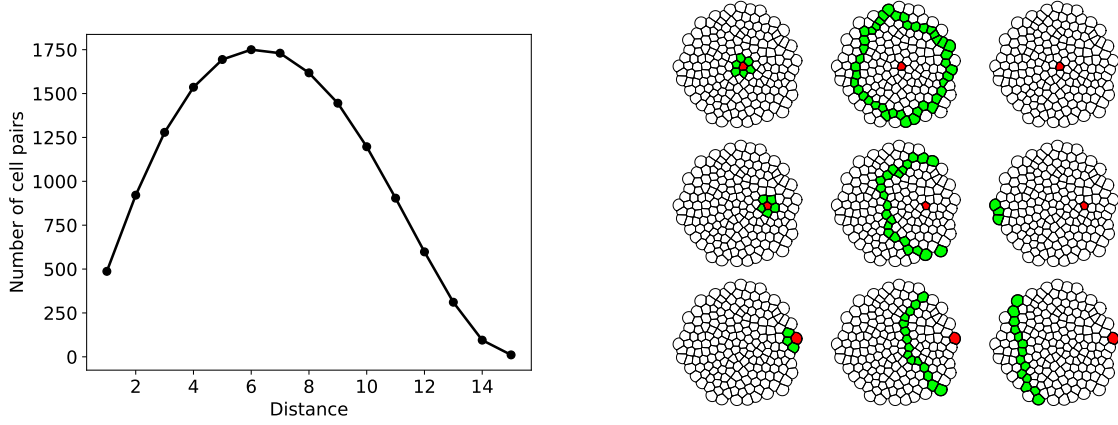


Figure 4: Distribution of all cell pairs  $|S_k|$  found for any distance  $k$  in the organoid presented in figure 3 (left). On the right, all cell pairs of equal distance relative to a single cell are visualized. The green cells represent all cells with distances  $k = 1$  (left),  $k = 6$  (center) and  $k = 12$  (right) towards the red cell. From top to bottom, the red cell was chosen closer to the boundary of the organoid.

Here,  $\bar{\mathbf{w}}$  denotes the mean of  $\mathbf{w}$ , whereas  $A$  is a matrix representation for the connectivity of the cells in an organoid. In this study,  $A$  is chosen as the adjacency matrix of graph  $G$ . Furthermore,  $\mathbf{w}$  is chosen to represent the fates of every cell. Therefore, the cell fates have to be interpreted as a numerical value. If there are only two cell fates like the  $u^+v^-$  and  $u^-v^+$  cells from the in silico organoids,  $\mathbf{w}$  can be written as a vector with  $w_i \in \{0, 1\}$  for  $i = 1, \dots, n$ . Here, the zeros and ones each stand for one of the cell fates. It is irrelevant which number is assigned to which fate, the resulting value of  $I$  does not change as a result.

**Regular grids** On a regular  $8 \times 8$  grid, it is possible to demonstrate how  $I$  behaves according to a given pattern (Fig. 5). In a checkerboard pattern, equal cells are never adjacent to each other as long as the number of neighbors do not exceed 4. Thus, there is no auto-correlation between the cells leading to the minimal value  $I = -1$  (Fig. 5 (a)). For randomly distributed cells,  $I$  takes values close to 0 (Fig. 5 (b)). In the case of spatial separation, cells are mostly neighbored to equal cell types leading to values of  $I$  close to 1 (Fig. 5 (c)).

**Limit example ( $I = 1$ )** The exact value  $I = 1$  is usually reserved for limit cases. Creating one such example is made possible by extending the pattern in figure 5 (c) to a  $k \times k$  grid. For simplicity, we can also assume a periodic boundary, such that every cell has exactly four neighbors. The number of cells is  $k^2$  and the sum of matrix entries yields

$$\sum_{i,j=1}^k = 4k^2. \quad (13)$$

Without loss of generality, assume that the cells on the left half have cell fate values  $w_i = 1$ . Since  $\mathbf{w}$  has equal amounts of ones and zeros, the values  $w_i - \bar{w}$  are  $1/2$  for cells on the left half and  $-1/2$  for cells on the right half. Altogether, the inner product yields

$$(\mathbf{w} - \bar{\mathbf{w}})^T (\mathbf{w} - \bar{\mathbf{w}}) = 1/4k^2. \quad (14)$$

Thus,  $I$  can be written as

$$I = \frac{k^2}{4k^2} \frac{(\mathbf{w} - \bar{\mathbf{w}})^T A (\mathbf{w} - \bar{\mathbf{w}})}{1/4k^2}, \quad (15)$$

$$= \frac{1}{k^2} (\mathbf{w} - \bar{\mathbf{w}})^T A (\mathbf{w} - \bar{\mathbf{w}}). \quad (16)$$

The final matrix vector product can be dissected into two parts. First, the cells that are in contact to cells of different types. There is one line of contact in the middle of the grid, and another one at the periodic boundary. This amounts to a total of  $4k$  cells that have exactly one neighbor that is different.



At the same time, the remaining  $k^2 - 4k$  cells have equal neighbors. This yields

$$(\mathbf{w} - \bar{\mathbf{w}})^T A(\mathbf{w} - \bar{\mathbf{w}}) = (k^2 - 4k) \left( \frac{1}{4} + \frac{1}{4} + \frac{1}{4} + \frac{1}{4} \right) + 4k \left( \frac{1}{4} + \frac{1}{4} + \frac{1}{4} - \frac{1}{4} \right). \quad (17)$$

$$= k^2 - 2k \quad (18)$$

Thus, the exact value of  $I$  is given by the following formula

$$I = \frac{k^2 - 2k}{k^2}. \quad (19)$$

Using l'Hospital's rule the limit  $k \rightarrow \infty$  yields  $I = 1$ .

**Irregular grids** On an irregular grid, where the number of neighbors of one cell  $i$  might be different to that of another cell  $j$ , i.e.  $N_G(i) \neq N_G(j)$  for some  $i, j \in 1, \dots, n$ , the possible minimum/maximum values of  $I$  and therefore the range of possible values changes. It is therefore necessary to classify a whole range of different patterns and draw reference values from them for comparison. Patterns made by spatial separation (Sec. 2.2) highlight a decreasing trend of  $I$ , as the number of spatial separations  $n_x$  and  $n_y$  increase (Fig. 6). This is consistent with the idea that clusters determine the value of  $I$ . By repeatedly separating the tissue into smaller compartments, cluster sizes and thus  $I$  will decrease. The span of  $I$  ranges from  $-0.23$  to  $0.75$ . For radially separated patterns, a similar behavior is observed for the first four organoids (Fig. 7). However, when  $n_r$  increases even further,  $I$  slightly increases again. Effects like these are to be expected when generating patterns in such a simplified way. In this specific case, the spatial resolution of the patterning given by  $n_r$  has simply become smaller than the actual cell sizes.

In conclusion,  $I$  can be used as a measure to determine the type of clustering of two different cell types in an organoid. Large values of  $I$  mean fewer but larger clusters, while low values mean more but smaller clusters. The values will especially make sense when comparing different types of organoids. However, Moran's  $I$  provides no further information about the spatial distribution of the cells. Combined with the PCFs, two- and three dimensional patterns can now be quantified at the local and global level.

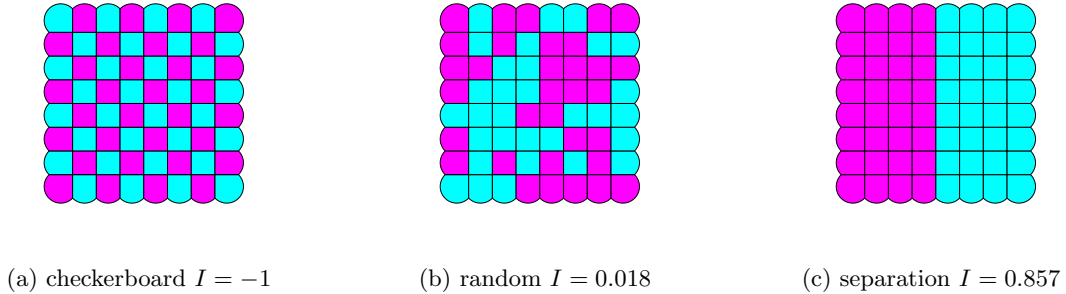


Figure 5

## 2.4 Data pre-processing

The goal of this section is to establish some familiarity with the experimental data. Furthermore, it will be explained in detail how the data are processed. The data in question was provided by [23]. In short, they were able to create organoids from ESCs of the mouse embryos ICM. The ESCs in question were extracted from the Epiblast portion of the mouse embryo and treated with doxycycline to reach a state of co-expression of NANOG and GATA6 [22]. Using these cells, aggregates of 200 mouse ESCs were seeded into concave wells that were additionally centrifuged to concentrate the cell aggregates in the middle. Afterwards, they were left to develop for 24 or 48 hours. Immunofluorescence staining of the resulting 3D organoids was performed according to [47]. Images were then acquired using a laser scanning confocal fluorescence microscope. The image pre-processing steps described in [23] together with the image analysis pipeline in [44] yield segmented images with fluorescence intensity values of NANOG and GATA6. Using  $k$ -means clustering, the cells were categorized into four different cell types according to [48]. These cell types are N+G-, N+G+, DN and DP. These abbreviations describe expressions of both NANOG and GATA6 in the cells. N+G- denote cells with high NANOG expression levels and low GATA6 expression levels. For N-G+ it is the other way around. DN or N-G- means "double negative", i.e. cells which exhibit low expressions of both NANOG and GATA6. On the other hand, DP or N+G+

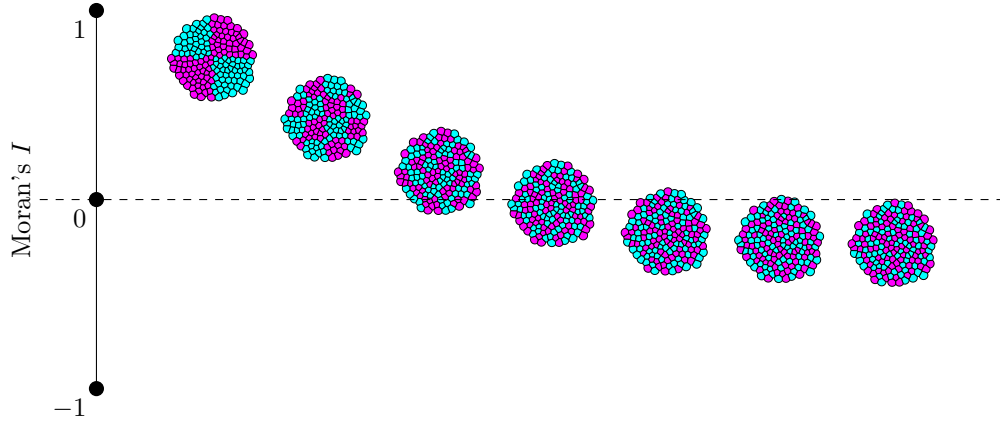


Figure 6: Graph showing Moran's  $I$  for organoids with cells separated in two spatial dimensions. The number of separations in each dimension increases by 2 for every new organoid from left to right. The center of the organoids correspond to their Moran's  $I$  and its corresponding position on the y-axis.

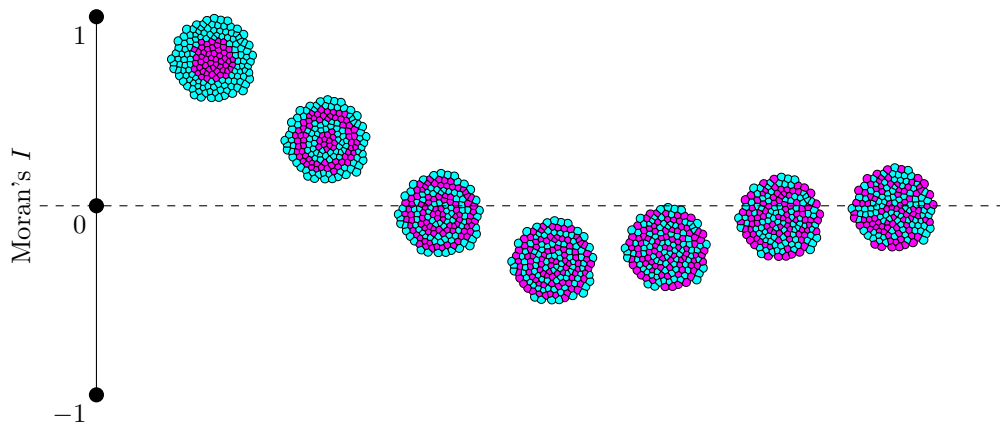


Figure 7: Graph showing Moran's  $I$  for organoids with radially separated cells. The number of separations in each dimension increases by 2 for every new organoid from left to right. The center of the organoids correspond to their Moran's  $I$  and its corresponding position on the y-axis.

stands for "double positive", i.e. high expressions of both transcription factors. N+G- and N-G+ cells shall describe Epi and PrE precursor cells respectively and will therefore be used interchangeably. In total, the dataset contains 76 different 3D organoids with 34 24h-organoids and 42 48h-organoids. For each organoid, the dataset provide the positional information of the centroid of each cell nucleus via their  $x$ -,  $y$ - and  $z$ -position. Furthermore, the cells have already been separated into different cell types N+G-, N-G+, DN and DP.

While N+G- and N-G+ cells are considered to be distinct cell precursors, DN and DP cells remain a source of uncertainty in the patterns in the ICM organoids. DP cells are mostly present in the early stages of the embryo, whereas DN populations increase towards later stages [48, 23]. Biologically, DN cells represent cells which are not yet committed to one of the cell fates. Contrary to this, DN cells are likely representing late stage Epi and PrE cells, that have started to down-regulate NANOG and GATA6 expressions [49, 50]. In the data, the combined percentages of DN and DP cells range from 10% to 78%. To take this uncertainty into consideration, DN and DP cell fates were randomly assigned to N+G- or N-G+. The probability  $p$  to assign a cell with N+G- fate equals the sum of all N+G- cells divided by the sum of all N+G- and N-G+ cells (Fig. 8). This ensures that the cell type proportions will not be changed by this assignment. This process is repeated 1000 times to generate a large number of samples of possible patterns created by a single organoid. This creates an envelope of values for the PCFs and Moran's  $I$  by taking the respective minimum/maximum as limiting values. Calculating the minimum/maximum values directly is in fact not feasible as this problem is NP-hard with a computational complexity of  $\mathcal{O}(2^n)$  with  $n$  denoting the combined number of DN and DP cells. As a fun side note, for  $n = 100$  on a computer with a 2GHz CPU this would amount to a computation time of roughly

$$\text{Time} = \frac{\text{Number of operations}}{\text{Operations per second}} = \frac{2^{100}}{2^9} \text{s} = 2^{91} \text{s} \approx 7.85 \cdot 10^{19} \text{y}. \quad (20)$$

Separate interactive 3D visualizations were created for the 24h (<https://schardts.github.io/Organoids24h>) as well as 48h (<https://schardts.github.io/Organoids48h>) organoids using the Plotly library in Python [51]. These contain the PCF envelope, the cell type proportions and a rotatable 3D model of the respective organoid (Fig. 9).

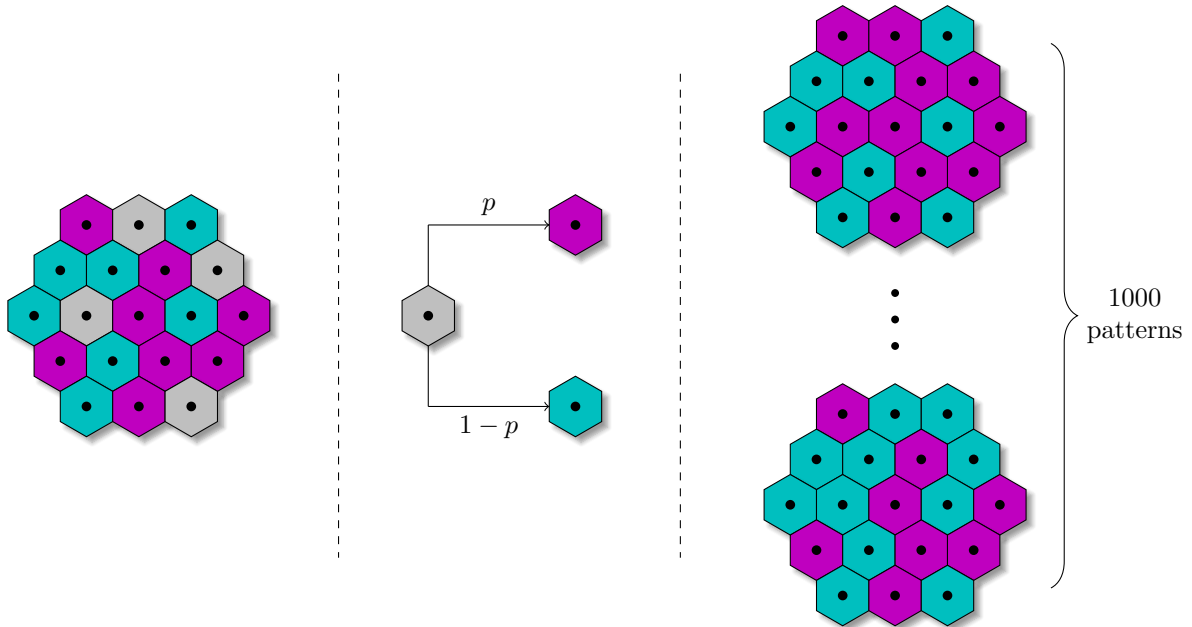


Figure 8: An illustration of the experimental data (left) shows three different cell types. N+G- cells are shown in magenta, N-G+ cells in cyan. Gray cells represent cells which are either DP or DN and are therefore not further identifiable. In the following (middle), each gray cell is randomly assigned a cell fate. Probability  $p$  reflects the proportions of N+G- cells found in the data. After 1000 repetitions, one obtains a variety of 1000 different patterns with only two cell types (right).

## 2.5 Computational methods

All computational tasks were implemented in Python 3. Graphs were processed using the *igraph* package. The graph distance was calculated using the `shortest_paths` function which uses an efficient

Organoid ID = 23, Number of cells = 1001

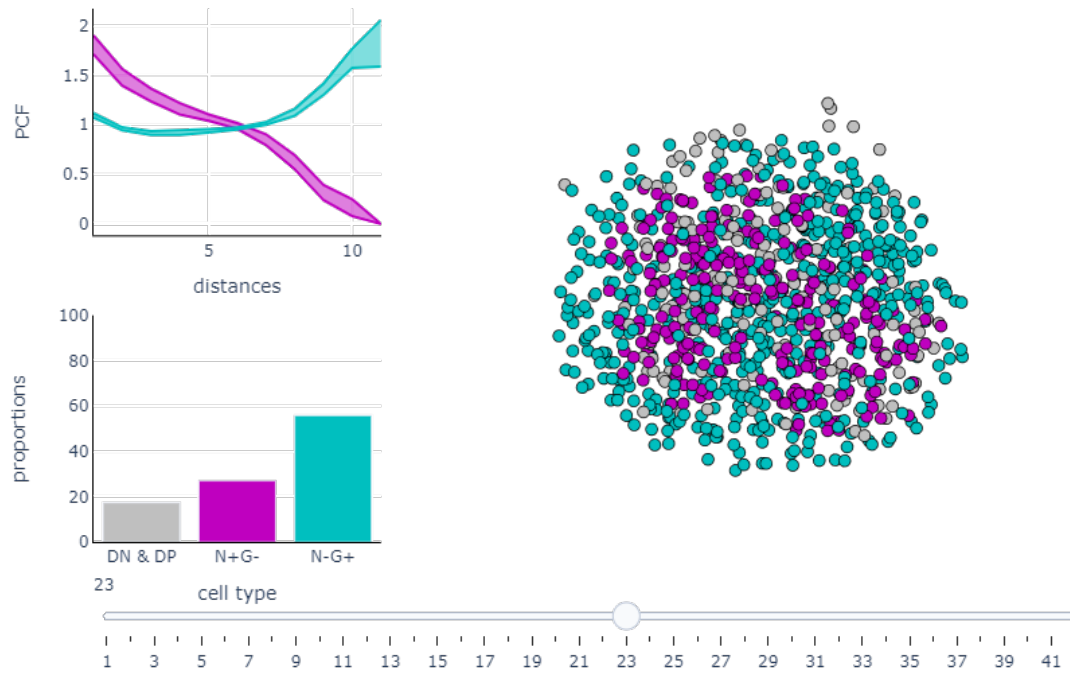


Figure 9: Snapshot of the 3D interactive plots generated from the experimental data. In the top left corner, an envelope of possible PCF values shows the spatial arrangement of the cells. Directly below, the cell proportions are shown divided into N+G-, N-G+ and the combined DN and DP proportions. On the right, the corresponding organoid is visualized as an interactive 3D model. A slider at the bottom allows the selection of different organoids. The color scheme remains the same for all three plots.

implementation of breadth first search (BFS) algorithm. *SciPy* [52] has been used for efficient Delaunay triangulations and Voronoi tessellations through `scipy.spatial`. Furthermore, solving nonlinear equations and the fitting of curves to data were performed via `scipy.optimize`. If possible, any mathematical operations were performed on arrays via *NumPy* [53]. Solving ODEs was done using the explicit Euler method. Computational visualizations were done either in *Matplotlib* [54] or *Plotly* [51]. Vector graphic illustrations were done in  $\text{\LaTeX}$  using the *TikZ* library [55] or Microsoft PowerPoint.

### 3 Models and analyses

#### 3.1 Transcriptional regulation

The fate of cells in the ICM during embryo development strongly correlates with the protein expression levels of NANOG and GATA6. In general, it is a reasonable assumption that a cell's fate is decided by various protein expression levels. The process of transcriptional regulation is driving the conversion of DNA to RNA, which is mainly responsible for the protein expression. In this section, the mystery of this regulation will be unraveled with the help of a mathematical basis that provides a better understanding and control of the underlying processes. Right at the center of this section is the gene regulatory network (GRN) surrounding NANOG and GATA6 (Fig. 10). At its core, the transcription factors NANOG, here represented as  $u$ , and GATA6, represented by  $v$ , inhibit each other [10, 11]. They are known to influence their own expression, leading to an auto-activation of both  $u$  and  $v$ . Through the FGF/Erk signaling pathway, ICM cells are able to influence their NANOG and GATA6 expressions intercellularly. FGF4 binding to its receptor FGFr2 increases the production of GATA6, while simultaneously inhibiting NANOG expressions [15, 16]. This process is greatly simplified in the GRN by an external signal  $s$ , which is not further specified at this time. The signal  $s$  acts as an inhibitor on  $u$  and an activator on  $v$ .

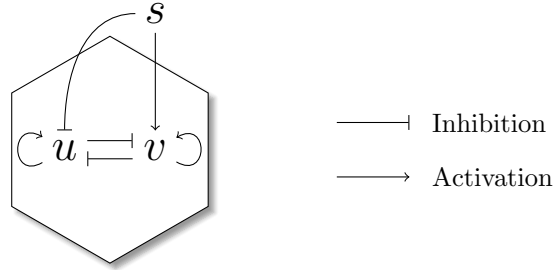


Figure 10: Illustration of the GRN used throughout this study. Transcription factors NANOG and GATA6 are represented by  $u$  and  $v$  respectively. Inside the cell,  $u$  and  $v$  mutually inhibit each other. An external signal  $s$  inhibits  $u$  and activates  $v$ .

##### 3.1.1 Generalized phenomenological models

**Two cell system:** This whole journey was initiated by the motivation that arose from the extensive analytical study of the Delta-Notch system in epithelial gut tissue of *Drosophila* in [26]. The Delta-Notch system is often referred to as a lateral inhibition model, due to cells of the neural cell fate inhibiting adjacent cells from adopting the same fate. However, the lateral mechanism involved is functionally not that of an inhibition, but rather an activation from adjacent cells. In [26], a single condition on two generic functions was able to decide if two cells are in two different states or in the same. A similar condition was pursued in this study. Thus, a simplified GRN, without auto-activation and adjacency-based inhibition, was considered in the first step (Fig. 11). This was translated into a generic system of ODEs

$$\begin{aligned}\frac{du_1}{dt} &= f(v_1) - u_1, \\ \frac{du_2}{dt} &= f(v_2) - u_2,\end{aligned}\tag{21}$$

$$\begin{aligned}\frac{dv_1}{dt} &= g(u_1) + h(u_2) - v_1, \\ \frac{dv_2}{dt} &= g(u_2) + h(u_1) - v_2.\end{aligned}$$

The equations reflect the GRN with  $\mathbf{u}$  and  $\mathbf{v}$ , now being vectors in  $\mathbb{R}^2$ , inhibiting each other inside the cell via the functions  $f$  and  $g$ . Simultaneously,  $v$  is activated by the expression of  $u$  in the neighboring cell, which is reflected by the function  $h$ . The type of interaction leads to conditions for trend of the respective functions. Inhibitions are strictly monotonically decreasing, whereas activations are strictly monotonically increasing, which for differentiable functions implies

$$f'(x) < 0, \quad g'(x) < 0, \quad h'(x) > 0\tag{22}$$

The adjacency-based activation replaces the activation through the external signal  $s$  and offers a slight foreshadowing of what is to come in this thesis. Auto-activation as well as the inhibition from the signal were neglected at this point. The last terms in ODE system (21) describe the exponential decay and therefore the lifetime of  $u$  and  $v$  with a decay rate of 1.

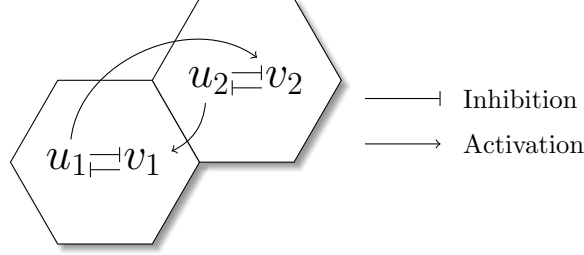


Figure 11: Schematic representation of the GRN describing two cells in contact to each other. Arrows inside the cell describe intracellular interactions of  $u$  and  $v$ , whereas arrows ranging from one cell to the other display the extracellular interactions, i.e. cell-cell communication.

The stability of a dynamic system provides valuable information about the functions and parameters of the system. An ODE system

$$\frac{dx_i}{dt} = F(x), \quad i = 1, \dots, M$$

is linearly stable in  $x^*$ , if its linearization matrix  $L^{\text{ODE}} = F'(x^*)$  has only eigenvalues with negative real part. For (21)  $L^{\text{ODE}}$  becomes

$$L^{\text{ODE}} = \begin{pmatrix} -1 & 0 & f'(v_1) & 0 \\ 0 & -1 & 0 & f'(v_2) \\ g'(u_1) & h'(u_2) & -1 & 0 \\ h'(u_1) & g'(u_2) & 0 & -1 \end{pmatrix} \quad (23)$$

The eigenvalues are then given as the roots of the characteristic polynomial  $\chi(\lambda) = \det(L^{\text{ODE}} - \lambda I_4)$ , where  $I_4$  denotes the identity matrix in  $\mathbb{R}^4$ . To simplify this, matrix  $L - \lambda I_4$  is divided into four  $2 \times 2$  block matrices

$$L^{\text{ODE}} - \lambda I_4 = \begin{pmatrix} -1 - \lambda & 0 & f'(v_1) & 0 \\ 0 & -1 - \lambda & 0 & f'(v_2) \\ g'(u_1) & h'(u_2) & -1 - \lambda & 0 \\ h'(u_1) & g'(u_2) & 0 & -1 - \lambda \end{pmatrix} =: \begin{pmatrix} A & B \\ C & D \end{pmatrix}. \quad (24)$$

The commutative properties  $CD = DC$  allows the determinant to be written as  $\det(L^{\text{ODE}} - \lambda I_4) = \det(AD - BC)$ , making it possible to calculate the characteristic polynomial by

$$\begin{aligned} \chi(\lambda) &= \det \left( \begin{pmatrix} (1 + \lambda)^2 & 0 \\ 0 & (1 + \lambda)^2 \end{pmatrix} - \begin{pmatrix} f'(v_1)g'(u_1) & f'(v_1)h'(u_2) \\ f'(v_2)h'(u_1) & f'(v_2)g'(u_2) \end{pmatrix} \right) \\ &= [(1 + \lambda)^2 - f'(v_1)g'(u_1)] [(1 + \lambda)^2 - f'(v_2)g'(u_2)] - f'(v_1)h'(u_2)f'(v_2)h'(u_1). \end{aligned} \quad (25)$$

In order to investigate if this system is capable of generating an inhomogeneous steady state, i.e. a state of two different cell types, it is sufficient to derive conditions for the instability of homogeneous steady states. A homogeneous steady state is reached, when both cells have equal expression values in equilibrium, i.e.

$$u_1 = u_2, \quad v_1 = v_2, \quad \frac{du}{dt} = \mathbf{0} = \frac{dv}{dt}.$$

This makes it possible to write the homogeneous steady state as

$$\begin{pmatrix} u_1 \\ u_2 \\ v_1 \\ v_2 \end{pmatrix} = \begin{pmatrix} u_1 \\ u_1 \\ v_1 \\ v_1 \end{pmatrix} = \begin{pmatrix} f(v_1) \\ f(v_1) \\ v_1 \\ v_1 \end{pmatrix} \quad (26)$$

Inserting this into (25) yields

$$\chi(\lambda) = [(1 + \lambda)^2 - (g \circ f)'(v_1)]^2 - [(h \circ f)'(v_1)]^2. \quad (27)$$

Here,  $(g \circ f)(u_1)$  denotes the chaining of two functions, i.e.  $f(g(u_1))$ . Using the chain rule  $g'(f(u_1))f'(u_1) = (f \circ g)'(u_1)$ . The eigenvalues  $\lambda$  are now determined by solving  $\chi(\lambda) = 0$  for  $\lambda$ . The first step is taking the square root of the equation and reorder it to find

$$(1 + \lambda)^2 = (g \circ f)'(v_1) \pm (h \circ f)'(v_1). \quad (28)$$

Taking again the square root finally yields the four different eigenvalues

$$\lambda_{1,2,3,4} = -1 \pm \sqrt{(g \circ f)'(v_1) \pm (h \circ f)'(v_1)}. \quad (29)$$

Due to the nature of the inhibition  $f' < 0$  and  $g' < 0$ . Contrary to that  $h' > 0$ . Combining the functions then yields

$$(g \circ f)' > 0, \quad (h \circ f)' < 0. \quad (30)$$

With this information, the eigenvalues can now be analyzed step by step. A condition for the instability of the homogeneous steady state is sought, this means  $\text{Re}(\lambda) > 0$ . Starting with the first "±" sign it can be seen that

$$\text{Re} \left( -1 - \sqrt{(g \circ f)'(v_1) \pm (h \circ f)'(v_1)} \right) \leq -1 < 0. \quad (31)$$

The inequality arises due to the square root being either positive or imaginary. For the second "±" sign, we know that  $(g \circ f)'(v_1) - (h \circ f)'(v_1) \geq (g \circ f)'(v_1) + (h \circ f)'(v_1)$ . Therefore, it is necessary to look at the "−" case. This results in the instability condition for the two cell system

$$(g \circ f)'(u_1) - (h \circ f)'(u_1) > 1. \quad (32)$$

For  $f = 0$ , the condition becomes identical to the one in [26], indicating that it was successfully extended by mutual inhibition. If condition (32) is fulfilled, the homogeneous steady state becomes unstable. In return, a heterogeneous steady state, i.e. a state with two different cell types, arises.

**One dimensional ring of  $n$  cells:** The above stability condition (32) can be extended to a one-dimensional cell line of  $n$  cells, where the first and last cell are connected to each other (Fig. 12). The resulting ODE system will then be

$$\begin{aligned} \frac{du_i}{dt} &= f(v_i) - u_i \\ \frac{dv_i}{dt} &= g(u_i) + h \left( \frac{u_{i+1} + u_{i-1}}{2} \right) - v_i, \quad i = 1, \dots, n. \end{aligned} \quad (33)$$

Here, the activation of  $v_i$  depends on the expression of  $u$  in the two neighboring cells. A periodic boundary condition,  $u_{n+1} = u_1$  and  $v_{n+1} = v_1$ , is applied, ensuring that the cell line has no loose ends leading to the representation as a ring of cells.

Assuming again a homogeneous steady state, i.e.  $u_i = u_j$  for  $i, j \in \{1, \dots, n\}$  and  $u_i = f(v_i)$ , the linearization of the system can be expressed via

$$L^{\text{ODE}} = \begin{pmatrix} -I_n & A \\ B & -I_n \end{pmatrix}. \quad (34)$$

The diagonals are the result of the decay in (33), whereas the interactions give rise to the off-diagonal matrices

$$A = \begin{pmatrix} f'(v_1) & 0 & \dots & 0 \\ 0 & \ddots & & \vdots \\ \vdots & & \ddots & 0 \\ 0 & \dots & 0 & f'(v_1) \end{pmatrix}. \quad (35)$$

$$B = \begin{pmatrix} g'(f(v_1)) & \frac{1}{2}h'(f(v_1)) & 0 & \dots & 0 & \frac{1}{2}h'(f(v_1)) \\ \frac{1}{2}h'(f(v_1)) & & & & & 0 \\ 0 & & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & & & 0 \\ 0 & & & & & \frac{1}{2}h'(f(v_1)) \\ \frac{1}{2}h'(f(v_1)) & 0 & \dots & 0 & \frac{1}{2}h'(f(v_1)) & g'(f(v_1)) \end{pmatrix} \quad (36)$$

$$(37)$$



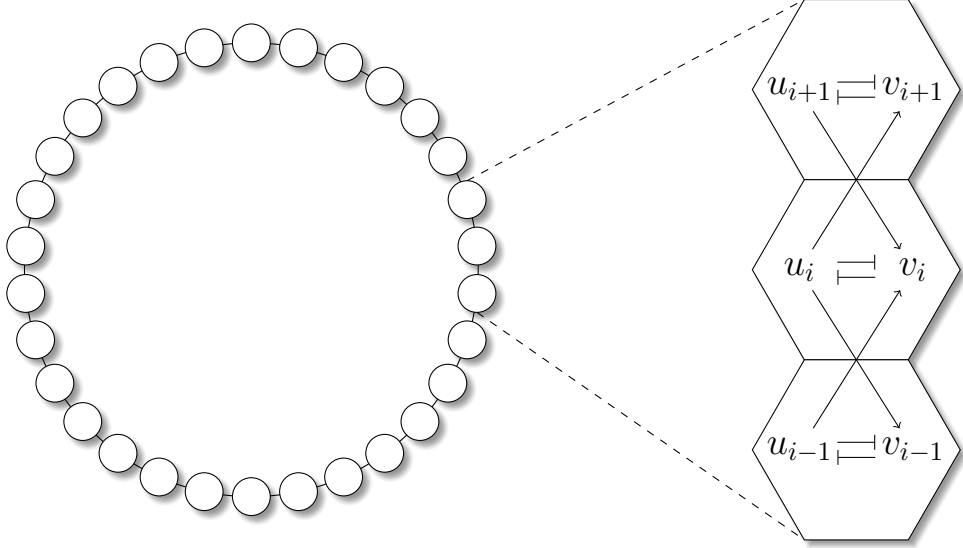


Figure 12: Schematic representation of a one-dimensional cell line arranged as a ring (left). The zoomed in portion depicts the GRN of three cells that are connected to each other (right).

Most importantly, the product  $AB$  yields the following

$$AB = \begin{pmatrix} (g \circ f)'(v_1) & \frac{1}{2}(h \circ f)'(v_1) & 0 & \dots & 0 & \frac{1}{2}(h \circ f)'(v_1) \\ \frac{1}{2}(h \circ f)'(v_1) & & & & & 0 \\ 0 & & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & & & 0 \\ 0 & & & & \frac{1}{2}(h \circ f)'(v_1) & (g \circ f)'(v_1) \\ \frac{1}{2}(h \circ f)'(v_1) & 0 & \dots & 0 & \frac{1}{2}(h \circ f)'(v_1) & (g \circ f)'(v_1) \end{pmatrix}. \quad (38)$$

This one will be helpful in calculating the determinant for the characteristic polynomial

$$\det(L^{\text{ODE}} - \lambda I_{2n}) = \det \begin{pmatrix} -(1+\lambda)I_n & A \\ B & -(1+\lambda)I_n \end{pmatrix}, \quad (39)$$

$$= \det((1+\lambda)^2 I_n - AB), \quad (40)$$

$$= \det(-AB - \mu I_n). \quad (41)$$

Again, the pairwise commutative properties of the involved block matrices lead to equality (40). With  $\mu = -(1+\lambda)^2$  the eigenvalue problem of  $L^{\text{ODE}}$  has been reduced to a different eigenvalue problem for  $-AB$  in (41). In order to calculate these, the following theorem from [56] (Theorem 3.1) is used.

**Theorem 1** Let  $A_n \in \mathbb{R}^n$  be a tridiagonal Toeplitz matrix with perturbation in the corners, i.e.

$$A_n = \begin{pmatrix} b+\gamma & c & 0 & \dots & 0 & \alpha \\ a & b & & & & 0 \\ 0 & & \ddots & \ddots & & \vdots \\ \vdots & & \ddots & & & 0 \\ 0 & & & & b & c \\ \beta & 0 & \dots & 0 & a & b+\delta \end{pmatrix}. \quad (42)$$

In addition to that, let  $\theta$  be a solution to

$$\left(\frac{a}{c}\right)^{n/2} \left[ ac \sin((n+1)\theta) + (\gamma\delta - \alpha\beta) \sin((n-1)\theta) - c \left(\frac{a}{c}\right)^{1/2} (\gamma + \delta) \sin(n\theta) - (c\alpha \left(\frac{a}{c}\right)^n + \alpha\beta) \sin(\theta) \right] = 0. \quad (43)$$

Then, the eigenvalues of  $A_n$  are given by

$$\mu = b + 2c \left(\frac{a}{c}\right)^{1/2} \cos(\theta). \quad (44)$$

When connecting (43) with  $-AB$ , the condition yields

$$\frac{1}{4}(h \circ f)'(v_1)^2 [\sin((n+1)\theta) - \sin((n-1)\theta) - 2\sin(\theta)] = 0. \quad (45)$$

The strict monotonicity of  $h$  and  $f$  guarantee that  $(h \circ f)'(v_1) \neq 0$ . In addition to that, the sine addition theorem,  $\sin(\alpha + \beta) = \sin(\alpha)\cos(\beta) + \cos(\alpha)\sin(\beta)$ , together with the symmetries  $\sin(\alpha) = -\sin(-\alpha)$  and  $\cos(\alpha) = \cos(-\alpha)$ , allow equation (45) to be reduced to

$$0 = \sin((n+1)\theta) - \sin((n-1)\theta) - 2\sin(\theta), \quad (46)$$

$$= \sin(n\theta)\cos(\theta) + \cos(n\theta)\sin(\theta) - \sin(n\theta)\cos(\theta) + \cos(n\theta)\sin(\theta) - 2\sin(\theta), \quad (47)$$

$$= 2\cos(n\theta)\sin(\theta) - 2\sin(\theta), \quad (48)$$

$$= 2\sin(\theta)[\cos(n\theta) - 1]. \quad (49)$$

Therefore, either  $\sin(\theta) = 0$  or  $\cos(n\theta) = 1$ . For  $\theta = \frac{2k\pi}{n}$  and  $k \in \mathbb{Z}$ , both conditions are covered. The resulting eigenvalues for  $-AB$  are then given by

$$\mu = -(g \circ f)'(v_1) - (h \circ f)'(v_1) \cos\left(\frac{2k\pi}{n}\right), \quad \text{for } k = 1, \dots, n. \quad (50)$$

Using  $\mu = -(1 + \lambda)^2$ , the eigenvalues for the initial problem of  $L^{\text{ODE}}$  are given by

$$\lambda = -1 \pm \sqrt{(g \circ f)'(v_1) + (h \circ f)'(v_1) \cos\left(\frac{2k\pi}{n}\right)}, \quad \text{for } k = 1, \dots, n \quad (51)$$

It can be seen that for  $n = 2$ , the solution from the two cell case (29) can be recovered. The homogeneous steady state then becomes unstable if

$$(g \circ f)'(v_1) + (h \circ f)'(v_1) \cos\left(\frac{2k\pi}{n}\right) > 1, \quad \text{for any } k = 1, \dots, n. \quad (52)$$

From a mathematician's point of view, it is a beautiful coincidence, that the circular arrangement of cells in figure 12 leads to the emergence of a trigonometric function. For instability, it suffices to fulfill (52) for a single  $k$ . The  $k \in \mathbb{Z}$  that lead to the largest value in (52) are the ones such that the value of  $\cos$  is closest to  $-1$ . For an even number of cells, this will be  $k = \frac{n}{2}$ , whereas for odd numbers, we get  $k = \frac{n+1}{2}$  or  $k = \frac{n-1}{2}$ , ultimately leading to the stability condition

$$\begin{aligned} (g \circ f)'(v_1) - (h \circ f)'(v_1) &> 1, & \text{if } n \text{ is even,} \\ (g \circ f)'(v_1) - (h \circ f)'(v_1) \cos\left(\frac{n+1}{n}2\pi\right) &> 1, & \text{if } n \text{ is odd.} \end{aligned} \quad (53)$$

With  $(h \circ f)' < 0$  and  $\cos < 1$ , this yields that the stability condition for the two cell system (32) is already sufficient to ensure stability for  $n$  cells. Extensions of this kind of stability analysis become increasingly more difficult when increasing the dimensions of the system. Thus, the focus will be shifted towards extensions of the GRN and the resulting system of equations.

**Two cell system with additional auto-activation and adjacency-based inhibition** In order to describe the actual GRN from figure 10, the auto-activation and the inhibition from the signal still have to be included. Similar to the activation, the inhibition is described as adjacency-based from  $u_1$  to  $u_2$  and vice-versa. So far, activatory and inhibitory effects were combined as a sum of two separate functions (21). At this point, this is taken one step further by combining all effects within a single generic function i.e.

$$\begin{aligned} \frac{du_1}{dt} &= f(u_1, u_2, v_1) - u_1, \\ \frac{du_2}{dt} &= f(u_2, u_1, v_2) - u_2, \end{aligned} \quad (54)$$

$$\begin{aligned} \frac{dv_1}{dt} &= g(u_1, u_2, v_1) - v_1, \\ \frac{dv_2}{dt} &= g(u_2, u_1, v_2) - v_2. \end{aligned}$$

In  $f$ , the first argument relates to the auto-activation, the second one the intercellular inhibition and the third one the intracellular inhibition. Similarly, in  $g$ , the first argument relates to the intracellular inhibition, the second one intercellular activation and the third one the auto-activation. The effect of the interactions can be summarized by their partial derivatives

$$\begin{aligned} f_1 &> 0, & f_2 &< 0, & f_3 &< 0 \\ g_1 &< 0, & g_2 &> 0, & g_3 &> 0. \end{aligned} \quad (55)$$

Here, the shorthand notation  $f_i := \frac{\partial}{\partial_i} f$  is used to denote the derivative with respect to the  $i$ -th argument. In order to increase readability, a homogeneous steady state will already be assumed at this point, i.e.  $u_1 = u_2$  and  $v_1 = v_2$ . Therefore the function arguments will be neglected in our further notation such that, e.g.  $f_1 = f_1(u_1, u_1, v_1)$ .

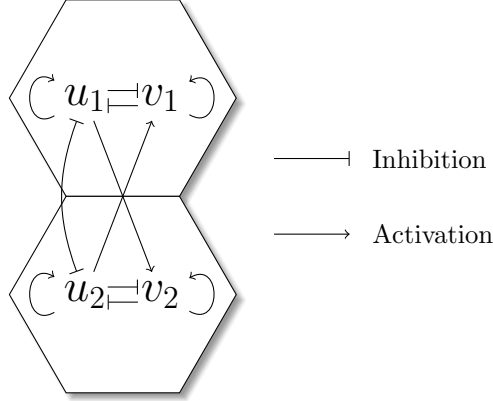


Figure 13: Schematic representation of the GRN for the fully generalized ODE system (54). Two cells in contact to each other communicate via intercellular activation and inhibition. Inside each cell,  $u$  and  $v$  mutually inhibit each other while activating themselves. From one cell to the other,  $u$  is activating  $v$  while simultaneously inhibiting  $u$ .

The stability analysis requires again the linearization of ODE system (54), which further leads to the problem of calculating the determinant of

$$L^{\text{ODE}} - \lambda I_4 = \begin{pmatrix} f_1 - 1 - \lambda & f_2 & f_3 & 0 \\ f_2 & f_1 - 1 - \lambda & 0 & f_3 \\ g_1 & g_2 & g_3 - 1 - \lambda & 0 \\ g_2 & g_1 & 0 & g_3 - 1 - \lambda \end{pmatrix} =: \begin{pmatrix} A & B \\ C & D \end{pmatrix} \quad (56)$$

Again, the commutativity  $CD = DC$  can be used to calculate the determinant

$$\chi(\lambda) = \det(L^{\text{ODE}} - \lambda I_4) = \det(AD - BC), \quad (57)$$

$$= \det \left( (g_3 - 1 - \lambda) \begin{pmatrix} f_1 - 1 - \lambda & f_2 \\ f_2 & f_1 - 1 - \lambda \end{pmatrix} - f_3 \begin{pmatrix} g_1 & g_2 \\ g_2 & g_1 \end{pmatrix} \right) \quad (58)$$

$$= \det \left( \begin{pmatrix} (g_3 - 1 - \lambda)(f_1 - 1 - \lambda) - f_3 g_1 & (g_3 - 1 - \lambda)f_2 - f_3 g_2 \\ (g_3 - 1 - \lambda)f_2 - f_3 g_2 & (g_3 - 1 - \lambda)(f_1 - 1 - \lambda) - f_3 g_1 \end{pmatrix} \right) \quad (59)$$

$$= [(g_3 - 1 - \lambda)(f_1 - 1 - \lambda) - f_3 g_1]^2 - [(g_3 - 1 - \lambda)f_2 - f_3 g_2]^2 \quad (60)$$

Setting  $\chi(\lambda) = 0$  and taking the square root of the equation yields

$$(g_3 - 1 - \lambda)(f_1 - 1 - \lambda) - f_3 g_1 \pm [(g_3 - 1 - \lambda)f_2 - f_3 g_2] = 0. \quad (61)$$

After rearranging all the terms, a somewhat complicated quadratic function is found

$$\lambda^2 + (2 - f_1 - g_3 \pm f_2)\lambda + [(g_3 - 1)(f_1 - 1) - f_3 g_1 \pm (g_3 f_2 - f_3 g_2 - f_2)]. \quad (62)$$

The use of  $\pm$  is supposed to show that both plus minus signs are connected such that their sign is always the same. Thus, there are two different quadratic equations, one for "+" and one for "-", resulting in four different eigenvalues given through the quadratic formula

$$\lambda_{1,2,3,4} = -\frac{2 - f_1 - g_3 \pm f_2}{2} \quad (63)$$

$$\pm \sqrt{\left(\frac{2 - f_1 - g_3 \pm f_2}{2}\right)^2 - [(g_3 - 1)(f_1 - 1) - f_3 g_1 \pm (g_3 f_2 - f_3 g_2 - f_2)]}. \quad (64)$$

As a small remark, it can be seen that equation (63) reduces to (29) by choosing

$$f_1 \rightarrow 0, \quad f_2 \rightarrow 0, \quad f_3 \rightarrow f', \quad (65)$$

$$g_1 \rightarrow g', \quad g_2 \rightarrow h', \quad g_3 \rightarrow 0. \quad (66)$$

There are two different cases to be considered. First, the one where the radical in (63) is less than zero. Second, the case where the radical is greater than zero.

**Radical less than zero** For radicals less than zero, the square root becomes imaginary, leading to the real part

$$\text{Re}(\lambda) = -\frac{2 - f_1 - g_3 \pm f_2}{2} \stackrel{!}{>} 0. \quad (67)$$

Using  $f_2 < 0$ , it is sufficient to look at one of the two cases

$$-\frac{2 - f_1 - g_3 - f_2}{2} > 0. \quad (68)$$

The terms can be rearranged to

$$f_1 + f_2 + g_3 < -1 \quad (69)$$

In the previous cases, without intercellular inhibition, i.e.  $f_2 = 0$ , this condition could not be fulfilled due to  $f_1 > 0$  and  $g_3 > 0$ . The radical in (63) being less than zero provides an additional condition that needs to be fulfilled

$$(g_3 - 1)(f_1 - 1) - f_3 g_1 \pm (g_3 f_2 - f_3 g_2 - f_2) > \left( \frac{2 - f_1 - g_3 \pm f_2}{2} \right)^2 > 0. \quad (70)$$

For the second inequality, the instability condition  $\text{Re}(\lambda) > 0$  was used. In total, this yields two inequalities which form the first instability condition when combined

$$\begin{aligned} f_1 + f_2 + g_3 &< -1 \\ (g_3 - 1)(f_1 - 1) - f_3 g_1 \pm (g_3 f_2 - f_3 g_2 - f_2) &> 0. \end{aligned} \quad (71)$$

**Radical greater than zero** For a generic quadratic equation,  $x^2 + px + q = 0$ , the solutions are given by

$$x = -\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q}. \quad (72)$$

If  $\left(\frac{p}{2}\right)^2 - q > 0$ , then the solutions  $x$  are positive if and only if

$$-\frac{p}{2} \pm \sqrt{\left(\frac{p}{2}\right)^2 - q} > 0 \quad (73)$$

$$\pm \sqrt{\left(\frac{p}{2}\right)^2 - q} > \frac{p}{2} \quad (74)$$

$$\left(\frac{p}{2}\right)^2 - q > \left(\frac{p}{2}\right)^2 \quad (75)$$

$$-q > 0. \quad (76)$$

Applied to (63) this yields the second instability condition for the homogeneous steady state

$$(g_3 - 1)(f_1 - 1) - f_3 g_1 \pm (g_3 f_2 - f_3 g_2 - f_2) < 0 \quad (77)$$

This inequality represents the exact opposite of the second inequality in (71). Thus, this is the first condition that needs to be checked, in order to prove instability of the homogeneous steady state.

In conclusion, it was proven that for the fully generalized ODE system (54) a stability/instability condition could be derived. From this point on however, it was decided that the results, as beautiful as they may be, do not justify the effort that must be made to handle any possible modifications of this system. Such modifications might include the jump from two to  $n$  cells, or from one to higher dimensions. Furthermore, cell-cell communication was so far limited to a single cell. While this provides enough space for further research, the objective of this thesis was shifted more in the direction of applications. Therefore, the remaining focus in this section will be spent on concrete examples of transcriptional regulation.

### 3.1.2 Phenomenological models

Currently, the most widespread element of modeling transcriptional regulation is the Hill function. Various examples of their use can be found for the regulation of NANOG and GATA6 in the mouse embryo [27, 28, 30, 31], but also in other system like the Delta-Notch signaling pathway in *Drosophila* [26]. Initially used to describe the dissociation curves of aggregates of hemoglobin molecules [36], the Hill equation has found its way into enzyme kinetics through the modeling of metabolic pathways [57]. In the context of transcriptional regulation, the rates of transcription are modified by the use of activatory and inhibitory Hill equations

$$\text{Activation:} \quad \frac{a_u u^k}{1 + a_u u^k}, \quad (78)$$

$$\text{Inhibition:} \quad \frac{1}{1 + a_u u^k}. \quad (79)$$

Here,  $u$  is the normalized concentration of the activator/inhibitor. The constant  $k \in \mathbb{N}$  is called Hill coefficient and describes the cooperativity  $u$ . For  $k = 1$ , this resembles the expressions from the Michaelis-Menten kinetics [58]. Another constant  $a_u$  describes the strength of the interaction. A simple ODE, in which the regulation of  $u$  is inhibited by another transcription factor  $v$ , can then be described as

$$\frac{du}{dt} = r_u \frac{1}{1 + a_v v^k} - \gamma_u u, \quad (80)$$

where  $r_u$  describes the basal transcription rate and  $\gamma_u$  the decay rate of  $u$ . For more complicated GRNs, equation (80) would have to include a combination of all the activations and inhibitions on the transcription of  $u$ . In literature, these are combined either additively or multiplicatively. Based on the qualitative reasoning mentioned in [27], the GRN (Fig. 10) yields

$$\begin{aligned} \frac{du}{dt} &= r_u \left( \frac{a_u u^k}{1 + a_u u^k} + \frac{1}{1 + a_s s^k} \right) \frac{1}{1 + a_v v^k} - \gamma_u u, \\ \frac{dv}{dt} &= r_v \left( \frac{a_v v^k}{1 + a_v v^k} + \frac{a_s s^k}{1 + a_s s^k} \right) \frac{1}{1 + a_u u^k} - \gamma_v v. \end{aligned} \quad (M1)$$

Although presentation slightly differs, ODE system (M1) is an exact replication of the first two equations of the ODE system found in [27, 28]. This system in particular has already been thoroughly investigated and is therefore only used as a reference system to compare the other systems.

### 3.1.3 Thermodynamic models

The main goal of this section is to establish a dynamical system describing the gene regulation in cells using methods from statistical thermodynamics. The basic concept of the model underlies only two assumptions:

1. Transcription determines the production of new protein.
2. Decay describes the lifetime of the protein.

These assumptions are translated into an ordinary differential equation (ODE) describing the concentration of protein  $u$  over time:

$$\frac{du}{dt} = r p_u - \gamma u. \quad (81)$$

The second term is the exponential decay with decay rate  $\gamma$ . The first term describes the rate of transcription of the corresponding gene. Here,  $p_u$  denotes the probability that RNA polymerase (RNAP) is bound to the promoter of  $u$ . The production rate  $r$  describes how much protein can be produced while RNAP is bound.

**RNAP binding probability** In the absence of activating/inhibiting transcription factors, the binding probability of RNAP to the promoter can be determined using statistical mechanics as in [37, 39]. To this end, the DNA is divided into  $\Omega$  fixed lattice sites, plus one site for the promoter of  $u$ . Considering  $X$  distinct RNAP enzymes, the number of possibilities that they are bound or unbound are

$$\text{Number of unbound states:} \quad \frac{\Omega!}{X!(\Omega - X)!} \quad (82)$$

$$\text{Number of bound states:} \quad \frac{\Omega!}{(X - 1)!(\Omega - X + 1)!} \quad (83)$$

For different state energies  $\varepsilon^{\text{unbound}}$  and  $\varepsilon^{\text{bound}}$ , the Boltzmann distributions

$$e^{-\beta\varepsilon^{\text{unbound}}} \quad \text{and} \quad e^{-\beta\varepsilon^{\text{bound}}} \quad (84)$$

describe the probability of adopting that state. The constant  $\beta$  is the inverse product of the Boltzmann constant and temperature  $\frac{1}{k_B T}$ . The state energies can be split into the sum of the state energies of the single RNAP enzymes

$$\varepsilon^{\text{unbound}} = X\varepsilon_x^{\text{unbound}} \quad (85)$$

$$\varepsilon^{\text{bound}} = (X-1)\varepsilon_x^{\text{unbound}} + \varepsilon_x^{\text{bound}}. \quad (86)$$

The partition function is used to describe the statistical properties of the given ensemble. It is defined as the sum of the Boltzmann distributions over all possible microstates

$$Z^{\text{total}} = \sum_{\text{microstates}} e^{-\beta\varepsilon^{\text{microstate}}}, \quad (87)$$

$$= \frac{\Omega!}{X!(\Omega-X)!} e^{-\beta\varepsilon^{\text{unbound}}} + \frac{\Omega!}{(X-1)!(\Omega-X+1)!} e^{-\beta\varepsilon^{\text{bound}}}, \quad (88)$$

$$= Z^{\text{unbound}} + Z^{\text{bound}}. \quad (89)$$

The binding probability  $p_u$  of  $X$  binding to the promoter of  $u$  is then given by the ratio of bound states and all states

$$p_u = \frac{Z^{\text{bound}}}{Z^{\text{unbound}} + Z^{\text{bound}}}. \quad (90)$$

It is safe to assume that the number of lattice sites on the DNA far exceeds the amount RNAP, i.e.  $\Omega \gg X$ . Therefore,  $\frac{\Omega!}{(\Omega-X)!} \approx \Omega^X$  is a valid approximation for the combinatorial terms. The probability then becomes

$$p_u = \frac{Z^{\text{bound}}/Z^{\text{unbound}}}{1 + Z^{\text{bound}}/Z^{\text{unbound}}} = \frac{\frac{X}{\Omega} e^{-\beta(\varepsilon_x^{\text{bound}} - \varepsilon_x^{\text{unbound}})}}{1 + \frac{X}{\Omega} e^{-\beta(\varepsilon_x^{\text{bound}} - \varepsilon_x^{\text{unbound}})}}. \quad (91)$$

In the terms corresponding to the bound state, only the difference of energies remains. Therefore they are replaced by  $\Delta\varepsilon_x = \beta(\varepsilon_x^{\text{bound}} - \varepsilon_x^{\text{unbound}})$ . Although,  $\Delta\varepsilon$  does not have the physical unit of energy, it will still be referred to as energy difference. The ratio  $x := \frac{X}{\Omega}$  describes the volume fraction of RNAP in the system. Together, this yields

$$p_u = \frac{e^{-\Delta\varepsilon_x x}}{1 + e^{-\Delta\varepsilon_x x}}. \quad (92)$$

For shorthand notation, the Boltzmann coefficients are replaced by an energy coefficient  $\eta_x := e^{-\Delta\varepsilon_x}$  to reveal

$$p_u = \frac{\eta_x x}{1 + \eta_x x} =: p_0. \quad (93)$$

This is equivalent to the rate of product formation in the case of the Michaelis-Menten kinetics. In this context,  $p_0$  describes the basal probability of RNAP binding.

**First order interactions** If a single interacting agent  $A$  influences the chance of binding  $X$  to the promoter by means of activation or inhibition, additional binding events have to be considered. This changes the corresponding number of different states as well as the state energies. From table 1, it can be seen that these are determined straightforward with the exception of the event where both  $X$  and  $A$  are bound. In this case, an additional energy term  $\Delta\varepsilon_{xa}$  was introduced to describe the effect  $A$  has on the binding of  $X$ .

Binding Event	Number of States	Coefficient
$X$ unbound $A$ unbound	$\frac{\Omega!}{X!A!(\Omega-X-A)!}$	1
$X$ bound $A$ unbound	$\frac{\Omega!}{(X-1)!A!(\Omega-X-A+1)!}$	$e^{-\Delta\varepsilon_x}$
$X$ unbound $A$ bound	$\frac{\Omega!}{X!(A-1)!(\Omega-X-A+1)!}$	$e^{-\Delta\varepsilon_a}$
$X$ bound $A$ bound	$\frac{\Omega!}{(X-1)!(A-1)!(\Omega-X-A+2)!}$	$e^{-\Delta\varepsilon_x - \Delta\varepsilon_a - \Delta\varepsilon_{xa}}$

Table 1: List of different binding events and the corresponding number of possible states and coefficients for a single interactor.

After defining again the energy coefficients  $\eta_\xi := e^{-\Delta\varepsilon_\xi}$  with  $\xi \in \{x, a, xa\}$ , the same procedure as before ultimately leads to the binding probability of RNAP at the promoter of  $u$  under the influence of an interactor  $a$

$$p_u = \frac{\eta_x x (1 + \eta_{xa} \eta_a a)}{1 + \eta_x x (1 + \eta_{xa} \eta_a a) + \eta_a a}. \quad (94)$$

At first glance, (94) looks intimidating. However, there lies some beauty in the interpretation of this equation. Any term in this equation describes a different binding event. The numerator collects all events in which RNAP  $x$  is bound, whereas the denominator is the collection of all possibilities. The coefficients  $\eta_x$  and  $\eta_a$  describe the strength of binding without any type of interaction. In contrast to that,  $\eta_{xa}$  describes only the effect of the interaction between  $x$  and  $a$ . To see the effect of these interactions, we rewrite equation (94) by dividing both numerator and denominator by  $(1 + \eta_a \eta_{xa} a)$  to obtain

$$p_u = \frac{\eta_x x}{f_u + \eta_x x} \quad \text{with} \quad f_u := \frac{1 + \eta_a a}{1 + \eta_a \eta_{xa} a}. \quad (95)$$

The regulatory term  $f_u$  enables us to characterize an inhibition or activation by checking its value in comparison to 1. This is directly influenced by the interaction coefficient  $\eta_{xa}$ :

$$\begin{aligned} \eta_{xa} < 1 &\implies f_u > 1 \implies p_u < p_0 \implies \text{Inhibition} \\ \eta_{xa} > 1 &\implies f_u < 1 \implies p_u > p_0 \implies \text{Activation} \\ \eta_{xa} = 1 &\implies f_u = 1 \implies p_u = p_0 \implies \text{No Interaction} \end{aligned} \quad (96)$$

A special case of inhibition is found for  $\eta_{xa} = 0$ . In this case  $-\Delta\varepsilon_{xa} = \infty$ , i.e. it is impossible for both  $X$  and  $A$  to bind together. Throughout this study, this case is referred to as blocking inhibition.

**Higher order interactions** Higher order interactions describe the cooperativity of two or more interactors participating in the binding of the RNAP. By adding a single interactor  $b$ , the realm of possibilities then increases by the additional single binding events, the additional interactions with the RNAP but also the interactions between the constituents themselves. The regulatory term  $f_u$  is updated using the following terms:

1.  $B$  bound  $\implies \eta_b b$
2.  $A$  and  $B$  bound  $\implies \eta_a \eta_b \eta_{ab} ab$
3.  $X$  and  $B$  bound  $\implies \eta_x \eta_b \eta_{xb} xb$
4.  $X$ ,  $A$  and  $B$  bound  $\implies \eta_x \eta_a \eta_b \eta_{xa} \eta_{xb} \eta_{ab} \eta_{xab} xab$

Combining the previously found terms with these yields

$$f_u = \frac{1 + \eta_a a + \eta_b b + \eta_a \eta_b \eta_{ab} ab}{1 + \eta_a \eta_{xa} a + \eta_b \eta_{bx} b + \eta_a \eta_b \eta_{xa} \eta_{xb} \eta_{ab} \eta_{xab} xab}. \quad (97)$$

The amount of coefficients becomes staggering at some point. For  $N$  interactors, this would amount to  $2^N$  terms in both numerator and denominator. Each of these but one introducing its own coefficient, leading to a total of  $2^{N+1} - 2$  coefficients.

**Independence** Although the number of coefficients increases with each new interactor, there are usually some kinds of interactions we might be able to exclude from the start. Binding sites on the DNA which are neither shared nor tampered with by interactors are independent of each other, i.e.  $\eta_{ab} = 1$ . If we assume in addition to that, that there is no interaction between RNAP and the combined interactors, i.e.  $\eta_{xab} = 1$ , the regulatory term becomes

$$f_u = \frac{(1 + \eta_a a)(1 + \eta_b b)}{(1 + \eta_a \eta_{ax} a)(1 + \eta_b \eta_{bx} b)}. \quad (98)$$

In general, we can summarize  $N$  independent interactors  $a_i$  with  $i = 1, \dots, N$  by using the product of their individual regulatory terms

$$f_u = \prod_{i=1}^N \frac{1 + \eta_{a_i} a_i}{1 + \eta_{a_i} \eta_{a_i x} a_i} \quad (99)$$

**Application to NANOG/GATA6** The problem of NANOG and GATA6 regulation is modeled according to the proposed GRN (Fig. 10). Each of its interactions has to be translated into corresponding regulatory terms that make up the binding probability. First,  $u$  acts as an activator on its own promoter. This regulatory effect is expressed via

$$f_u^{(1)} = \frac{1 + \eta_u u}{1 + \eta_u \eta_{xu} u}. \quad (100)$$

Both  $u$  and  $s$  serve as an inhibitor on the promoter of  $u$ . For the sake of simplicity, we assume that the binding of RNAP gets completely blocked, which means the interaction coefficient is zero, leading to

$$f_u^{(2)} = \frac{1 + \eta_v v}{1}, \quad f_u^{(3)} = \frac{1 + \eta_s s}{1}. \quad (101)$$

In total, this yields a regulatory term

$$f_u = f_u^{(1)} \cdot f_u^{(2)} \cdot f_u^{(3)} = \frac{(1 + \eta_u u)(1 + \eta_v v)(1 + \eta_s s)}{1 + \eta_u \eta_{xu} u}. \quad (102)$$

It is assumed that the binding of  $u$ ,  $v$  and  $s$  is equally likely in the vicinity of both promoters. This way, the coefficients  $\eta_u$ ,  $\eta_v$  and  $\eta_s$  can be directly transferred to the promoter of  $v$ . Since,  $v$  is also able to auto-regulate and is inhibited by  $u$ , the first two regulatory effects are

$$f_v^{(1)} = \frac{1 + \eta_v v}{1 + \eta_v \eta_{xv} v}, \quad f_v^{(2)} = \frac{1 + \eta_u u}{1}. \quad (103)$$

The additional activation from the signal  $s$  yields

$$f_v^{(3)} = \frac{1 + \eta_s s}{1 + \eta_s \eta_{xs} s}. \quad (104)$$

The complete regulatory term then reads

$$f_v = f_v^{(1)} \cdot f_v^{(2)} \cdot f_v^{(3)} = \frac{(1 + \eta_u u)(1 + \eta_v v)(1 + \eta_s s)}{(1 + \eta_v \eta_{xv} v)(1 + \eta_s \eta_{xs} s)}. \quad (105)$$

Expressions (102) and (105) allow us to describe the temporal evolution of  $u$  and  $v$  inside a single cell via

$$\begin{aligned} \frac{du}{dt} &= r_u \frac{\eta_{xx}}{f_u + \eta_{xx}} - \gamma_u u & \text{with } f_u &= \frac{(1 + \eta_u u)(1 + \eta_v v)(1 + \eta_s s)}{1 + \eta_u \eta_{xu} u}, \\ \frac{dv}{dt} &= r_v \frac{\eta_{xx}}{f_v + \eta_{xx}} - \gamma_v v & \text{with } f_v &= \frac{(1 + \eta_u u)(1 + \eta_v v)(1 + \eta_s s)}{(1 + \eta_v \eta_{xv} v)(1 + \eta_s \eta_{xs} s)}. \end{aligned} \quad (M2)$$

**How the signal influences the steady states** Mathematically, the results of inhibition and activation in (M2) manifest in the steady states of the ODE system deviating from the basal expression levels. This means that the regulatory terms  $f_u$  and  $f_v$  are either less or greater than one in steady state. Based on these regulatory terms, the aim is to show that the signal  $s$  determines the emergence of steady states, where  $u$  is high and  $v$  is low, i.e.  $f_u < 1$  and  $f_v > 1$ , or vice-versa. This can be accomplished by finding a parameter combination that allows basal expression values for both  $u$  and  $v$  to be generated for a given signal value despite the interactions. Thus, increasing the signal value will increase  $f_v$  while simultaneously decreasing  $f_u$ . The basal expression levels for  $u$  and  $v$  are the steady state solutions for  $f_u = 1 = f_v$

$$u_0 = \frac{r_u}{\gamma_u} \frac{\eta_{xx}}{1 + \eta_{xx}}, \quad v_0 = \frac{r_v}{\gamma_v} \frac{\eta_{xx}}{1 + \eta_{xx}}. \quad (106)$$

The goal is to control parameter values  $\eta_s$  and  $\eta_{xs}$  in such a way that for a given signal  $s_0$  the system will end up at  $u_0$  and  $v_0$ . Given that  $f_u(u, v)$  and  $f_v(u, v)$  control the deviation from the basal activity, it is sufficient to find parameter relations fulfilling

$$f_u(u_0, v_0, s_0) = 1, \quad (107)$$

$$f_v(u_0, v_0, s_0) = 1. \quad (108)$$

This yields two equations that are rearranged to

$$(1 + \eta_u u_0)(1 + \eta_v v_0)(1 + \eta_s s_0) = 1 + \eta_u \eta_{xu} u_0, \quad (109)$$

$$(1 + \eta_u u_0)(1 + \eta_v v_0)(1 + \eta_s s_0) = (1 + \eta_v \eta_{xv} v_0)(1 + \eta_s \eta_{xs} s_0). \quad (110)$$



This nonlinear equation system can be solved in two steps. First (109) is solved for  $\eta_s$  to reveal

$$\eta_s = \frac{1}{s_0} \left( \frac{1 + \eta_u \eta_{xu} u_0}{(1 + \eta_u u_0)(1 + \eta_v v_0)} - 1 \right). \quad (111)$$

In the second step, (110) is solved for the second coefficient

$$\eta_{xs} = \frac{1}{\eta_s s_0} \left( \frac{(1 + \eta_v v_0)(1 + \eta_u u_0)(1 + \eta_s s_0)}{1 + \eta_u \eta_{xv} v_0} - 1 \right) =: \eta_{xs}^*. \quad (112)$$

This enables the characterization of different behaviors regarding ODE system (M2) by varying a few key parameters (Fig. 14). Increases in  $\eta_x x$  also increase the basal expression values. Together with the basal expressions, the overall values for both  $u$  and  $v$  at given signal values also increase, as indicated by the curve of steady states being shifted towards the upper right corner. The areas that characterize the cell fate are determined by the basal expression values  $u_0$  and  $v_0$  which only depend on  $\eta_x x$ . Hence, the increase of  $\eta_x x$  increases the area for the  $u^- v^-$  fate, whereas the area for the  $u^+ v^+$  fate decreases. The regions deciding the cell fates  $u^+ v^-$  and  $u^- v^+$  do not change in their areas but in their respective boundaries. The increase of the signal  $s$  displaces the steady states from the  $u^- v^+$  towards the  $u^+ v^-$  region. For  $\eta_{xs} = \eta_{xs}^*$  these pass right through the point where the basal expressions meet, as expected by condition (107). This effectively generates a system, where only the two states are possible. When deviating from  $\eta_{xs}$  the steady states first need to pass through one of the other two regions. A decrease of  $\eta_{xs}$  allows cells to be in  $u^- v^-$  state, whereas an increase enables the possibility of  $u^+ v^+$  cells.

### 3.1.4 Auto-activation dominant model

Model (M2) was derived purely from physical principles fulfilling the goal of replacing the phenomenological model (M1). However, at the analytical level, both models are difficult to handle in terms of stability. In this section, an additional model for the transcriptional regulation of  $u$  and  $v$  is introduced. For this model, the generality of (M2) is traded for an analytical description of the steady states of the system. This is realized by the assumption that the auto-activations of both  $u$  and  $v$  are dominant compared to the basal activity of RNAP. Hence, the initiation of transcription will only be associated with the binding of  $u$  or  $v$  to their respective promoters. Under this assumption, the interactions at the promoter are no longer described at the RNAP level, but only for the initiators of the transcription, e.g. if  $v$  inhibits  $u$  the interaction is described via  $\eta_{uv} < 1$  or even  $\eta_{uv} = 0$ . This way, all the interactions from GRN 10 are

$$\eta_{uv} = 0, \quad \eta_{vu} = 0, \quad (113)$$

$$\eta_{us} = 0, \quad \eta_{vs} > 1. \quad (114)$$

Higher order interactions are again neglected, e.g.  $\eta_{uvs} = 0$ . As before, states with a single binding event result in the addition of the terms  $\eta_\alpha \alpha$  with  $\alpha \in \{u, v, s\}$ . It remains a single state where  $v$  and  $s$  are bound cooperatively, yielding the term  $\eta_v \eta_s \eta_{vs} v s$ . The binding probability of  $u$  is described by the ratio of events where  $u$  is bound and all possible events

$$p_u = \frac{\eta_u u}{1 + \eta_v v(1 + \eta_s \eta_{vs} s) + \eta_u u + \eta_s s}. \quad (115)$$

Likewise, the binding probability of  $v$  is

$$p_v = \frac{\eta_v v(1 + \eta_s \eta_{vs} s)}{1 + \eta_v v(1 + \eta_s \eta_{vs} s) + \eta_u u + \eta_s s}. \quad (116)$$

For  $n$  different cells, the transcriptional regulation is then described by the ODE system:

$$\begin{aligned} \frac{du_i}{dt} &= r_u \frac{\eta_u u_i}{1 + \eta_v v_i(1 + \eta_s \eta_{vs} s_i) + \eta_u u_i + \eta_s s_i} - \gamma_u u_i \\ \frac{dv_i}{dt} &= r_v \frac{\eta_v v_i(1 + \eta_s \eta_{vs} s_i)}{1 + \eta_v v_i(1 + \eta_s \eta_{vs} s_i) + \eta_u u_i + \eta_s s_i} - \gamma_v v_i, \quad i = 1, \dots, n. \end{aligned} \quad (M3)$$

As before, the absorbed signals  $s_i$  are considered to be as general as possible. Thus, it will be described as a function

$$\mathbf{s} : \mathbb{R}^n \times \mathbb{R}^n \rightarrow \mathbb{R}^n : (\mathbf{u}, \mathbf{v}) \mapsto \mathbf{s}(\mathbf{u}, \mathbf{v}) \quad (117)$$

In the following, a thorough analysis of this model reveals several advantages over the other models considered throughout this study.

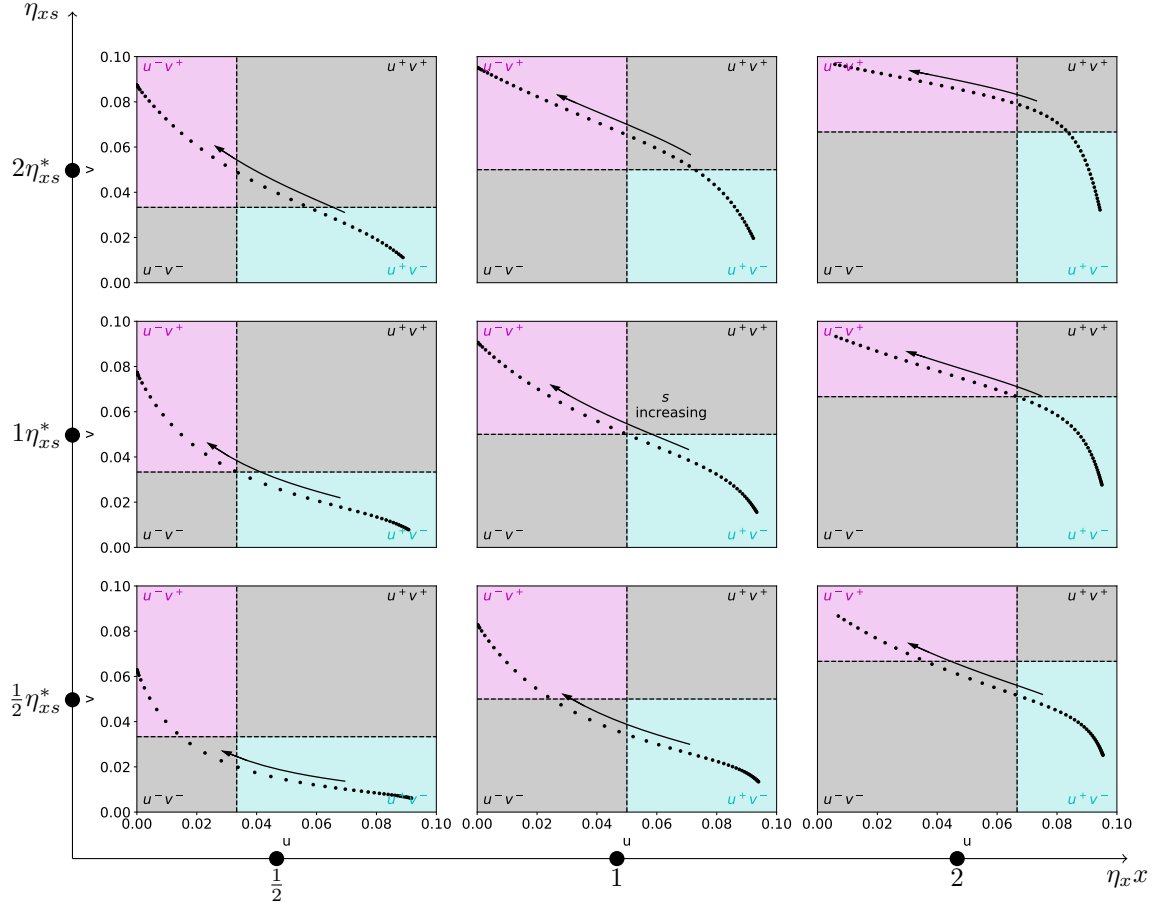


Figure 14: Steady states of the thermodynamic equations for transcriptional regulation ( $M2$ ) under variation of  $\eta_x x$  and  $\eta_{xs}$ . From left to right  $\eta_x x$  increases. From bottom to top  $\eta_{xs}$  increases. Dashed lines indicate the basal expressions  $u_0$  and  $v_0$ , separating regions of high expressions from regions of low expressions. Regions where both  $u$  and  $v$  are expressed either high are low are shaded in gray. Regions where  $u$  is high and  $v$  is low are shown in cyan, for the opposite case in magenta. Any dot represents the steady state of the system for one of 50 logarithmically spaced signal values  $s \in [10^{-3}, 1]$ . A black arrow indicates in which direction the steady states move upon increasing  $s$ . The remaining model parameters were chosen as  $\gamma_u = \gamma_v = 10$ ,  $r_u = r_v = 1$ ,  $-\Delta\varepsilon_u = -\Delta\varepsilon_v = 5$ ,  $-\Delta\varepsilon_{xu} = 4$ ,  $-\Delta\varepsilon_{vx} = 1$ ,  $s_0 = 0.05$ . Equations (111) and (112) yield  $\eta_s \approx 53.40$  and  $\eta_{xs} \approx 5.17$  which translates to  $-\Delta\varepsilon_s \approx 3.98$  and  $-\Delta\varepsilon_{xs} \approx 1.64$ .

**Steady states** The steady states of system (*M3*) can explicitly be calculated by setting  $\frac{du_i}{dt} = 0 = \frac{dv_i}{dt}$ . This yields

$$\frac{\eta_u u_i}{1 + \eta_u v_i(1 + \eta_s \eta_{vs} s_i) + \eta_u u_i + \eta_s s_i} = \frac{\gamma_u}{r_u} u_i, \quad (118)$$

$$\frac{\eta_v v_i(1 + \eta_s \eta_{vs} s_i)}{1 + \eta_v v_i(1 + \eta_s \eta_{vs} s_i) + \eta_u u_i + \eta_s s_i} = \frac{\gamma_v}{r_v} v_i. \quad (119)$$

Rearranging (118) and (119), reveals two possible solutions for  $u_i$  and  $v_i$ , respectively

$$u_i = \begin{cases} 0 \\ \frac{r_u}{\gamma_u} - \frac{1 + \eta_v v_i(1 + \eta_s \eta_{vs} s_i) + \eta_s s_i}{\eta_u} \end{cases}, \quad v_i = \begin{cases} 0 \\ \frac{r_v}{\gamma_v} - \frac{1 + \eta_u u_i + \eta_s s_i}{\eta_v(1 + \eta_s \eta_{vs} s_i)} \end{cases} \quad (120)$$

Combining the different solutions from (120) leads to a total of four different steady states. Three of these show either no expression of  $u$  and  $v$ , i.e.  $u^-v^-$ , or high expression in one of the transcription factors and none in the other, i.e.  $u^+v^-$  or  $u^-v^+$

$$u_i = 0, \quad v_i = 0 \quad (121)$$

$$u_i = \frac{r_u}{\gamma_u} - \frac{1 + \eta_s s_i}{\eta_u}, \quad v_i = 0 \quad (122)$$

$$u_i = 0, \quad v_i = \frac{r_v}{\gamma_v} - \frac{1 + \eta_s s_i}{\eta_v(1 + \eta_s \eta_{vs} s_i)} \quad (123)$$

When combining the non-zero solutions for  $u_i$  and  $v_i$  in (120), a fourth steady state is found. However, it turns out that  $u_i$  and  $v_i$  cancel each other out leading to an over-determined system when the following condition is met

$$\eta_v(1 + \eta_s \eta_{vs} s_i) = \eta_u \frac{r_u \gamma_v}{r_v \gamma_u}. \quad (124)$$

This enables infinitely many solutions for the steady states of (*M3*). Hence, the exact values for  $u_i$  and  $v_i$  cannot be further identified. Inserting condition (124) in the steady state solution  $v_i \neq 0$  from (120) reveals that the combination of  $u_i$  and  $v_i$  is subject to a restriction that resembles the right hand side of the steady state (122):

$$u_i + \frac{r_u \gamma_v}{r_v \gamma_u} v_i = \frac{r_u}{\gamma_u} - \frac{1 + \eta_s s_i}{\eta_u}. \quad (125)$$

Interestingly, this restriction can also be rewritten using (124) to resemble the right hand side of (123):

$$\frac{r_v \gamma_u}{r_u \gamma_v} u_i + v_i = \frac{r_v}{\gamma_v} - \frac{1 + \eta_s s_i}{\eta_v(1 + \eta_s \eta_{vs} s_i)}. \quad (126)$$

This means that expression levels of  $u_i$  and  $v_i$  in this state must be lower or equal to the expressions in (122) and (123). Thus, this is an intermediate state that lies between the desired cell fates  $u^+v^-$  and  $u^-v^+$ . It is particularly interesting to see what role the signal  $s_i$  plays in this. Therefore,  $s_i$  is isolated in equation (124), leading to a critical signal value  $s^*$

$$s^* = \frac{r_u \gamma_v \eta_u - r_v \gamma_u \eta_v}{r_v \gamma_u \eta_v \eta_s \eta_{vs}}. \quad (127)$$

For  $s_i = s^*$ , this intermediate steady state will always occur. Hence, this critical signal value allows the description of the switching behavior in the ODE system (Fig. 15). For values  $s_i < s^*$ , a cell ends up in states (122) ( $u^+v^-$ ), whereas  $s_i > s^*$  leads to the opposite state (123) ( $u^-v^+$ ). At exactly  $s = s^*$ , both expression levels  $u$  and  $v$  approach the straight line defined by (125). Thereby, no unique steady state is attained. Altogether, the steady states (121)-(123) of ODE system (*M3*) were successfully derived and the signal was identified triggering cell differentiation in this model.

**Stability analysis** In contrast to the generalized phenomenological models, ODE system (*M3*) allows a linear stability analysis on the level of single cells as well as complete tissues for a generic description of the signaling mechanism. The goal here is to determine parameter restrictions that lead to stability/instability of the system. Specifically, parameters that allow the formation of a tissue that is inhomogeneous with respect to cell fates. For this reason, the linearization matrix is set up:

$$L^{\text{ODE}} = \begin{pmatrix} r_u A_{uu} - \gamma_u I_n & r_u A_{uv} \\ r_v A_{uv} & r_v A_{vv} - \gamma_v I_n \end{pmatrix}, \quad (128)$$

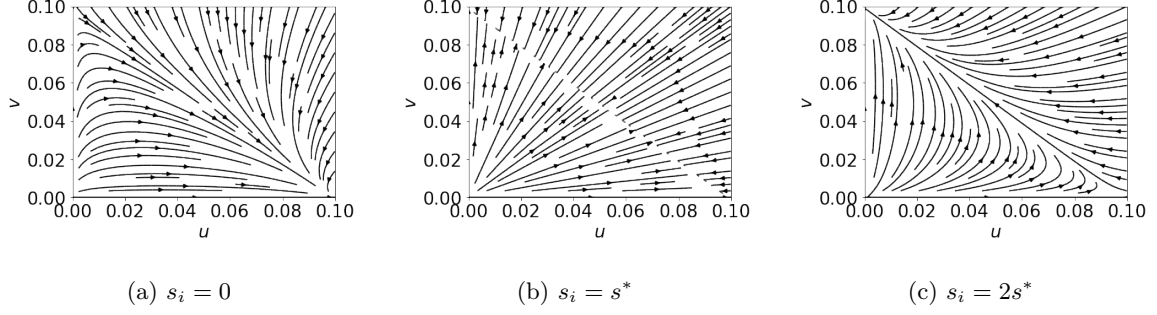


Figure 15: Streamline phase portraits of ODE system (M3) for a single cell and three different values for  $s_i$ . Arrows show the path from the initial condition towards the respective steady states (122), (125) and (123).

Here,  $I_n$  again described the  $n$ -dimensional identity matrix. The block matrices  $A_{xy}$ ,  $x, y \in \{u, v\}$  arising in  $L^{\text{ODE}}$  are defined by the total derivatives of  $p_u$  and  $p_v$  with respect to  $u_i$  and  $v_i$ . Interpreting the signal as a function of  $\mathbf{u}$  and  $\mathbf{v}$ , the chain rule leads to

$$A_{uu} = \frac{\partial p_u}{\partial u} + \frac{\partial p_u}{\partial s} \frac{\partial s}{\partial u}, \quad A_{uv} = \frac{\partial p_u}{\partial v} + \frac{\partial p_u}{\partial s} \frac{\partial s}{\partial v}, \quad (129)$$

$$A_{vu} = \frac{\partial p_v}{\partial u} + \frac{\partial p_v}{\partial s} \frac{\partial s}{\partial u}, \quad A_{vv} = \frac{\partial p_v}{\partial v} + \frac{\partial p_v}{\partial s} \frac{\partial s}{\partial v}. \quad (130)$$

Here, the derivative  $\frac{\partial p_u}{\partial u} := \left( \frac{\partial p_u}{\partial u_j}(u_i, v_i, s_i) \right)_{i,j=1,\dots,n}$  denotes a matrix whose entries are comprised of the respective combinations of partial derivatives. Other block matrices are defined analogously. The partial derivatives regarding the the binding probabilities are

$$\frac{\partial}{\partial u_j} p_u(u_i, v_i, s_i) = \begin{cases} \frac{\eta_u(1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_s s)}{(1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_u u_i+\eta_s s_i)^2}, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (131)$$

$$\frac{\partial}{\partial v_j} p_u(u_i, v_i, s_i) = \begin{cases} -\frac{\eta_v \eta_u u_i(1+\eta_s \eta_{vs} s_i)}{(1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_u u_i+\eta_s s_i)^2}, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (132)$$

$$\frac{\partial}{\partial u_j} p_v(u_i, v_i, s_i) = \begin{cases} -\frac{\eta_v \eta_u v_i(1+\eta_s \eta_{vs} s_i)}{(1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_u u_i+\eta_s s_i)^2}, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (133)$$

$$\frac{\partial}{\partial v_j} p_v(u_i, v_i, s_i) = \begin{cases} \frac{\eta_v(1+\eta_s \eta_{vs} s_i)(1+\eta_u u_i+\eta_s s_i)}{(1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_u u_i+\eta_s s_i)^2}, & \text{if } i = j \\ 0, & \text{if } i \neq j \end{cases} \quad (134)$$

$$\frac{\partial}{\partial s_i} p_u(u_i, v_i, s_i) = -\frac{\eta_u \eta_s u_i(1+\eta_v \eta_{vs} v_i)}{(1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_u u_i+\eta_s s_i)^2} \quad (135)$$

$$\frac{\partial}{\partial s_i} p_v(u_i, v_i, s_i) = \frac{\eta_v \eta_s v_i(\eta_{vs} + \eta_u \eta_{vs} u_i - 1)}{(1+\eta_v v_i(1+\eta_s \eta_{vs} s_i)+\eta_u u_i+\eta_s s_i)^2} \quad (136)$$

Stability again requires the eigenvalues of  $L^{\text{ODE}}$  to have only negative real parts. The eigenvalues are the roots of the characteristic polynomial, which is in general described by

$$\chi(\lambda) = \det(L^{\text{ODE}} - \lambda I_{2n}). \quad (137)$$

Again,  $I_{2n}$  denotes the identity matrix in  $2n$  dimensions. For the generalized setting, the eigenvalues cannot be accurately determined. However, the characteristic polynomial becomes highly simplified when it is restricted to one of the steady states (121)-(123). This allows each state to be analyzed individually.

**Steady state (121)** The first steady state (121) is characterized by the absence of any expressions. Therefore, it resembles the  $u^-v^-$  fate of a cell. The system is supposed to differentiate between two cell fates where the expressions of one protein is high and the other is low and vice-versa. Therefore, it is necessary to find ways to exclude the emergence of the  $u^-v^-$  fate for any single cell. Without loss of generality, it is assumed that  $u_{n+1} = 0 = v_{n+1}$ . Focusing on row  $n+1$  of  $L^{\text{ODE}}$ , the relevant derivatives are (132) - (136) with (131) being the only one left out. Fortunately, all derivatives yield 0 with the

exception of  $\frac{\partial}{\partial v_1} p_v(u_1, v_1, s_1)$ . This remaining entry results in

$$r_v \frac{\partial}{\partial v_1} p_v(0, 0, s_1) - \gamma_v - \lambda = r_v \eta_v \frac{1 + \eta_s \eta_{vs} s_1}{1 + \eta_s s_1} - \gamma_v - \lambda. \quad (138)$$

The Laplace expansion allows the determinant of a matrix  $A \in \mathbb{R}^{n \times n}$  to be calculated as

$$\det(A) = \sum_{j=1}^{n-1} (-1)^{i+j} a_{ij} \det(A_{ij}), \quad i = 1, \dots, n, \quad (139)$$

where  $a_{ij}$  denote the matrix entries and  $A_{ij} \in \mathbb{R}^{(n-1) \times (n-1)}$  the submatrices that arise, when row  $i$  and column  $j$  are removed from  $A$ . This expansion is valid for any row  $i$ , especially in the case of  $L^{\text{ODE}}$  for row  $n+1$ , where only a single summand is not equal to 0

$$\det(L^{\text{ODE}}) = \left( r_v \eta_v \frac{1 + \eta_s \eta_{vs} s_1}{1 + \eta_s s_1} - \gamma_v - \lambda \right) \det(L_{ij}^{\text{ODE}}), \quad i = 1, \dots, n, \quad (140)$$

The first factor in this equation is sufficient to provide a single eigenvalue by setting

$$r_v \eta_v \frac{1 + \eta_s \eta_{vs} s_1}{1 + \eta_s s_1} - \gamma_v - \lambda \stackrel{!}{=} 0. \quad (141)$$

For the eigenvalue  $\lambda$ , this can be rearranged to

$$\lambda = r_v \eta_v \frac{1 + \eta_s \eta_{vs} s_1}{1 + \eta_s s_1} - \gamma_v. \quad (142)$$

This state becomes unstable for  $\lambda > 0$ . In this case, the parameters follow the relation

$$\eta_v > \frac{\gamma_v}{r_v} \frac{1 + \eta_s s_1}{1 + \eta_s \eta_{vs} s_1}. \quad (143)$$

For a physically realistic signal,  $s_i \geq 0$ , and considering the signal to be activating, i.e.  $\eta_{vs} > 1$ , a sufficient condition for instability is

$$\eta_v > \frac{\gamma_v}{r_v}. \quad (144)$$

For the sake of readability, the notation using energy coefficients  $\eta$  is preferred over the energy differences  $-\Delta\varepsilon$ . However, when using actual values for said parameters, it is preferable to describe them as energy differences. Therefore, (144) is also translated to

$$-\Delta\varepsilon_v > \ln \left( \frac{\gamma_v}{r_v} \right). \quad (145)$$

This provides the necessary condition for instability of steady state (121) and consequently  $u^-v^-$  cell fates. With the exclusion of this steady state, the focus will shift towards steady states (122) and (123), such that only the two cell fates  $u^+v^-$  and  $u^-v^+$  are made relevant.

**Homogeneous steady state (122)** Now that only two cell fates remain, it is desirable to achieve an inhomogeneous mixture of  $u^+v^-$  and  $u^-v^+$  cells in the tissue, in order to generate organoids that consist of both, PrE and Epi precursor cells. Hence, two additional stability analyses are performed to identify parameter regions in which homogeneous states become unstable. If the system is unstable in both homogeneous steady state (122) and (123), then the system will inevitably assume an inhomogeneous state, i.e. a mixture of two different cell types. First, steady state (122) is assumed for every cell, i.e.  $u_i = \frac{r_u}{\gamma_u} - \frac{1 + \eta_s s_i}{\eta_u}$  and  $v_i = 0$  for all  $i$ . Inserting these expressions into the derivatives (131)-(136) results in a simplification of  $L^{\text{ODE}}$ . Expressions (133) and (136) yield 0 for every  $i, j$ . Hence, the off-diagonal block matrix  $A_{vu} = \mathbf{0}$ . The full determinant can therefore be determined by the product of the block matrix determinants on the diagonal of  $L^{\text{ODE}}$ . It suffices to focus on the lower right block matrix, which contains the matrix  $A_{vv}$ . Since (136) amounts to 0,  $A_{vv}$  becomes diagonal with diagonal entries

$$(A_{vv})_i = \frac{\eta_v(1 + \eta_s \eta_{vs} s_i)}{1 + \eta_u u_i + \eta_s s_i}, \quad i = 1, \dots, n \quad (146)$$

Inserting  $u_i$  yields

$$(A_{vv})_i = \frac{\gamma_u \eta_v}{r_u \eta_u} (1 + \eta_s \eta_{vs} s_i), \quad i = 1, \dots, n. \quad (147)$$

Making use of this allows to determine  $n$  factors of the characteristic polynomial

$$\chi(\lambda) = \det(r_u A_{uu} - (\gamma_u + \lambda)I_n) \det(r_v A_{vv} - (\gamma_v + \lambda)I_n) \quad (148)$$

$$= \det(r_u A_{uu} - (\gamma_u + \lambda)I_n) \left[ \prod_{i=1}^n \gamma_u \frac{r_v \eta_v}{r_u \eta_u} (1 + \eta_s \eta_{vs} s_i) - \gamma_v - \lambda \right] \quad (149)$$

Hence,  $n$  eigenvalues are given by the second factor in (149). If a single one of these is greater than zero, the inhomogeneous state is unstable. This means it suffices to show

$$\gamma_u \frac{r_v \eta_v}{r_u \eta_u} (1 + \eta_s \eta_{vs} s_i) > \gamma_v \quad (150)$$

for one  $i = 1, \dots, n$ . Instead of any  $i = 1, \dots, n$ , a necessary condition is given by the maximum of all signal values. After rearranging the equation, a parameter restriction for  $\eta_u$  is found

$$\eta_u < \eta_v \frac{r_v \gamma_u}{r_u \gamma_v} (1 + \eta_s \eta_{vs} \max_i s_i). \quad (151)$$

At this point, restriction (151) describes the case of a generalized system which cannot be further simplified. Depending on the exact type of cell-cell communication, one might find a more precise expression. A different formulation is again obtained by looking at the energy differences, which leads to

$$-\Delta \varepsilon_u < -\Delta \varepsilon_v + \ln \left( 1 + e^{-\Delta \varepsilon_s - \Delta \varepsilon_{vs}} \max_i s_i \right) + \ln \left( \frac{r_v \gamma_v}{r_u \gamma_u} \right). \quad (152)$$

As this result only relies on  $n$  of a total of  $2n$  eigenvalues, it remains possible that there is an even tighter restriction on the parameters. However, throughout this whole study, no evidence emerged suggesting that this is the case.

**Homogeneous steady state (123)** Analogous to the previous analysis, the analysis of homogeneous steady state (123) starts by setting  $u_i = 0$  and  $v_i = \frac{r_v}{\gamma_v} - \frac{1 + \eta_s s_i}{\eta_v (1 + \eta_s \eta_{vs} s_i)}$ . This time, (132) and (135) become zero for all  $i, j$  and thus  $A_{uv} = \mathbf{0}$ . The block matrix  $A_{uu}$  is diagonal with diagonal entries

$$(A_{uu})_i = \frac{\eta_u}{1 + \eta_v v_i (1 + \eta_s \eta_{vs} s_i) + \eta_s s_i}, \quad i = 1, \dots, n. \quad (153)$$

Inserting  $v_i$  yields

$$(A_{uu})_i = \frac{\gamma_v \eta_u}{r_v \eta_v} \frac{1}{1 + \eta_s \eta_{vs} s_i}, \quad i = 1, \dots, n. \quad (154)$$

Again, only  $n$  factors of the characteristic polynomial are determined

$$\chi(\lambda) = \det(r_u A_{uu} - (\gamma_u + \lambda)I_n) \det(r_v A_{vv} - (\gamma_v + \lambda)I_n) \quad (155)$$

$$= \left[ \prod_{i=1}^n \gamma_v \frac{r_u \eta_u}{r_v \eta_v} \frac{1}{1 + \eta_s \eta_{vs} s_i} - \gamma_u - \lambda \right] \det(r_v A_{vv} - (\gamma_v + \lambda)I_n). \quad (156)$$

Exploiting the instability condition that one eigenvalue must be positive yields again to a restriction which must be fulfilled for a single  $i = 1, \dots, n$

$$\eta_u > \frac{r_v \gamma_u}{r_u \gamma_v} \eta_v (1 + \eta_s \eta_{vs} s_i). \quad (157)$$

This time this transforms into a necessary condition if the values of  $s_i$  assumes its minimum

$$\eta_u > \frac{r_v \gamma_u}{r_u \gamma_v} \eta_v (1 + \eta_s \eta_{vs} \min_i s_i). \quad (158)$$

In terms of energy differences, this reads

$$-\Delta \varepsilon_u > -\Delta \varepsilon_v + \ln \left( 1 + e^{-\Delta \varepsilon_s - \Delta \varepsilon_{vs}} \min_i s_i \right) + \ln \left( \frac{r_v \gamma_u}{r_u \gamma_v} \right). \quad (159)$$

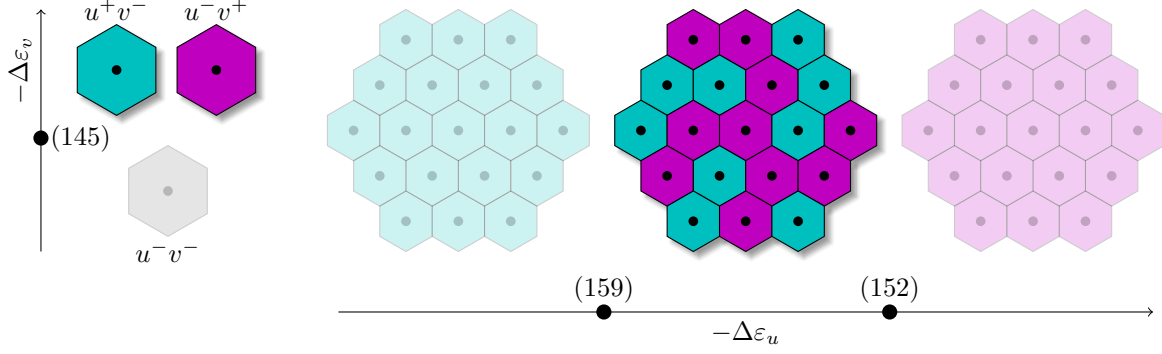


Figure 16: Illustration of the different steady states at the single cell level (left) and the tissue level (right). The states we are aiming for are highlighted with higher opacity. Nodes and their corresponding number on the axes reference the relevant equation for the transition from one state to another.

**Inhomogeneous steady state** The previous analyses revealed two conditions for the instability of homogeneous steady states. Consequently, condition (152) and (159) combined define a stability interval for the inhomogeneous steady states depending on  $-\Delta\epsilon_u$ . This interval is given by

$$\Delta\epsilon_{min} < -\Delta\epsilon_u < \Delta\epsilon_{max} \quad (160)$$

with

$$\Delta\epsilon_{min} := -\Delta\epsilon_v + \ln \left( 1 + e^{-\Delta\epsilon_s - \Delta\epsilon_{vs}} \min_i s_i \right) + \ln \left( \frac{r_v \gamma_u}{r_u \gamma_v} \right) \quad (161)$$

$$\Delta\epsilon_{max} := -\Delta\epsilon_v + \ln \left( 1 + e^{-\Delta\epsilon_s - \Delta\epsilon_{vs}} \max_i s_i \right) + \ln \left( \frac{r_v \gamma_u}{r_u \gamma_v} \right) \quad (162)$$

The reproduction rates  $r_u, r_v$  and decay rates  $\gamma_u, \gamma_v$  shift this interval by  $\ln \left( \frac{r_u \gamma_v}{r_v \gamma_u} \right)$ . The length of the interval is determined by the minimum and maximum signal values combined with the associated energy differences  $-\Delta\epsilon_s$  and  $-\Delta\epsilon_{vs}$ . The complete stability analysis can be summarized by figure 16. At the single cell level, the formation of  $u^-v^-$  cells was successfully excluded by inequality (145). Therefore, a distinction of three different steady states takes place at the tissue level. Two states where all cells are either  $u^+v^-$  or  $u^-v^+$  and the state in between, where an inhomogeneous mixture of the two cell types is found.

### 3.1.5 Transcriptional regulation summary

The question if a mathematical model is able to describe cell fate decisions is difficult to answer and far surpasses the scope of this thesis. However, when the cell fates are associated with the steady states of a dynamical system, it becomes possible to construct fitting ODEs that are able to generate states that resemble certain properties of given cell fates to some extent, such as protein expression levels. The idea behind the transcriptional regulation models presented in this section, is to provide models that lead to the emergence of two different types of cells, one with high expression in  $u$  and low expression in  $v$  ( $u^+v^-$ ) and vice-versa ( $u^-v^+$ ).

For a model to generate differences in expression levels, it must be able to adopt heterogeneous states, which is made possible by the instability of homogeneous states. In a first step, the well known instability condition provided in [26] was first extended by the addition of mutual inhibition. Thus a condition for the derivatives of the inhibitory and activatory functions in the ODE system guarantees the emergence of heterogeneous states. Furthermore it was shown, that the instability condition for two cells is sufficient to prove instability of the homogeneous state in a ring of  $n$  cells. In the second step, auto-activation as well as an adjacency based inhibition were included in order to reflect the GRN 10 that was proposed to describe the transcriptional regulation of NANOG and GATA6. It was possible to derive two conditions that lead to the instability of the homogeneous state. Thus, heterogeneity of a two cell system can be guaranteed. Extensions to higher dimensions and different types of cell-cell communication were considered to exceed the goal of this thesis, yet offer opportunities for further research.

Phenomenological models based on the Hill equation are already widely used to describe the temporal evolution of markers for cell differentiation [27, 28, 30, 31, 26]. Although the Hill equation was

never explicitly formulated for cell differentiation [36], its use in describing cooperative binding events is surprisingly similar to the thermodynamic approach considered in this study. However, it lacks some of the detail by basically assuming only a state where  $k$  constituents are bound simultaneously. In the case of oxygen-hemoglobin binding, improved approximations for the dissociation curves were established through the Adair equation [59] and later on the Monod-Wyman-Changeux (MWC) model [60]. These models differ in that they also consider states in which fewer than four of the four total oxygen binding sites of hemoglobin may be occupied. A further generalization of these models was presented by the thermodynamic models in [39] on which the thermodynamic models in this work are also based.

Another drawback of the phenomenological approach is that the combination of several activatory and inhibitory Hill functions via addition or multiplication is not trivial. The decision in [27] and thus in (M1) to introduce the mutual inhibition as multiplicative terms was described as "qualitative reasoning". It aims to prevent high expressions of one transcription factor in the presence of the other. In doing so, this decision was resolved on a phenomenological level rather than emerging from an intrinsic mechanism of the GRN. In the scope of this study, the phenomenological model (M1) serves only as a reference system that will be compared to the others in a simple computational example.

Contrary to phenomenological models, mechanistic models aim to describe a phenomenon using physical, chemical or biological principles. The thermodynamic model introduced in this study was motivated by [37, 38, 39]. It uses well known methods from statistical thermodynamics to describe the probability of RNAP binding to a desired promoter. Although more complicated in its derivation than phenomenological models, it goes beyond their capabilities by incorporating different orders of interaction. It requires energy differences that describe how advantageous a state is compared to the base state by being greater or less than one. This approach spawns multiple different possibilities for GRN 10. In this study, a model incorporating only first order and mostly independent interactions was derived and further investigated. An analysis of several model parameters revealed that the model is capable of describing cells in either two ( $u^+v^-$  and  $u^-v^+$ ) or three (additionally  $u^-v^-$  or  $u^+v^+$ ) different states based on the received signal. Furthermore, it provides its own definition of the basal activity of a given promoter. In doing so, a mathematical definition of activations and inhibitions arises that is characterized through the deviation from the basal expression values. An experimental verification of this system would require a gene regulatory system with exactly one promoter, activator and inhibitor, in which the cells could be deprived of the interactors. It is beyond the aim of this thesis to judge whether this would actually be realizable. However, the insight gained by providing an exact definition of different cell states ( $u^+v^+$ ,  $u^+v^-$ ,  $u^-v^+$ ,  $u^-v^-$ ) is invaluable and will hopefully pave the way for numerous subsequent studies.

Under the assumption that auto-activation and transcription can be used synonymously with transcription, a different ODE system (M3) could be derived. Although the assumption already invalidates the mechanistic nature of the model, the remaining derivation is based on the mechanistic description of transcription factor bindings. It might therefore be best described as a mix between a phenomenological and mechanistic model. Surprisingly, it was shown that this model can be described in much greater detail at the analytical level. The steady states of the ODE system were determined explicitly leading to three relevant states  $u^-v^-$ ,  $u^+v^-$  and  $u^-v^+$ . In a stability analysis, a parameter restriction was identified that allowed the exclusion of the  $u^-v^-$  fate, leaving only a binary choice for the cell fate. Furthermore, a stability interval was derived enabling the formation of an inhomogeneous mixture of cell types. On a practical level, the model sets itself far apart from the others, since the parameters for inhomogeneous states can easily be determined in advance. Moreover, the transition from the homogeneous  $u^+v^-$  state to the homogeneous  $u^-v^+$  state suggests that there is a monotonic increase in  $u^-v^+$  cells in between. This confers unexpectedly strong control over the cell type proportion in simulations of cell fate decisions at the tissue level.

### 3.2 Cell-cell communication

Cell-cell communication is made possible by the secretion and absorption of chemical signals. How exactly these signals disperse throughout a tissue or organoid depends on a wide variety of factors. Some of these might be physical properties, like diffusivity or the viscosity of interstitial fluid. Others might directly be related to the production, uptake and decay of the signaling molecules themselves. In the mouse embryo, the FGFr/Erk signaling pathway describes one possible cell-cell communication mechanism involved in the expressions of NANOG and GATA6. Thereby, FGF4 binds to the receptor FGFr2, leading to an activation of GATA6, while simultaneously inhibiting NANOG [15, 16]. The synthesis of Fgf4 is activated by NANOG [17]. In this section, a set of different signals will be introduced as simple mathematical functions of  $\mathbf{u} = (u_1, \dots, u_n)^T$ .



### 3.2.1 Averaged nearest neighbor signaling

Slow diffusion, rapid decay and fast signal uptake are three reasons that can cause a paracrine signal to reach only adjacent cells. When this happens, the absorbed signal  $s_i$  only depends on the adjacent cells  $N_G(i)$ . Assuming additionally that the signal is distributed evenly among the neighboring cells, the absorbed signal of cell  $i$  can be written as

$$s_i = \frac{1}{|N_G(i)|} \sum_{j \in N_G(i)} u_j. \quad (163)$$

In this averaged nearest neighbor (NN) signal, a linear dependence of the secreted signal on  $u_j$  was assumed as the most simple way to describe the signal. This way, the signal  $\mathbf{s}$  is given by a linear function  $\mathbf{s}(\mathbf{u}) = \mathbf{A}\mathbf{u}$  with the adjacency matrix

$$A = (A_{i,j})_{i,j=1,\dots,M}, \quad A_{i,j} = \begin{cases} \frac{1}{|N_G(i)|} & \text{if } j \in N_G(i) \\ 0 & \text{if } j \notin N_G(i) \end{cases}. \quad (164)$$

### 3.2.2 Distance-based global signaling

One of the main parts of this thesis is the investigation of a possible long range cell-cell communication mechanism and its effects on the pattern formation in tissues. This communication would be made possible by a paracrine signal that reaches cells beyond the nearest neighbors. Considering a larger diffusivity, longer lifetime or slow uptake of the signal, it is reasonable to describe the absorbed signal  $s_i$  as the collective effect of all cells throughout the organoid. As the signal disperses in all directions, its intensity decreases with the distance traveled. Hence, the signal will be defined as the weighted sum of secreted signals over the remaining cells

$$s_i = \left( \sum_{j \neq i} u_j q^{d_{ij}-1} \right) / \left( \max_k \sum_{j \neq k} q^{d_{kj}-1} \right), \quad q \in [0, 1]. \quad (165)$$

The distance  $d_{ij}$  is the previously defined graph distance. At first glance, the Euclidean distance seems to be the better choice to describe the radial propagation of a signal. However, the graph distance allows to imitate the absorption of the signal by each cell it passes by. The normalization in the denominator of (165) ensures that the signals will be equally strong, independent of a cells location or neighbors. The weights  $q^{d_{ij}-1}$  were chosen purely as a mathematical construct to define the fractions of the signal that gets transported from cell to cell. For  $q = 0.1$ , the second nearest neighbors of a cell receive only 10% of the total signal of the direct neighbors. In comparison,  $q = 0.5$  leads to a halving of the signal every time it reaches cells one graph distance further away (Fig. 17). Thus, the dispersion parameter  $q$  can be used to describe a transition from a direct neighbor signal to an equally dispersed signal. This can be demonstrated by setting  $q = 0$ . The weights then become

$$q^{d_{ij}-1} = 0^{d_{ij}-1} = \begin{cases} 1, & \text{for } d_{ij} = 1 \\ 0, & \text{for } d_{ij} > 1, \end{cases}, \quad (166)$$

i.e. the weights for all cells  $j$  that are not adjacent to cell  $i$  are 0. Hence, the signal becomes a NN signal similar to (163) without averaging over the number of neighbors. At the other end,  $q = 1$  yields

$$q^{d_{ij}-1} = 1^{d_{ij}-1} = 1. \quad (167)$$

In this case, every cell has the same impact on other cells, independent of the distance between them. In conclusion, the signal parameter  $q$ , which resembles a measure for the signal dispersion, can be used to describe a continuous transition from an NN signal at  $q = 0$ , through a distance-based global signal for  $q \in (0, 1)$  to an evenly distributed signal at  $q = 1$ . As for (163), the signal is again a linear function  $\mathbf{A}\mathbf{u}$  with the dispersion weight matrix

$$A = (A_{i,j})_{i,j=1,\dots,M}, \quad A_{i,j} = \begin{cases} a q^{d_{ij}-1} & \text{if } i \neq j \\ 0 & \text{if } i = j \end{cases}, \quad (168)$$

and the normalization factor

$$a = \left( \max_k \sum_{l \neq k} q^{d_{kl}-1} \right)^{-1}. \quad (169)$$

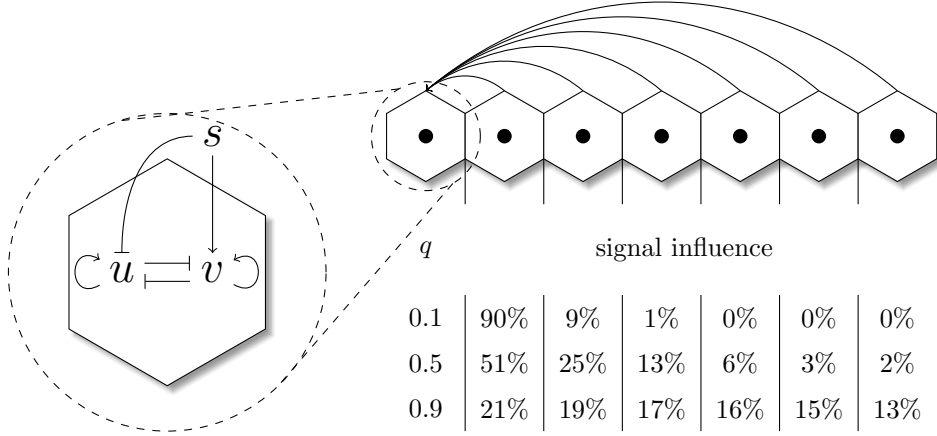


Figure 17: Illustration of the effects of the signal from cells of different distances for an exemplary one-dimensional cell line. The cell on the left resembles the previously established GRN, where  $u$  and  $v$  mutually inhibit each other. In addition to that,  $s$  activates  $v$  while inhibiting  $u$ . Curved arrows indicate that the first cell in the cell line receives a signal from the remaining six cells. The table highlights the respective contributions of each cell to the received signal for different dispersions  $q \in \{0.1, 0.5, 0.9\}$ . Percentages are rounded to the nearest integer.

### 3.2.3 Estimation of the stability interval of model (M3)

For the two concrete examples of signaling, the stability interval (160) can be estimated by specifying the lower and upper bounds for the expression values  $u_i$ . Given the steady states (122) and (123), this yields

$$0 \leq u_i \leq \frac{r_u}{\gamma_u} - \frac{1 + \eta_s s_i}{\eta_u} < \frac{r_u}{\gamma_u} \quad (170)$$

Using this inequality to replace every  $u_j$  with either 0 or  $\frac{r_u}{\gamma_u}$  in the NN signal (163), leads to

$$s_i \geq \frac{1}{N_G(i)} \sum_{j \in N_G(i)} 0 = 0, \quad (171)$$

$$s_i < \frac{1}{N_G(i)} \sum_{j \in N_G(i)} \frac{r_u}{\gamma_u} = \frac{r_u}{\gamma_u}. \quad (172)$$

Likewise, for the distance-based signal (165), the same estimate is found

$$s_i \geq \left( \sum_{j \neq i} 0 q^{d_{ij}-1} \right) / \left( \max_k \sum_{j \neq k} q^{d_{kj}-1} \right) = 0, \quad (173)$$

$$s_i < \left( \sum_{j \neq i} \frac{r_u}{\gamma_u} q^{d_{ij}-1} \right) / \left( \max_k \sum_{j \neq k} q^{d_{kj}-1} \right) < \frac{r_u}{\gamma_u}. \quad (174)$$

In both cases, this means that  $s_i \in [0, r_u/\gamma_u)$ . Using the bounds of this interval in the stability interval (160), leads to a simplified parameter restriction for  $-\Delta\varepsilon_u$

$$-\Delta\varepsilon_v + \ln \left( \frac{r_v \gamma_u}{r_u \gamma_v} \right) < -\Delta\varepsilon_u < -\Delta\varepsilon_v + \ln \left( 1 + e^{-\Delta\varepsilon_s - \Delta\varepsilon_{vs}} \frac{r_u}{\gamma_u} \right) + \ln \left( \frac{r_v \gamma_u}{r_u \gamma_v} \right). \quad (175)$$

For specific parameter combinations, (175) becomes an almost accurate representation of the true stability interval. Such a case is given, if the steady state values of  $u_i$  are close to its upper bound, i.e.

$$u_i = \frac{r_u}{\gamma_u} - \frac{1 + \eta_s s_i}{\eta_u} \approx \frac{r_u}{\gamma_u}. \quad (176)$$

Hence, given parameter combinations, such that

$$\frac{r_u}{\gamma_u} \gg \frac{1 + \eta_s \frac{r_u}{\gamma_u}}{\eta_u} \quad (177)$$

is fulfilled, the approximated stability interval (175) can be used to accurately determine the remaining parameters such that a heterogeneous mixture of cells is guaranteed.

### 3.3 Organoid growth

Having already introduced transcriptional regulation, as well as cell-cell communication, the final step is to provide the cells on which to run the transcriptional regulation models. The goal of this section is to introduce several methods that in combination form an organoid growth model. This is roughly divided into the following aspects:

- Growth of single cells
- Division of cells and organoid growth
- Adhesion and repulsion of cells in contact

The section concludes by an additional analysis of the models to identify key parameters needed for the simulations.

#### 3.3.1 Cell growth

One part of the growth of an organoid is the growth of individual cells. Interestingly, the best fitting mathematical description for cell growth is still controversial [61]. There are hints of linear [62] as well as exponential [63] growth in cells. In addition, there are indications of a cell size control, limiting the maximum cell size [64]. More recently, the growth rates of cells have been shown to vary between different stages of the cell cycle [65]. These include phases of constant, but also decreasing as well as increasing growth rates. Because of the complexity of cell growth in general, only a simplified case is considered here. By absorbing nutrients from the surrounding medium, cells are able to increase in size. Thereby, the larger a cell becomes, the more nutrients it will be able to absorb, suggesting an exponential growth of the cell. However, the acquired nutrients are also necessary to sustain the metabolism of the cell at its current size. Under this assumption, there must be a limitation to the growth of the cells. One prominent model that captures this behavior is the logistic growth model, which is here applied to the cell's radius, i.e.

$$\frac{dr}{dt} = \lambda r(r^* - r) \quad (178)$$

with constant growth rate  $\lambda$  and maximum radius  $r^*$ . For low values of  $r$  (178) approximately describes exponential growth. Likewise, for values of  $r$  close to  $r^*$  it closely resembles a bounded growth, i.e.  $\frac{dr}{dt} = r^+ - r$ . The analytical solution to equation (178) is

$$r(t) = \frac{r^*}{1 + \left(\frac{r^* - r_0}{r_0}\right) e^{-\lambda r^* (t - t_0)}}, \quad (179)$$

with initial radius and time  $r_0$  and  $t_0$ . Depending on the ratio of initial radius to maximum radius, the cell initially enters an exponential growth or starts directly in the bounded growth state (Fig. 18).

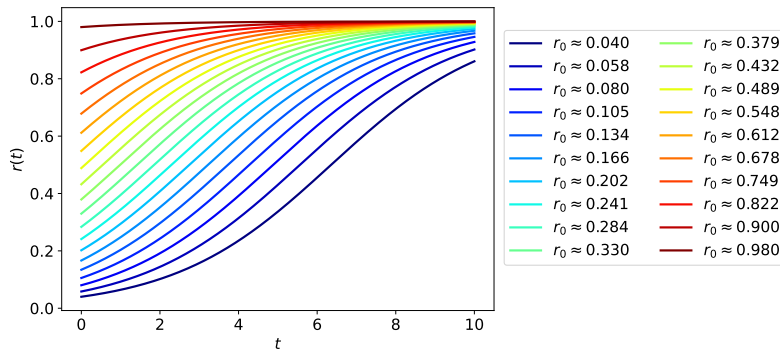


Figure 18: Visualization of the time evolution of a cell's radius given different initial conditions  $r_0$ . The spacing between the values of  $r_0$  increases quadratically. The growth rate chosen in this example was  $\lambda = 0.5$  and the maximum radius  $r^* = 1$ .

### 3.3.2 Radius dependent cell division

Cell division is usually described as a function of either the elapsed time [66] or the size of a cell [43]. In the latter, cell division was already used for the in silico generation of mouse ICM organoids. The division itself was described as a stochastic process of a cell's radius. Likewise, a similar strategy is chosen in this thesis, by defining a cell division probability based on a cell's radius. In order to realize this, a cumulative distribution function (CDF)  $P(R \leq r) = F_R(r)$  is needed. For a random variable  $R$ , the CDF describes the probability that a cell will have divided upon reaching a radius  $r$ . We assume that the cell division radii will be normally distributed. However, the maximum radius  $r^*$  already prohibits the model to use the right tail of the distribution. Therefore, the CDF chosen in this model is from a truncated normal distribution on an interval  $[r_{\min}, r_{\max}]$  with  $r_{\max} \leq r^*$ , i.e.

$$F_R(r) = \frac{\text{erf}(\rho) - \text{erf}(\rho_{\min})}{\text{erf}(\rho_{\max}) - \text{erf}(\rho_{\min})}, \quad r \in [r_{\min}, r_{\max}], \quad (180)$$

$$\rho = \frac{r - \mu_{\text{div}}}{\sqrt{2}\sigma_{\text{div}}}, \quad \rho_{\min} = \frac{r_{\min} - \mu_{\text{div}}}{\sqrt{2}\sigma_{\text{div}}}, \quad \rho_{\max} = \frac{r_{\max} - \mu_{\text{div}}}{\sqrt{2}\sigma_{\text{div}}}.$$

Values  $\mu_{\text{div}}$  and  $\sigma_{\text{div}}$  are the respective mean and standard deviation of this distribution if  $\mu_{\text{div}} = \frac{1}{2}(r_{\min} + r_{\max})$ , allowing us to preserve the symmetry of the distribution despite truncation. The function erf is the error function defined as

$$\text{erf}(x) = \frac{2}{\pi} \int_0^x e^{-x^2} dx. \quad (181)$$

Outside of  $[r_{\min}, r_{\max}]$  we define

$$F_R(r) = 0, \quad r < r_{\min}, \quad (182)$$

$$F_R(r) = 1, \quad r > r_{\max}. \quad (183)$$

This way, there will be no chance of cells dividing up to a radius  $r_{\min}$ , whereas cells with a radius above  $r_{\max}$  always divide (Fig. 19).

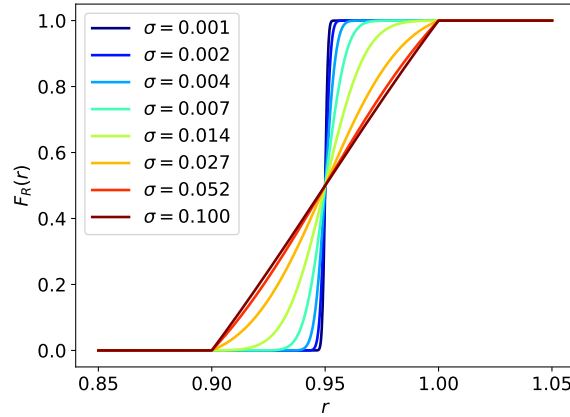


Figure 19: Cumulative distribution function for a truncated normal distribution on the interval  $[0.9, 1]$ . The mean lies in the center of the interval  $\mu_{\text{div}} = 0.95$ .

**Radius after cell division** For the cell division, cells are assumed to have homogeneous density. Conservation of mass states that the total volume of the two daughter cells with radii  $r_1$  and  $r_2$  must equal the volume of the mother cell with radius  $r$

$$\frac{4}{3}\pi r^3 = \frac{4}{3}\pi r_1^3 + \frac{4}{3}\pi r_2^3. \quad (184)$$

For a symmetric cell division, i.e.  $r_1 = r_2 =: r_0$ , this results in the cell radius after division

$$r_0 = \frac{r}{2^{\frac{1}{\text{dim}}}}. \quad (185)$$

This definition of the cell division radius applies for 3D. Since a 2D model is also considered in this study, a generalized expression will be used, such that

$$r_0 = \kappa r, \quad \text{with } \kappa = 2^{-\frac{1}{\text{dim}}}. \quad (186)$$

Here, dim describes the dimension of the system, i.e. in this study either 2 or 3.

**Time discrete cell division** When combining cell growth and division probability,  $F_R(r)$  implicitly provides the probability that a cell has divided up to some point in time. The elapsed time is exactly the time needed for a cell to grow to a radius  $r$ . For the logistic growth of the cell radii (178), an analytical solution is available, thus posing no problem in this regard. However, the upcoming cell motion can only be solved numerically, so the concept is already taken up at this point. This means that parts of the full organoid growth model will have to be solved numerically at discrete points in time. Hence, instead of the CDF, time discrete probabilities are needed. To this end, we take a look at the probabilities  $1 - F_R(r^{(i)})$  that a cell does not divide up to time step  $i$  with radius  $(r^{(i)})$ . This equals the product of all probabilities  $p_i$ , that cells do not divide in that exact time step.

$$\begin{aligned}
i = 1 : \quad & q^{(1)} = 1 - F_R(r^{(1)}) \\
i = 2 : \quad & q^{(1)}q^{(2)} = 1 - F_R(r^{(2)}) \implies q_2 = \frac{1 - F_R(r^{(2)})}{q^{(1)}} = \frac{1 - F(r^{(2)})}{1 - F_R(r^{(1)})} \\
i = 3 : \quad & q^{(1)}q^{(2)}q^{(3)} = 1 - F_R(r^{(3)}) \implies q^{(3)} = \frac{1 - F_R(r^{(3)})}{q^{(1)}q^{(2)}} = \frac{1 - F_R(r^{(3)})}{1 - F_R(r^{(2)})} \\
& \vdots \\
i = N : \quad & \prod_{i=1}^N q^{(i)} = 1 - F_R(r^{(N)}) \implies q_N = \frac{1 - F_R(r^{(N)})}{\prod_{i=1}^{N-1} q^{(i)}} = \frac{1 - F_R(r^{(N)})}{1 - F_R(r^{(N-1)})}
\end{aligned}$$

The probability  $p^{(N)}$  of dividing in time step  $N$  is therefore given by:

$$p^{(N)} = 1 - q^{(N)} = 1 - \frac{1 - F_R(r^{(N)})}{1 - F_R(r^{(N-1)})} = \frac{F_R(r^{(N)}) - F_R(r^{(N-1)})}{1 - F_R(r^{(N-1)})}. \quad (187)$$

**Estimation of organoid growth rate** The ICM organoids developed in [23] were seeded with 200 cells. After 24 h or 48 h, the number of cells was measured, leading to the respective means and standard deviations

$$n_0 \pm \sigma_0 = 200 \pm 0, \quad (188)$$

$$n_{24} \pm \sigma_{24} = 441.74 \pm 148.98, \quad (189)$$

$$n_{48} \pm \sigma_{48} = 1041 \pm 306.82. \quad (190)$$

On average, this corresponds to slightly more than one cell division in 24 hours. In a parameter study of 500 simulations with randomly selected growth rates  $k \in [0.05, 0.1]$ , an approximate growth rate of  $k = 0.083$  was identified to fit to an average of 441.74 cells after 24 hours (Fig. 20 (a)). The initial cell radii were chosen to be uniformly distributed between the lowest possible radius after cell division  $\alpha r_{\min}$  and the largest possible radius before division  $r_{\max}$ . A linear model was used to fit a curve to the data. The organoid growth follows an exponential growth

$$n(t) = e^{\Lambda t} n(0) \quad (191)$$

An exact relation between cell growth rate  $\lambda$  and organoid growth rate  $\Lambda$  has yet to be found. Computationally, this cell growth rate can be fit to the exponential model (191). Since the radius reduction during cell division differs for 2D and 3D, different growth rates  $\Lambda_{2D} = 0.0245$  and  $\Lambda_{3D} = 0.0296$  are found in the respective simulations (Fig. 20 (b)). The idea behind the 2D organoids is not to simulate the cross-section of a 3D organoid, but simply to create a model system, which can be visualized vividly. As such, the 2D growth rate only serves to create 2D organoids that roughly have a similar diameter to the 3D organoids. Given the organoid growth rates, the respective number of cell divisions after 24 hours are

$$\frac{24\Lambda_{2D}}{\ln(2)} = 0.85, \quad \frac{24\Lambda_{3D}}{\ln(2)} = 1.03. \quad (192)$$

### 3.3.3 Cell motion

The growth of a cell inevitably causes the displacement of itself and all adjacent cells. In this section, the equations of motion that describe this displacement will be derived step by step using the Lagrangian approach. The Lagrangian for a single cell is given as the difference of the kinetic and potential energy

$$L = \frac{1}{2}m \left( \frac{d\mathbf{x}}{dt} \right)^2 - V(\mathbf{x}), \quad (193)$$

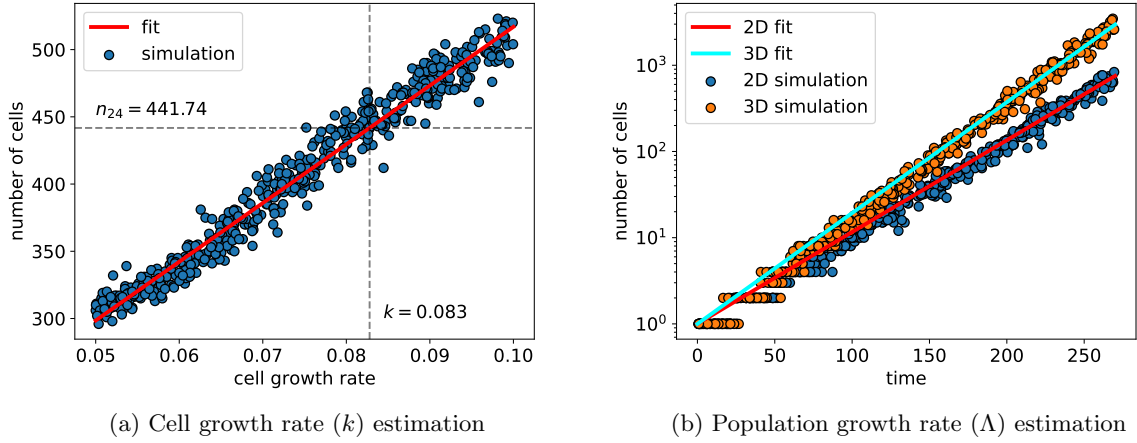


Figure 20: Simulations of 500 organoids with varying cell growth rates  $\lambda \in [0.05, 0.1]$  and fixed evolution time  $t = 24$  (a). The red line shows the fit of a linear model to the data. Dashed black lines indicate the average cell number of the experimental data after the first 24 hours and the estimated cell growth rate based on that value. Additional 500 organoids for each 2D and 3D respectively (b). Evolution times vary with  $t \in [0, 270]$ . Cell growth rate is fixed to  $\lambda = 0.083$ . Colored lines show the different fits to the exponential model.

where  $m$  describes the mass of the cell,  $\mathbf{x}$  the location of its centroid in two or three dimensions and  $V$  a generic potential for the time being. Cells moving in a fluid are subject to friction. Therefore, a linear damping force is assumed, which is described by the Rayleigh dissipation function

$$G = -\frac{b}{2} \left( \frac{d\mathbf{x}}{dt} \right)^2. \quad (194)$$

The viscous damping coefficient  $b$  describes the magnitude of the friction. The Lagrange equation

$$\frac{d}{dt} \frac{\partial L}{\partial \dot{\mathbf{x}}} - \frac{\partial L}{\partial \mathbf{x}} - \frac{\partial G}{\partial \dot{\mathbf{x}}} = 0 \quad (195)$$

describes the equations of motion. For simplicity, the shorthand notation  $\dot{\mathbf{x}} = \frac{d\mathbf{x}}{dt}$  is used. With (193) and (194), this yields

$$m \frac{d^2 \mathbf{x}}{dt^2} + \frac{dm}{dt} \frac{d\mathbf{x}}{dt} + b \frac{d\mathbf{x}}{dt} - \nabla V(\mathbf{x}) = 0. \quad (196)$$

The unusual part about (196) is the term including  $\frac{dm}{dt}$ . In most physics textbook examples, the mass of an object does not change over time. However, due to (178), a cell's radius and therefore its size and mass change over time. For the purpose of this study however, this will not be of concern as we only consider the overdamped approximation of (196). This can be justified by looking at the nondimensionalization of (196). For this, variables  $x$ ,  $t$  and  $m$  are replaced by dimensionless variables times some reference value

$$t \rightarrow T \cdot t, \quad x \rightarrow 2r^* \cdot x, \quad m \rightarrow M \cdot m. \quad (197)$$

In doing so, the relevant time and space derivatives also change according to

$$\frac{d}{dt} \rightarrow \frac{1}{T} \frac{d}{dt}, \quad \frac{d^2}{dt^2} \rightarrow \frac{1}{T^2} \frac{d^2}{dt^2}, \quad \nabla \rightarrow \frac{1}{2r^*} \nabla. \quad (198)$$

The nondimensionalized equation yields

$$\frac{2Mr^*}{T^2} m \frac{d^2 \mathbf{x}}{dt^2} + \frac{2Mr^*}{T^2} \frac{dm}{dt} \frac{d\mathbf{x}}{dt} + \frac{2br^*}{T} \frac{d\mathbf{x}}{dt} - \frac{1}{2r^*} \nabla V(2r^* \mathbf{x}) = 0. \quad (199)$$

The reference time is  $T = 3600$  s. The reference length was chosen as twice the maximum cell radius  $2r^*$ . The radius in turn was chosen to be half the maximum cutoff distance for the Delaunay triangulation used in [23], i.e.  $r^* = 1.5 \cdot 10^{-5}$  m. Since cells are primarily composed of water, 60-80% [67], their densities are typically given relative to water. Based on the abundance of proteins and water, the typical cell densities

are found between 1 and 1.3 times the density of water. The reference density used for the ICM cells will therefore be given by the density of water. For a spherical cell this leads to the single cell mass of

$$M = \frac{4}{3}\pi(r^*)^3 \cdot 10^3 \text{ kg} \cdot \text{m}^{-3} \approx 1.41 \cdot 10^{-11} \text{ kg}. \quad (200)$$

The coefficient for the inertial and the mass variation terms then yields

$$\frac{2Mr^*}{T^2} \approx 3.27 \cdot 10^{-23} \text{ N}. \quad (201)$$

The viscous damping coefficient according to Stoke's law is given by  $b = 6\pi\mu r^*$ , where  $\mu$  is the dynamic viscosity of the fluid in which the cell is located. A simple example is again just water, with  $\mu = 8.9 \cdot 10^{-4} \text{ Pa} \cdot \text{s}$ . The coefficient of the damping force then becomes

$$\frac{2br^*}{T} \approx 2.10 \cdot 10^{-15} \text{ N}. \quad (202)$$

With eight orders of magnitude separating the damping force from the inertia, the system is largely overdamped, which allows the inertia to be completely neglected. This overdamped equation of motion can be written as

$$\frac{d\mathbf{x}}{dt} = \frac{T}{2br^*} \frac{E_0}{r^*} \nabla V(\mathbf{x}). \quad (203)$$

In this equation, the potential was also replaced by a dimensionless one  $V \rightarrow E_0 V$  with reference energy  $E_0$ . All coefficients together can be combined to a scaling force coefficient

$$F_0 := \frac{T}{2br^*} \frac{E_0}{r^*}. \quad (204)$$

For two cells interacting with each other, the potential depends on the distance of the two cells  $V(|\mathbf{x}_j - \mathbf{x}_i|)$ . Using the chain rule for the gradient yields the force terms

$$\nabla_{\mathbf{x}_j} V(|\mathbf{x}_j - \mathbf{x}_i|) = V'(|\mathbf{x}_j - \mathbf{x}_i|) \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_j - \mathbf{x}_i|}, \quad (205)$$

where the first factor describes the magnitude of the force, and the second one its direction. For  $n$  cells, the equations of motion become

$$\frac{d\mathbf{x}_i}{dt} = F_0 \sum_{\substack{j=1 \\ j \neq i}}^n V'(|\mathbf{x}_j - \mathbf{x}_i|) \frac{\mathbf{x}_j - \mathbf{x}_i}{|\mathbf{x}_j - \mathbf{x}_i|}, \quad \text{for } i = 1, \dots, n. \quad (206)$$

**Morse potential** The Morse potential has been initially used to describe an exact solution of the Schroedinger equations representing the motion of nuclei in a diatomic molecule [42]. However, it has also found its way into biology. Here, it is used to describe the adhesion and repulsion of cells in contact with each other [40, 41]. Following this, the Morse potential is given as a function of the distance between two cells  $i$  and  $j$

$$V(|\mathbf{x}_j - \mathbf{x}_i|) = \begin{cases} e^{-2\alpha(|\mathbf{x}_j - \mathbf{x}_i| - \sigma(r_j + r_i))} - 2e^{-\alpha(|\mathbf{x}_j - \mathbf{x}_i| - \sigma(r_j + r_i))} & \text{for } |\mathbf{x}_j - \mathbf{x}_i| \leq r_j + r_i, \\ e^{-2\alpha(r_i + r_j - \sigma(r_i + r_j))} - 2e^{-\alpha(r_j + r_i - \sigma(r_j + r_i))} & \text{for } |\mathbf{x}_j - \mathbf{x}_i| > r_j + r_i. \end{cases} \quad (207)$$

The parameter  $\alpha$  describes the stiffness of the cells, whereas  $\sigma \in (0, 1]$  defines the optimal distance between two cells in contact as a fraction of the sum of their radii. The resulting norm of the force is then just given by its weak derivative

$$V'(|\mathbf{x}_j - \mathbf{x}_i|) = \begin{cases} 2\alpha (e^{-\alpha(|\mathbf{x}_j - \mathbf{x}_i| - \sigma(r_j + r_i))} - e^{-2\alpha(|\mathbf{x}_j - \mathbf{x}_i| - \sigma(r_j + r_i))}) & \text{for } |\mathbf{x}_j - \mathbf{x}_i| \leq r_j + r_i, \\ 0 & \text{for } |\mathbf{x}_j - \mathbf{x}_i| > r_j + r_i. \end{cases} \quad (208)$$

The effects of the potential now depends on the relative position of two cells. If the cells are too close, they will repel 21 (a)), whereas they will adhere to each other if they are too far from each other but still in contact 21 (c)). In between, an optimal state will be found 21 (b)). If the distance between the cells exceeds the sum of their radii, there can be no physical interaction 21 (d)). This desired behavior can be observed with the Morse potential and its resulting force. The potential starts by decreasing up to its energetic minimum (Fig. 21 (e)). This decrease leads to a negative force causing cells to repel each other (Fig. 21 (f)). The energetic minimum is reached when the distance between two cells  $|\mathbf{x}_i - \mathbf{x}_j|$

equals the sum of their radii  $r_i + r_j$  times the optimality factor  $\sigma \in (0, 1]$ . In this case, no forces act on the cells. As the distance increases again, the potential starts to increase, leading to an attractive force on the cells. Once the distance of the cell centroids exceeds the sum of their radii, the potential will not change anymore. Therefore, there are no longer any forces acting on the cells. Increasing the cell stiffness  $\alpha$  leads to an overall increase of the potential. This mostly increases the norm of the force. However, it also moves the maximum of the force closer to the location of the potential minimum. The subsequent decrease in the adhesive force fits nicely with the idea, that less cell to cell contact leads to less force.

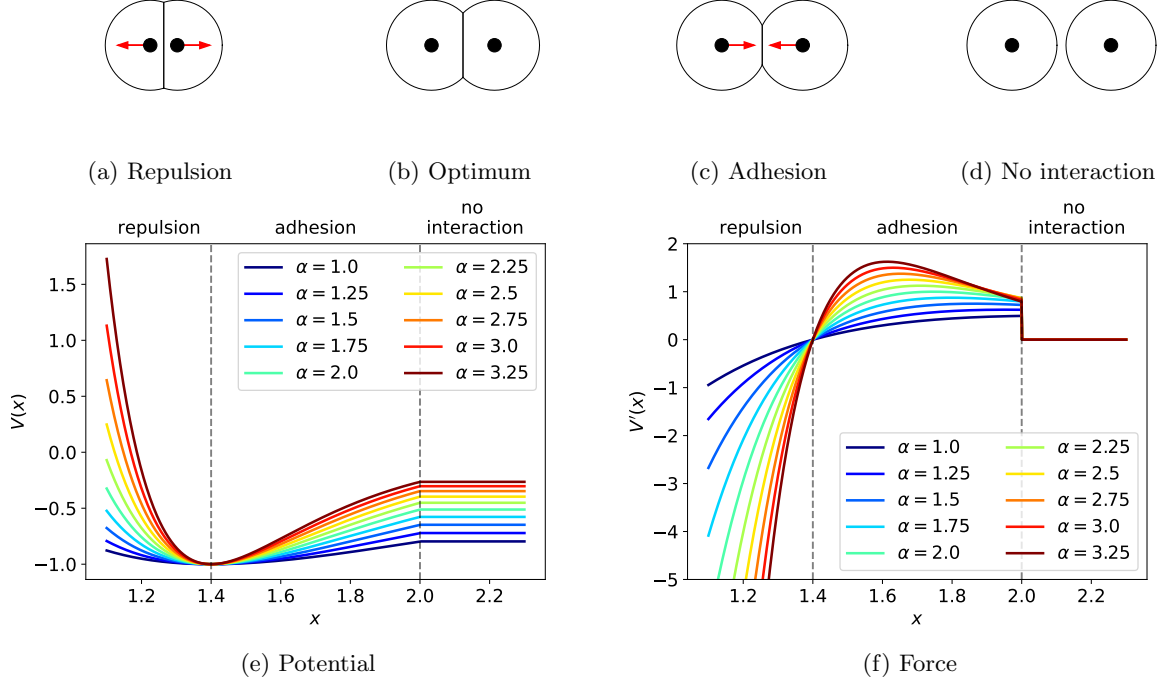


Figure 21: Illustration of the four different cases of interaction. Black dots represent the centroid of the cells. Black lines indicate their cell membranes. For the repulsion (a) and adhesion (c), red arrows show the direction of the respective forces acting on each cell. For the optimal distance (b) as well as disconnected cells (d), no forces are present. In (e) and (f), the potential and its resulting force are depicted respectively for ten different values of the cell stiffness  $\alpha$ . The additional parameters are  $r_i + r_j = 2$  and  $\sigma = 0.7$ .

**Distance after cell division** The two daughter cells emerging after cell division are subject to the same mechanical interactions and their resulting forces as all other cells in the organoid. However, immediately after cell division, their distance towards each other is so small that interactions from other cells become negligible in comparison. Considering only these two cells, their distance after a time  $\Delta t$  is described by

$$\frac{d|\mathbf{x}_1 - \mathbf{x}_2|}{dt} = \frac{d(\mathbf{x}_1 - \mathbf{x}_2)}{dt} \cdot \frac{\mathbf{x}_1 - \mathbf{x}_2}{|\mathbf{x}_1 - \mathbf{x}_2|} \quad (209)$$

$$= \left( \frac{d\mathbf{x}_1}{dt} - \frac{d\mathbf{x}_2}{dt} \right) \cdot \frac{\mathbf{x}_1 - \mathbf{x}_2}{|\mathbf{x}_1 - \mathbf{x}_2|} \quad (210)$$

$$= F_0 V'(|\mathbf{x}_2 - \mathbf{x}_1|) \left( \frac{\mathbf{x}_2 - \mathbf{x}_1}{|\mathbf{x}_2 - \mathbf{x}_1|} - \frac{\mathbf{x}_1 - \mathbf{x}_2}{|\mathbf{x}_1 - \mathbf{x}_2|} \right) \cdot \frac{\mathbf{x}_1 - \mathbf{x}_2}{|\mathbf{x}_1 - \mathbf{x}_2|} \quad (211)$$

$$= -2F_0 V'(|\mathbf{x}_2 - \mathbf{x}_1|) \frac{(\mathbf{x}_1 - \mathbf{x}_2) \cdot (\mathbf{x}_1 - \mathbf{x}_2)}{|\mathbf{x}_1 - \mathbf{x}_2|^2} \quad (212)$$

$$= -2F_0 V'(|\mathbf{x}_2 - \mathbf{x}_1|). \quad (213)$$

In this computation,  $\cdot$  is used to denote the inner product of two vectors. Defining  $h := |\mathbf{x}_2 - \mathbf{x}_1|$ , the distance between daughter cells for a short time scale following cell division follows the initial value



problem (IVP)

$$\begin{aligned}\frac{dh}{dt} &= -2F_0V'(h), & t \in (0, \Delta t), \\ h(0) &= 0.\end{aligned}\tag{214}$$

This IVP is solved by separation of variables and subsequent integration

$$\int_0^{\Delta t} dt = - \int_0^h \frac{1}{2F_0V'(h')} dh'.$$
(215)

It is possible to solve both integrals. The trivial one on the left hand side yields

$$\int_0^{\Delta t} dt = \Delta t.$$
(216)

The remaining integral first requires us to find a primitive function for  $1/V'(h)$ . This is done in several steps starting with a simplified function

$$f(x) = \frac{1}{e^{-x} - e^{-2x}} = \frac{e^{2x}}{e^x + 1}.$$
(217)

When integrating this function, we can substitute  $y = e^x$  and  $dy = e^x dx$ , which leads to

$$\int f(x) dx = \int \frac{e^{2x}}{e^x - 1} dx = \int \frac{y}{y - 1} dy$$
(218)

$$= \int \frac{1}{y - 1} - 1 dy$$
(219)

$$= \int \frac{1}{y - 1} dy - \int 1 dy$$
(220)

$$= \ln(y - 1) + y$$
(221)

$$= \ln(e^x - 1) + e^x$$
(222)

For  $x < 0$  the logarithm becomes complex. However, due to

$$\ln(z) = \ln(-z) + i\pi \quad \text{for } z \in \mathbb{R}, z < 0$$
(223)

we know that  $\ln(1 - e^x) + e^x$  also becomes a valid primitive function of the integral of  $f$ , i.e.

$$\int \frac{e^{2x}}{e^x - 1} dx = \begin{cases} \ln(e^x - 1) + e^x, & \text{for } x > 0, \\ \ln(1 - e^x) + e^x, & \text{for } x < 0 \end{cases}$$
(224)

After cell division, both daughter cells will have the same radius, i.e.  $r_1 = r_2 =: r$ . Therefore, their sum can be replaced by  $2r$ . In the next step,  $x$  is substituted by  $x = \alpha(h' - 2\sigma r)$ . For the definite integral on the right hand side of (215), the lower integral limit is 0. At the same time, the steady state of (214) is reached at  $h = 2\sigma r$ . Thus the values of  $h$  are restricted to the interval  $[0, 2\sigma r]$ . For the evaluation of the integral, this means that only the case of  $x < 0$  is relevant. Following procedure (218)-(222) again for  $1/(2F_0V'(h'))$  yields

$$\int \frac{1}{2F_0V'(h')} dh' = \frac{1}{2\alpha F_0} \left( \ln(1 - e^{\alpha(h' - 2\sigma r)}) - e^{-2\alpha\sigma r} \right).$$
(225)

Using the limits of integration, i.e.  $h' = 0$  and  $h' = h$ , and reorganizing the individual terms, the solution of the initial value problem (214) is given by the solution of the nonlinear equation

$$2\alpha F_0 \Delta t + e^{\alpha(h - 2\sigma r)} + \ln(1 - e^{\alpha(h - 2\sigma r)}) - e^{-2\alpha\sigma r} - \ln(1 - e^{-2\alpha\sigma r}) = 0.$$
(226)

In the following, equation (226) is solved numerically using Newton's method. As expected, the cell division distance  $h$  increases with increasing time steps  $\Delta t$  (Fig. 22). The cell stiffness  $\alpha$  increases the repulsive force, leading to larger distances for shorter time steps. At a distance of  $2\sigma r$ , the potential  $V(h)$  reaches its minimum, i.e. (214) reaches its steady state and no further forces are applied.

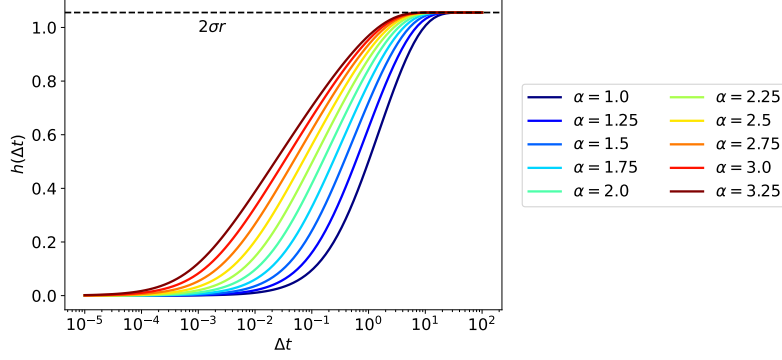


Figure 22: Numerical results for the distance after cell division  $h$  depending on the elapsed time step  $\Delta t$ . Along the color gradient the distance  $h$  for ten different increasing values for  $\alpha$  are visualized. The dashed line at the top is placed at the potential minimum which is also the steady state solution of (214). Used parameters were  $F_0 = 0.1$ ,  $\sigma = 0.7$  and  $r = 0.95 \cdot 2^{-1/3} \approx 0.754$ .

**Cell division direction** The direction of cell division is chosen randomly. In 2D a single angle  $\varphi$  randomly chosen to be in  $[0, 2\pi]$  suffices to describe any direction. In 3D an additional angle  $\theta \in [0, \pi]$  is needed. For a cell at position  $\mathbf{x}$ , two daughter cells emerge at  $\mathbf{x} \pm \delta\mathbf{x}$ . The displacement vector  $\delta\mathbf{x}$  is

$$\begin{aligned} \text{2D: } \delta\mathbf{x} &= \frac{h}{2} \begin{pmatrix} \cos(\varphi) \\ \sin(\varphi) \end{pmatrix}, \\ \text{3D: } \delta\mathbf{x} &= \frac{h}{2} \begin{pmatrix} \sin(\theta) \cos(\varphi) \\ \sin(\theta) \sin(\varphi) \\ \cos(\theta) \end{pmatrix}. \end{aligned} \quad (227)$$

### 3.3.4 Implementation

**Organoid growth standalone** In the first step, the organoid growth model is implemented without the inclusion of transcriptional regulation. In order to reduce the computational load, the first task in the simulation is to select a fixed cell division distance based on (226) for any cell. Therefore, the radius is set to the mean division radius  $\mu_{\text{div}}$  times the cell division factor  $\kappa = 2^{-\frac{1}{\text{dim}}}$ , i.e.  $r = \kappa \mu_{\text{div}}$ . Afterwards, the growth of the radius is directly implemented through equation (179). In every time step and for every cell  $i$ , a random number is chosen from a uniform distribution between 0 and 1, i.e.  $\mathcal{U}(0, 1)$ . If this number is smaller than (187), the cell divides. The cell is replaced by two cells with new radius according to (186). While the distance of cell division is given by the previously fixed value, the direction of the division is chosen randomly according to (227) (Alg. 1). After the division steps, the distances  $|\mathbf{x}_j - \mathbf{x}_i|$  between all cells are calculated. These are used in combination with an explicit Euler scheme to update the cell positions via (203) (Alg. 2).

---

#### Algorithm 1 Cell division pseudocode

---

```

Input:  $R = \{r_1, \dots, r_n\}$ ,  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,  $\delta\mathbf{x}$ 
for  $i = 1, \dots, n$  do
     $p \leftarrow (187)$  ▷ compute division probability
    Choose random number  $\xi \sim \mathcal{U}(0, 1)$ 
    if  $\xi \leq p$  then ▷ cell division criterion
         $r_{i,1}, r_{i,2} \leftarrow (186)$  ▷ cell radii after division
         $R \leftarrow (R \setminus \{r_i\}) \cup \{r_{i,1}\} \cup \{r_{i,2}\}$ 
         $X \leftarrow (X \setminus \{\mathbf{x}_i\}) \cup \{\mathbf{x}_i + \delta\mathbf{x}\} \cup \{\mathbf{x}_i - \delta\mathbf{x}\}$ 
    end if
end for
Output:  $R, X$ 

```

---

**Organoid growth with transcriptional regulation** When coupling the organoid growth model together with the transcriptional regulation model, the general procedure does not change. The transcriptional regulation relies on the cell graph for the cell-cell signaling. Therefore, in each time step,

---

**Algorithm 2** Organoid growth pseudocode

---

$\delta \mathbf{x} \leftarrow (227)$  with  $r = \kappa \mu_{\text{div}}$  ▷ fixed cell division distance  
**Input:**  $R = \{r_1, \dots, r_n\}$ ,  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$   
**for**  $i = 1, \dots, N$  **do** ▷ loop over time  
     $R \leftarrow (179)$  ▷ radius growth  
     $R, X \leftarrow \text{Alg. 1 with } (R, X, \delta \mathbf{x})$  ▷ cell division  
     $X \leftarrow (203)$  ▷ cell displacement  
**end for**  
**Output:**  $R, X$

---

before transcription is initialized, a new cell graph first needs to be generated. Afterwards, the signal value can be determined from the graph, as well as the values of  $u_i$  from the previous time step. In each time step a single step of the explicit Euler method is used to solve the underlying ODE system, e.q. (*M3*). For cell division, symmetric division is assumed, i.e. both transcription factors  $u$  and  $v$  will be distributed evenly among the daughter cell (Alg. 3). Combined with the rest, the routine follows the given steps:

- Update radius
- Check cell division criterion
- Cell division
- Cell displacement
- Compute cell graph
- Compute signal
- Transcriptional regulation

These steps are repeated in every time step to yield the overall method used to generate in silico organoids (Alg. 4). A full list of model parameters is found in table 2. The transcriptional regulation model chosen in this table is represented by *M3*. Both (*M1*) and (*M2*) together with the choice for the parameters of *M3* will be discussed in detail in a later section.

---

**Algorithm 3** Cell division with transcriptional regulation pseudocode

---

**Input:**  $R = \{r_1, \dots, r_n\}$ ,  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,  $\delta \mathbf{x}$ ,  $U = \{u_1, \dots, u_n\}$ ,  $V = \{v_1, \dots, v_n\}$   
**for**  $i = 1, \dots, n$  **do**  
     $p \leftarrow (187)$  ▷ compute division probability  
    Choose random number  $\xi \sim \mathcal{U}(0, 1)$   
    **if**  $\xi \leq p$  **then** ▷ cell division criterion  
         $r_{i,1}, r_{i,2} \leftarrow (186)$  ▷ cell radii after division  
         $R \leftarrow (R \setminus \{r_i\}) \cup \{r_{i,1}\} \cup \{r_{i,2}\}$   
         $X \leftarrow (X \setminus \{\mathbf{x}_i\}) \cup \{\mathbf{x}_i + \delta \mathbf{x}\} \cup \{\mathbf{x}_i - \delta \mathbf{x}\}$   
         $U \leftarrow (U \setminus \{u_i\}) \cup \{\frac{u_i}{2}\} \cup \{\frac{u_i}{2}\}$   
         $V \leftarrow (V \setminus \{v_i\}) \cup \{\frac{v_i}{2}\} \cup \{\frac{v_i}{2}\}$   
    **end if**  
**end for**  
**Output:**  $R, X, U, V$

---

### 3.3.5 Summary

The question of how organoids grow triggered a cascade of various other questions to be answered. Starting on the single cell level, the growth of a cell was described by a logistic growth model. After some time has passed and the cell reaches a certain size, it starts to divide. For the purpose of this model, the stochasticity in the cell division was described by a truncated normal distribution. For the time discrete case, the cell division system was corrected by looking at the division probabilities in every single time step. By using the experimental values as benchmarks, the single cell growth rate and organoid growth rates could be determined. On the organoid level, intercellular mechanical interactions were described by rigorously derived equations of motion together with the use of the Morse potential for adhesion/repulsion similar to [41]. This was extended by addressing the time step dependent distance of cells after division. Overall, the combined model allows the creation of 2D and 3D organoids in the form of cell radii and positions.

---

**Algorithm 4** Organoid growth with transcriptional regulation pseudocode

---

$\delta \mathbf{x} \leftarrow (227)$  with  $r = \kappa \mu_{\text{div}}$  ▷ fixed cell division distance  
**Input:**  $R = \{r_1, \dots, r_n\}$ ,  $X = \{\mathbf{x}_1, \dots, \mathbf{x}_n\}$ ,  $U = \{u_1, \dots, u_n\}$ ,  $V = \{v_1, \dots, v_n\}$   
**for**  $i = 1, \dots, N$  **do** ▷ loop over time  
     $R \leftarrow (179)$  ▷ radius growth  
     $R, X \leftarrow \text{Alg. 3 with } (R, X, \delta \mathbf{x}, U, V)$  ▷ cell division  
     $X \leftarrow (203)$  ▷ cell displacement  
    Compute cell graph  $G$   
     $\mathbf{s} \leftarrow (163)$  or  $(165)$  ▷ Calculate signal  
     $U, V \leftarrow (M1), (M2)$  or  $(M3)$  ▷ transcriptional regulation  
**end for**  
**Output:**  $R, X, U, V$

---

Model parameter	Fixed value	Description
$T$	234.74	Evolution time
$N$	5000	Number of time steps
$r^*$	1	Maximum radius
$\lambda$	0.083	Cell growth rate
$F_0$	0.01	Displacement force scaling factor
$\alpha$	3	Cell stiffness
$\sigma$	0.7	Cell-cell distance optimality factor
$-\Delta \varepsilon_u$	6-7.87	Energy difference w.r.t. binding of $u$
$-\Delta \varepsilon_v$	6	Energy difference w.r.t. binding of $v$
$-\Delta \varepsilon_s$	2	Energy difference w.r.t. binding of $s$
$-\Delta \varepsilon_{uv}$	2	Energy difference w.r.t. combined binding of $v$ and $s$
$r_u$	1	Transcription rate of $u$
$r_v$	1	Transcription rate of $v$
$r_u$	10	Decay rate of $u$
$r_v$	10	Decay rate of $v$

Table 2: List of model parameters and their description. Fixed values are used throughout the remainder of this thesis unless stated otherwise. Parameters downwards of  $-\Delta \varepsilon_u$  are representative for the transcriptional regulation model  $M3$ .

## 4 Computational Results

### 4.1 Analysis of experimental data

Quantifying and understanding experimental data is vital to any scientific investigation. An initial data analysis for the ICM organoids already provided quantitative results on the different cell populations N+G<sup>-</sup>, N-G<sup>+</sup>, DN and DP [23]. There, it was found that the proportions of N+G<sup>-</sup> cells of 24 h- and 48 h-organoids are comparable to that of the mid (64-90 cells) and late (90-120 cells) stage embryos provided by [48]. Additionally, the relative increase in proportions of N-G<sup>+</sup> cells between 24 h- and 48 h-organoids could be compared to the increase between mid and late stage blastocysts. In a neighborhood analysis, it was found that cells of one type tend to have an increased amount of equal cell neighbors when compared to a random distribution. This suggests a local clustering of the different cell types.

In this section, the investigation of the ICM organoid data is continued by applying the chosen quantification methods (PCF and Moran's  $I$ ). Moran's  $I$  allows the effect of local clustering in ICM organoids to be quantified on an individual level. The PCFs, on the other hand, are designed to tell us how the different cell types are distributed spatially in each organoid. Combined this will lay the foundation for the quantitative comparison of in vitro and in silico data.

#### 4.1.1 24 h-organoids show no clear distribution of cell types

The goal of this section is to identify a common spatial pattern among the 24 h-organoids. The PCF envelopes of the 24 h-organoids were qualitatively observed and classified (Fig. 23). Out of 34 organoids, 13 of those (green frames) have PCFs that are mostly close to 1 for all distances. This is an indication for a random distribution of the two cell types. However, a small number of five organoids (blue frames) exhibit a pattern where N+G<sup>-</sup> and N-G<sup>+</sup> cells get radially separated close to the maximum distance. Surprisingly, the N+G<sup>-</sup> cells are the ones found at long distance, even though it was already shown that at later stages, the average GATA6 expression levels are the ones that are greater at distances further away from the mouse blastocyst centroid [18]. What these have in common however, is the relatively low amount of cells ranging from 295 to 444. On the other hand, an organoid like, e.g. ID = 5 shows signs of a radial separation that is the exact opposite. At the same time, it already went through several cell divisions ending up with 679 cells.

The width of the PCF envelopes is influenced by two different factors. The first one, is the amount of undetermined cells in the data, i.e. DN or DP cells. The more of these cells were found pairing up at some distance, the wider the range of sampled PCFs will be. The second factor are the cell type proportions. If there are only few N+G<sup>-</sup> cells present compared to N-G<sup>+</sup>, then the PCF corresponding to N+G<sup>-</sup> varies stronger within the sampling process. This can be seen, e.g. for IDs 15 and 17 or in the extreme cases with only one cell type 10 and 11 (red frames). In summary, an overall matching pattern could not be accurately determined. The largest overlap was found for patterns that either resemble a random distribution of cells or could simply not be determined accurately enough using the PCFs.

#### 4.1.2 48 h-organoids exhibit radially distributed cell types

Applying the same procedure as before to generate PCF envelopes for the 48 h-organoids leads to drastically different results (Fig. 24). A total of 34 out of 42 organoids exhibit a strong similarity to each other. The pattern they represent is characterized by a small drop at the beginning and a sharp rise towards the maximum pair distances for N-G<sup>+</sup>. For N+G<sup>-</sup> there is a monotonous decrease to be found. These are signs for an engulfing pattern, where N+G<sup>-</sup> cells in the center of the organoid are surrounded by N-G<sup>+</sup> cells at its boundary. This proves on a quantitative level that the organoids from [23] are indeed in some kind of sorted state after 48 h. The question on how this state is ultimately achieved, cannot be answered at this point. The values of the PCFs at distance 1 that are slightly above 1 already hint an increased cluster formation that is further quantified in the following section.

#### 4.1.3 Experimental data reveals cluster formation

In [23], it was shown that on average cells of equal type exhibit more neighbors of the equal type, relative to the prevalent proportions. This can be further quantified using Moran's  $I$  which was already established as a way of describing the spatial auto-correlation of the two cell types. Here, we investigate its changes with respect to the elapsed time since the organoids were seeded and the cell number  $n$  (Fig. 25). At 24 h, the organoids consistently reveal that  $I > 0$ . This indicates that even at this point, the distribution of the cells is not completely random. On average,  $I$  increases even further when reaching the 48 h mark.

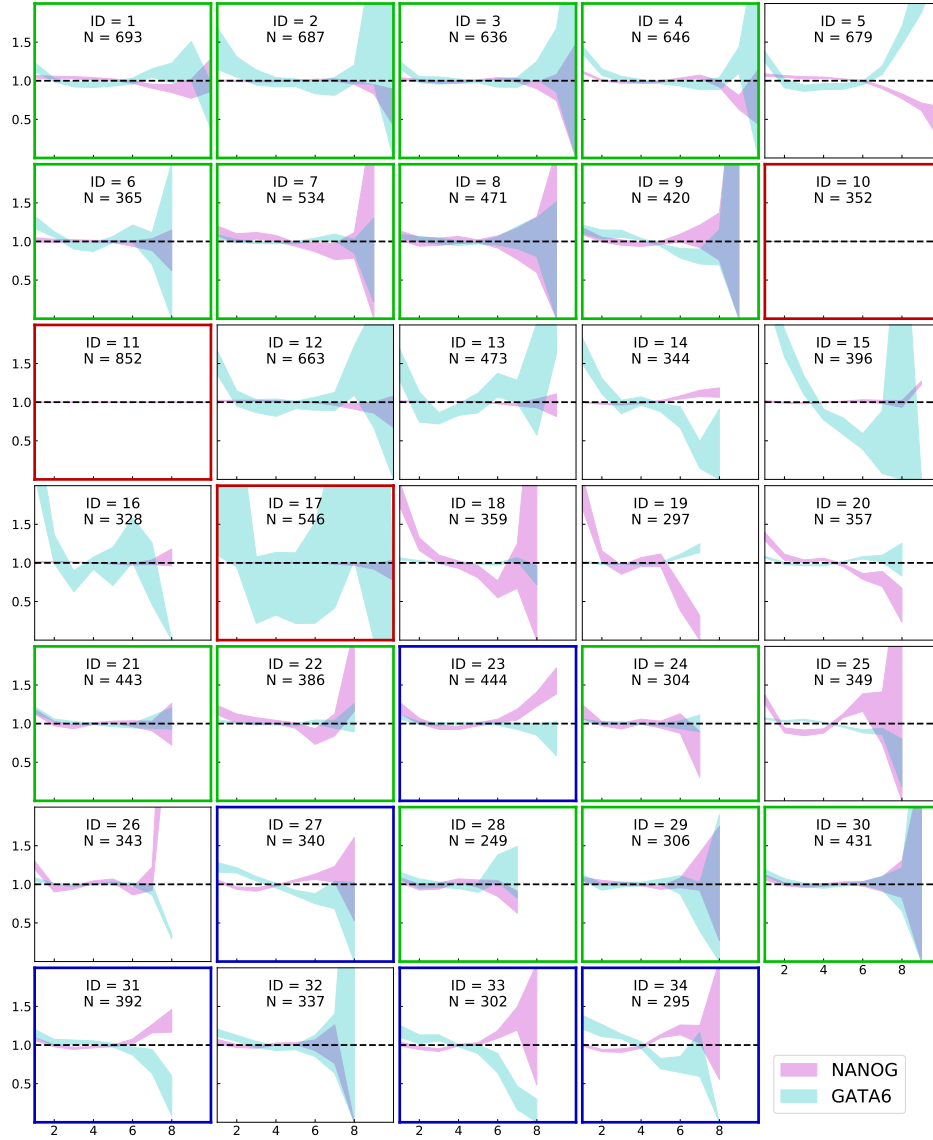


Figure 23: PCF envelopes for the data of the 24h-organoids. For each envelope, 1000 samples are used. Green borders highlight the organoids that show the most common pattern. In contrast to this, red borders highlight organoids, where cell type proportions are extremely skewed up to the point of no cells of one type. In blue, organoids with a similar pattern, that is not random, are highlighted.

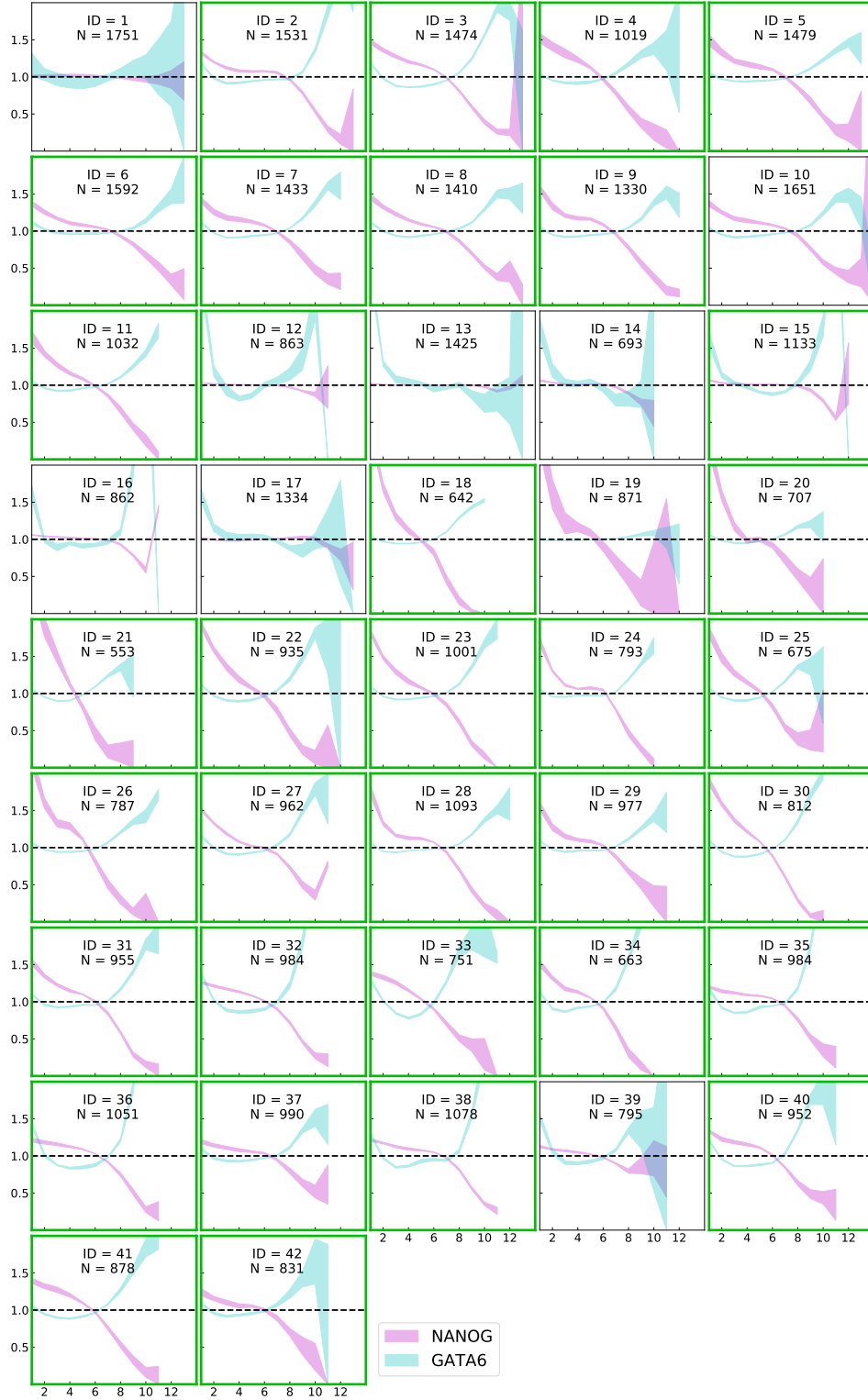


Figure 24: PCF envelopes for the ICM organoid data that developed for 48 h. For each envelope, 1000 samples are used. Green borders highlight the organoids that show the most common pattern. This assignment was done by qualitative observation. Two or more outliers, or envelopes that are too wide, were not specifically highlighted.

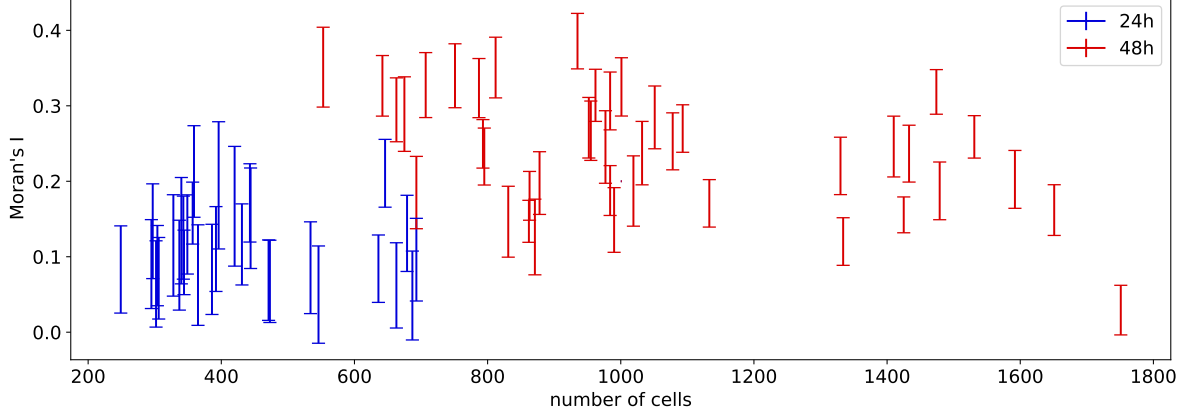


Figure 25: Visualization of minimum and maximum values of Moran’s  $I$  with respect to  $n$ . Each bar shows the resulting range of  $I$  for 1000 samples of randomly chosen fates for DN and DP cells.

Assuming symmetric division, i.e. each cell division leads to an equal distribution of NANOG and GATA6 between the daughter cells, this is not surprising. But even in asymmetric divisions, e.g. an N+G− cell exhibits low expressions of GATA6 such that both daughter cells cannot have high expressions of GATA6. Under these conditions, one would expect an increased clustering with increasing cell numbers. However, the correlations between the number of cells and  $I$  paint a different picture (Tab. 3). There is no correlation found in the 24h-organoids alone, whereas at 48 h, there is slightly negative correlation. The correlations of 24 h and 48 h combined are still slightly positive.

$r_{n,X}$	24 h	48 h	24 h & 48 h
$X = I_{\min}$	−0.02	−0.41	0.43
$X = I_{\text{mean}}$	−0.08	−0.44	0.38
$X = I_{\max}$	−0.14	−0.49	0.28

Table 3: Pearson correlation coefficients  $r_{n,X}$  between the number of cells  $n$  and values  $X$ . From the top to the bottom row,  $X$  was chosen to represent the minimum, arithmetic mean and maximum values of  $I$  found for the sampled organoids. The columns denote which part of the dataset was used. Either only the organoids labeled as 24 h, 48 h or both.

#### 4.1.4 Conclusion

The analysis of the experimental data from [23] shows that there is a clear distinction between the patterns for 24 h organoids and 48 h organoids. At 24 h no clear pattern could be determined. At 48 h however, the PCFs show a clear trend towards a radial separation of the two cell types at long distances. A similar observation was already made in the ICM of the mouse embryo, where the expression levels of GATA6 increase with the distance to the ICM centroid, whereas NANOG expressions decrease [18]. With the use of Moran’s  $I$  it could be determined that the clustering of cell types in any of the organoids is enough for it to not be considered random. It was also found that there is an increase of  $I$  with increasing number of cells up to some point. One hypothesis why it does not increase further is that the amount of cluster formation by cell division is overshadowed by the sorting mechanism. For this to be feasible, this would need to be a mechanism that gets increasingly more dominant, else it would not be able to compete with the exponential growth of the organoid. A different hypothesis might be a position biased differentiation, i.e. cells are able to rapidly switch their cell fates based on their position. Possible mechanism that could influence this could be of mechanical nature, like the internal pressure of the organoid, or based on cell-cell communication. In this thesis, the focus lies only on cell-cell communication. Since the 24 h organoids are not classifiable, a special emphasis will be on the 48 h organoids.

## 4.2 Transcriptional regulation model comparison

### 4.2.1 Comparison setup

So far, three models were introduced with the same purpose of describing the temporal evolution of transcription factors corresponding to GRN 10. In this section, some of the advantages and disadvantages



of each model will be highlighted. Furthermore, the major question of whether the models are capable to generate two cell types in terms of their expression levels, will be answered. In terms of attractiveness, the auto activation dominant model (*M3*) outshines the other models by far due to the successful derivation of the stability interval (160). This makes it the most comfortable candidate for any further simulations. At this point however, it is unclear if this model fulfills the same purpose as the phenomenological model (*M1*) or (*M2*). Therefore, all three models are put to the test. For a single cell, each model should be able to assume two different equilibrium states, which are supposed to represent the  $u^+v^-$  and  $u^-v^+$  state. For low values of  $s$ , this means that the model parameters have to be chosen, such that  $u$  is high and  $v$  is low. For large values of  $s$ , it should be the exact opposite. Therefore, arbitrary values are defined as

$$s = 10^{-3} \implies \begin{cases} u \approx 0 \\ v \approx 0.1 \end{cases}, \quad s = 1 \implies \begin{cases} u \approx 0.1 \\ v \approx 0 \end{cases}. \quad (228)$$

These values are of no biological relevance and serve only as reference values for high and low signal or expression values. By changing the relevant coefficients  $\eta_u$  and  $\eta_s$ , or  $a_u$  and  $a_s$ , these values could be arbitrarily scaled. For model (*M3*), the model parameters can be chosen directly. The nonzero steady states in (122) and (123) are given by

$$u = \frac{r_u}{\gamma_u} - \frac{1 + \eta_s s}{\eta_u}, \quad v = \frac{r_v}{\gamma_v} - \frac{1 + \eta_s s}{\eta_v(1 + \eta_s \eta_{vs} s)}. \quad (229)$$

Choosing  $r_u = r_v = 1$  and  $\gamma_u = \gamma_v = 10$  yields  $r_u/\gamma_u = r_v/\gamma_v = 0.1$ . The remaining parameters now need to be chosen, such that the second term becomes significantly smaller than the first one. The full list of parameters is given in table 4

$r_u, r_v$	$\gamma_u, \gamma_v$	$-\Delta\varepsilon_u$	$-\Delta\varepsilon_v$	$-\Delta\varepsilon_s$	$-\Delta\varepsilon_{vs}$
1	10	7	6	2	2

Table 4: Model parameters used for the simulations of the auto activation dominant model (*M3*).

Using these parameters in (122) and (123), yields

$$\begin{cases} u(10^{-3}) = 0.0991 \\ v(10^{-3}) = 0 \end{cases}, \quad \begin{cases} u(1) = 0 \\ v(1) = 0.0996 \end{cases}. \quad (230)$$

For comparison, most of the parameters of models (*M1*) and (*M2*) were fitted to match the same expression values for both states respectively. To this end, the quadratic error was minimized with regard to the model parameters, i.e.

$$\min_{\text{parameters}} (u(10^{-3}) - 0.0991)^2 + ((v(10^{-3}) - 0)^2 + (u(1) - 0)^2 + (v(1) - 0.0996)^2). \quad (231)$$

This was realized using *SciPy*'s `curve_fit` function. Functions  $u$  and  $v$  were provided as the steady state solutions of (*M1*) and (*M2*), i.e.  $\frac{du}{dt} = 0 = \frac{dv}{dt}$ , calculated via Newton's method. The resulting parameters for both models are found in table 5. This first result proves that each model is capable to generate states of different expression. Furthermore, it was shown that the models could be fit to rather arbitrary steady state values. In future research, this would allow the models to be adjusted according to possible experimental measurements of NANOG and GATA6 concentrations under activation and inhibition through an external signal. It remains to show the complete picture of the steady states achieved by the models as well as the temporal evolution towards these.

$r_u, r_v$	$\gamma_u, \gamma_v$	$\ln(a_u)$	$\ln(a_v)$	$\ln(a_s)$	$k$
1	10.72	1.95	2.03	6.85	2

$r_u, r_v$	$\gamma_u, \gamma_v$	$\eta_{xx}$	$-\Delta\varepsilon_u$	$-\Delta\varepsilon_v$	$-\Delta\varepsilon_{xu}$	$-\Delta\varepsilon_{xv}$	$-\Delta\varepsilon_s$ (111)	$-\Delta\varepsilon_{xs}$ (112)	$s_0$ (127)
2.25	9.99	0.07	5.8	6.12	2.73	0.08	3.13	2.43	0.03

Table 5: Fitted model parameters for the phenomenological model (*M1*) (above) and the thermodynamic model (*M2*) (below). The Hill coefficient was predefined as  $k = 2$  to remain consistent with literature [27, 28]. The thermodynamic quantities  $-\Delta\varepsilon_s$  and  $-\Delta\varepsilon_{xs}$  are the result of (111) and (112). The value  $s_0$  resembles the critical signal value (127) from the auto activation dominant model (*M3*)

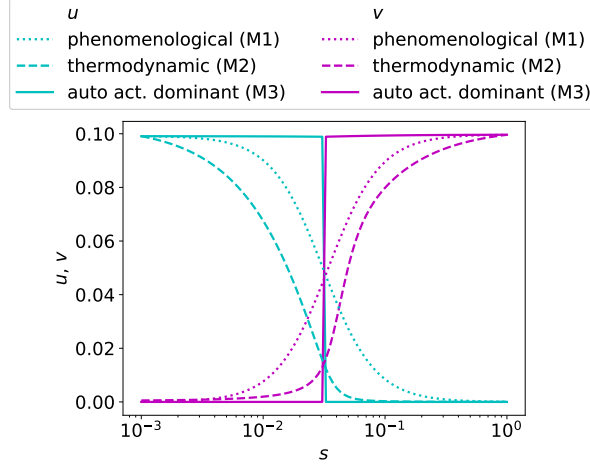


Figure 26: Expression values  $u$  and  $v$  in steady state with respect to different signal values  $s \in [10^{-3}, 1]$ . All lines represent the simulations performed for a single cell. Different line styles correspond to the three models compared in this study (solid: (M3), dashed: (M2), dotted: (M1)).

#### 4.2.2 Comparison of expression values

In each model, the steady states change with respect to  $s$  (Fig. 26). While (M3) exhibits a discontinuous transition at  $s = 0.03$ , models (M1) and (M2) show a smooth transition towards the extreme states. On the other hand, (M3) and (M1) are nearly point symmetric at the point of transition, whereas (M2) is not symmetric. In the time evolution of  $u$  and  $v$ , each model ends up in nearly the same steady state over the time  $T = 2$  (Fig. 27). The differentiation times of (M3) and (M1) are almost perfectly matching. The same holds true for the actual trend of  $u$  and  $v$ . Model (M2) exhibits an initial decrease in both expressions before one is finally overtaking the other. This suggests a comparable development of  $u$  and  $v$  up to a value close to that of equal expressions found in figure 26, i.e. where both lines intersect. A more detailed visualization of these phenomena is found in the phase portraits of the respective models (Fig. 28). Each model shows a unique reaction towards changes of  $s$ . With increasing  $s$ , the steady state of (M1) moves from  $u^+v^-$  to  $u^-v^+$  in what is almost a straight line (Fig. 28 (a)). Likewise, its stream lines also seem to navigate towards the steady state in a straight line. Contrary to this, (M2)'s steady state exhibits a hyperbolic like trajectory passing through regions of comparatively low expressions of both  $u$  and  $v$  (Fig. 28 (b)). Its streamlines also first point towards lower expressions of  $u$  and  $v$  before finally navigating towards the steady state. The phase portraits in figure 28 (c) have already been discussed in section 3.1.4 and were only included here for direct visual comparison.

#### 4.2.3 Conclusion

The comparison of the three models (M1), (M2) and (M3) showed that all three models are capable of generating cells that are either in  $u^+v^-$  or  $u^-v^+$  state. The mechanistic thermodynamic model (M2) is believed to be the one that most accurately captures the dynamics of transcriptional regulation. Hence, one could argue that any deviation in expression characteristics leads to a neglect of any intermediate states before the state of maximum/minimum expressions. The temporal evolution of  $u$  and  $v$  however, shows only slight deviations that mainly arise due to the lowered basal expression values in (M2). The overall trend remains the same. This leads to the conclusion that all three models are equally well suited to decide between the two cell fates. However, it should not be neglected that the model parameters were chosen specifically for this purpose. For different needs, (M3) or possibly even (M1) might lack some necessary information about the transcriptional regulation that is included in (M2). Further research and data, e.g. highly time resolved expression levels of NANOG and GATA6, would be required to determine which of these models best reflects biology. For the research pursued in this thesis, the focus lies mostly on the steady states of the systems such that the evolution towards the steady states is less important in comparison. Hence, for the remainder of this thesis, it is sufficient to use the auto activation dominant model (M3) for further computational investigations. Unless stated otherwise, the associated model parameters are the ones found in table 4.

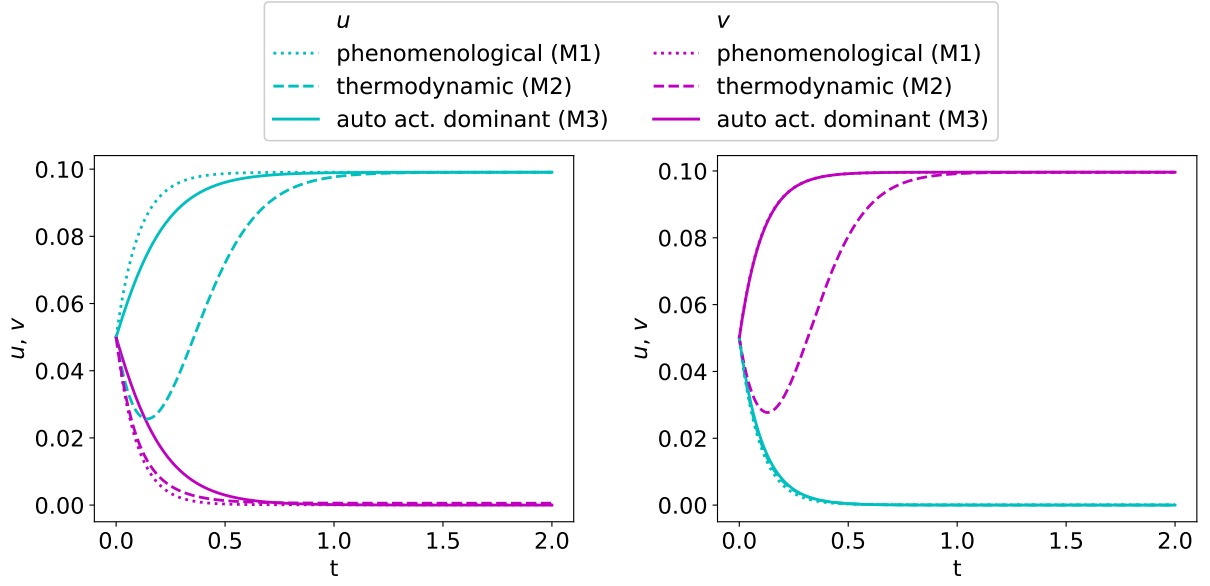


Figure 27: Temporal evolution of  $u$  and  $v$  for a single cell with  $s = 10^{-3}$  (left) and  $s = 1$  (right). Different line styles correspond to the three models compared in this study (solid: (M3), dashed: (M2), dotted: (M1)).

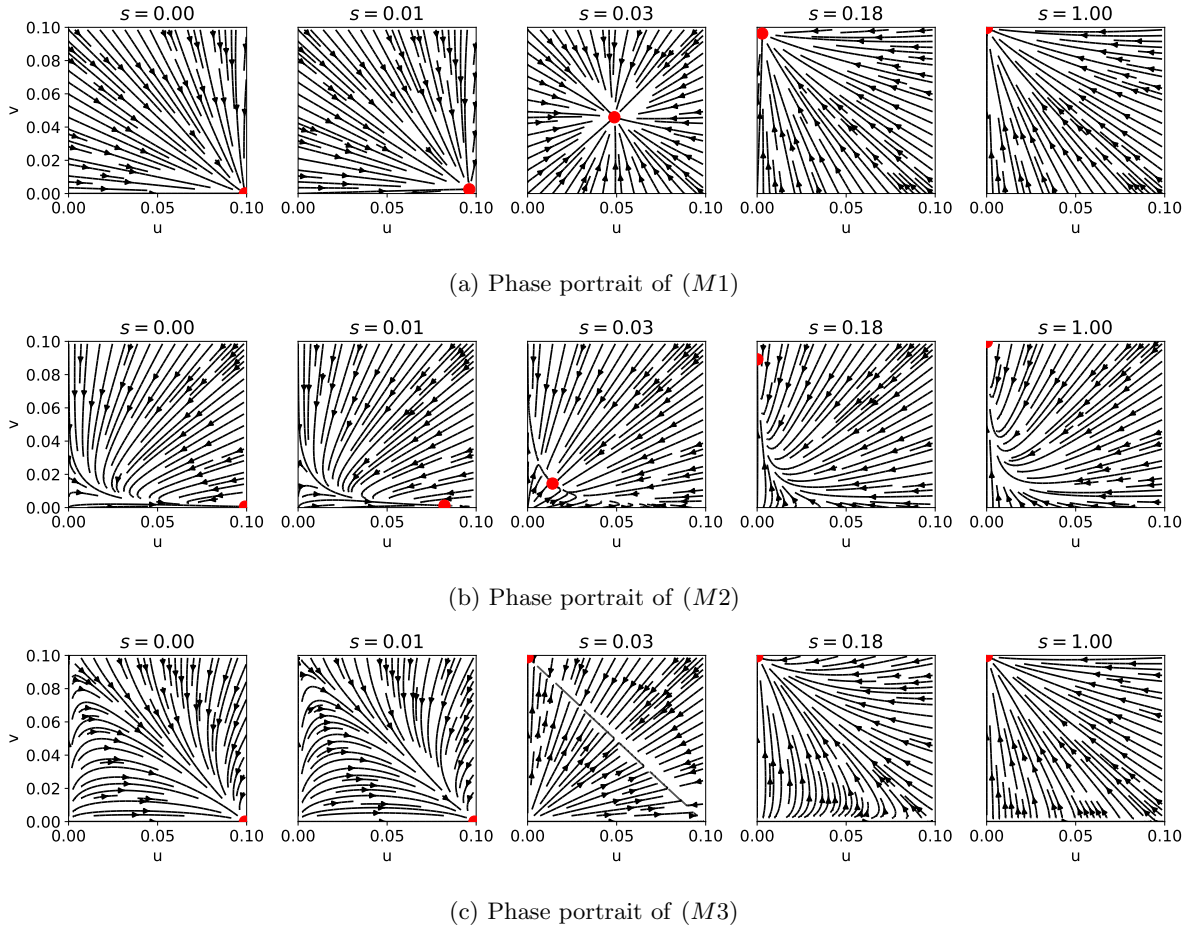


Figure 28: Phase portrait comparison of the three different transcriptional regulation models with different values of  $s$ . Arrows show the path from any position towards the respective steady state marked as a red dot.

### 4.3 Nearest neighbor signaling leads to checkerboard patterns

#### 4.3.1 Color coding

In the following sections, cells will usually be in  $u^+v^-$  or  $u^-v^+$  state. However, sometimes cells might not be in steady state at the end of the simulation, therefore the cells will be visualized using their current expression levels of  $v$  as well as a color range from cyan ( $u^+v^-$ ) to magenta ( $u^-v^+$ ) (Fig. 29). This colorbar will only appear here in this instance but the coloring carries the same meaning for any of the simulated organoids. This choice was based on personal preference and the attempt to improve the "signal to noise ratio" of the upcoming visualizations according to [68].



Figure 29: Colorbar used throughout this thesis to visualize the expression values of  $v$ . The values range from the lowest possible value 0 to the upper bound given by  $r_v/\gamma_v = 0.1$ .

#### 4.3.2 Checkerboard pattern

Models of cell differentiation described as lateral inhibition, which are characterized by an adjacency-based activation, tend to form patterns of alternating cell types [26]. One goal of this section is to show that the additional introduction of interactions in the proposed transcriptional regulation model leads to the same type of pattern. Furthermore, the resulting patterns will be classified with the use of PCFs and Moran's  $I$ . For the simulations in this section, organoid growth and transcriptional regulation were implemented sequentially, i.e. the organoid was grown a set amount of time, before transcriptional regulation was activated. Upon activation, growth is deactivated, such that the transcriptional regulation takes place on a static geometry. The growth time  $T = 273.74$  resembles the time needed for a single cell to yield a 3D organoid with roughly the same amount of cells as the 48h ICM organoids in [23]. In 3D, this would be around 1041 cells. In 2D, the same time is used to generate an organoid with 302 cells (Fig. 30). For  $-\Delta\varepsilon_u = 6.1$ , the organoid mostly consists of  $u^-v^+$  cells with  $u^+v^-$  spread evenly throughout the organoid, never touching any cell of its own type (Fig. 30 (a)). The number of  $u^+v^-$  cells increases for  $-\Delta\varepsilon_u = 7.3$  (Fig. 30 (b)). Hence, the cells cannot fully avoid being neighbored to cells of their own type. Instead, the cells manage to avoid being adjacent to too many cells of equal type. Finally,  $-\Delta\varepsilon_u = 7.8$  leads to the same type of pattern as (a), with the roles of  $u^+v^-$  and  $u^-v^+$  reversed (Fig. 30 (c)). For quantification of the arising patterns, the interval (6, 7.87) was divided into ten different values, excluding the interval boundaries. These values were used as  $-\Delta\varepsilon_u$  to generate a variety of different cell type proportions. The corresponding PCFs  $\rho_u$  and  $\rho_v$  can be interpreted at several key distances. At distance  $k = 1$ , low proportions of  $u^+v^-$  cells, i.e.  $-\Delta\varepsilon_u$  close to 6, lead to  $\rho_u(1)$  close to 0. This means that there are no or only few pairs of equal cells found directly adjacent to each other. In other words,  $u^+v^-$  avoid being adjacent to themselves. As proportions increase, cells are no longer able to avoid adjacency to the same type. Hence, the PCF values move closer to 1. Conversely,  $\rho_v$  exhibits the same behavior for  $u^-v^+$  cell pairs. For distances  $k = 2$ , an increased amount of equal cell pairs is found. Since the cells avoid being adjacent to cells of equal type, this means that the optimal way to arrange the cells in space is to spread them evenly throughout the tissue. Apparently, as seen by the PCF, this means higher occurrences of equal cell types at distance  $k = 2$ . The increase of  $u^+v^-$  proportions leads again to the peak moving closer to 1. Also in both PCFs, the medium ranges from  $k = 3$  to  $k = 17$  show almost no deviations from 1 independent of the cell type proportions. This is perfectly characteristic for a checkerboard pattern, i.e. the previously mentioned pattern of alternating  $u^+v^-$  and  $u^-v^+$  cells. At long distances above  $k = 18$ , the PCFs exhibit no clear trend. The low amount of cell pairs found in this region are able to drastically change PCF values based on the a few single cell fates. However, this was to be expected since the cell signaling only incorporates communication between neighbors. As such, they are incapable of influencing the cell fates of cells far beyond their reach.

Moran's  $I$  shows clear signs of a low spatial auto-correlation of the two cell types (Fig. 32). Values are found roughly between  $-0.4$  and  $-0.3$  in 2D, or  $-0.25$  and  $-0.15$  in 3D. A values of  $-1$  is cannot be achieved. A simple explanation for this is that if one cell type is completely isolated from cells of the same type, the other cell type must have neighbors of equal type. Thus, there will always be an auto-correlation between neighboring cells. This can only be eliminated using an ideal grid like e.g. in figure 5. Therefore, the values of  $I$  can never attain  $-1$  if cells have more than four neighbors, leading to values of  $I$  larger than  $-1$ . Furthermore, 3D cells have even larger values of  $I$ . Following the same reasoning, this is attributed to the increased number of neighbors when comparing 2D and 3D.

By increasing  $-\Delta\varepsilon_u$ , the amount of  $u^+v^-$  cells increases which is covered in more detail in the following section. This increase in proportions also influences  $I$ . In 3D,  $I$  decreases until the minimum is reached where cell type proportions are equal. Afterwards,  $I$  increases again. In 2D, this behavior cannot be observed, which is likely connected to large jumps in cell type proportions with respect to  $-\Delta\varepsilon_u$  that is observed in the next section.

In conclusion, the PCFs hint towards a low spatial auto-correlation which is additionally confirmed by the overall low values found for Moran's  $I$ . Furthermore, PCF values with distances above 1 hint towards an even distribution of the given cell types. These observations were performed for various cell type proportions, which will be carefully examined in the next step.

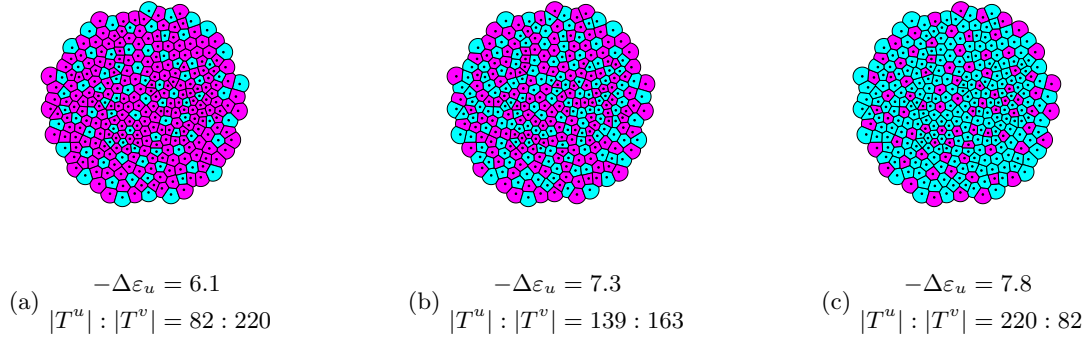


Figure 30: Simulated patterns on a 2D organoid with 302 cells for three different values of  $-\Delta\varepsilon_u$ . The colors of the cells emerge from a color gradient ranging from the minimum (cyan) to the maximum (magenta) expression values of  $v$ . This translates to  $u^+v^-$  cells being colored in cyan and  $u^-v^+$  cells in magenta.

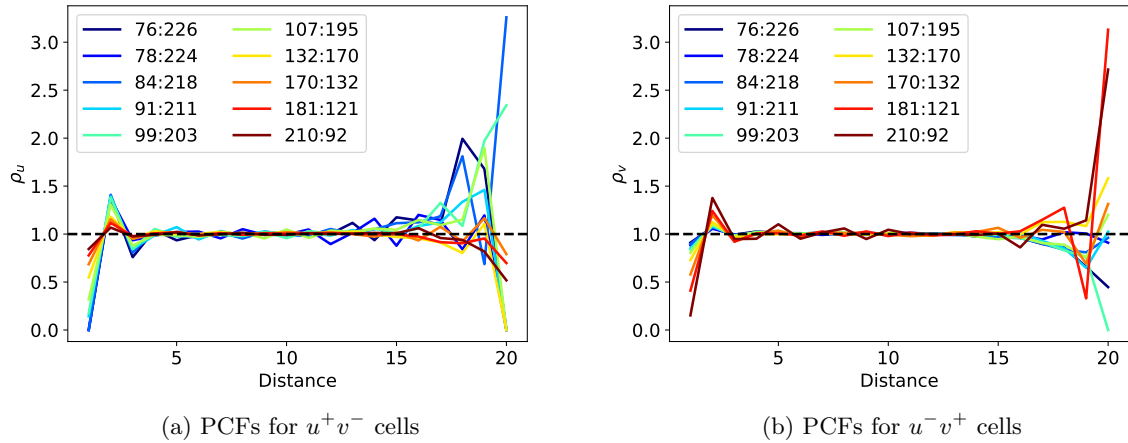


Figure 31: PCFs  $\rho_u$  and  $\rho_v$  for the simulated organoids using averaged NN signaling. Different ratios of cell types  $|T^u| : |T^v|$  (cf. (6), (7)) are depicted in the legend. These arise as the result of ten equidistantly increasing choices for  $-\Delta\varepsilon_u$ . The horizontal and dashed black line at 1 resembles the PCF values of an ideal uniform distribution of two different cell types. Values above 1 mean that more pairs are found at the respective distance, values below mean less.

#### 4.3.3 Cell type proportions

In the mouse embryo, Epi and PrE cells appear in robust proportions, on average which are assumed to be a crucial part of development [48, 30]. The data provided by the former yields an average composition in the ICM of  $50.94\% \pm 13.42\%$  Epi cells and  $29.63\% \pm 18.06\%$  PrE cells for late stage embryos ( $\geq 90$  cells). Therefore, the capability of model ( $M3$ ) to generate different cell type proportions, needs to be explored. The stability interval (160) determines the range of all inhomogeneous steady states and implicitly provides all the possible cell type proportions, the model is able to generate. Therefore, the

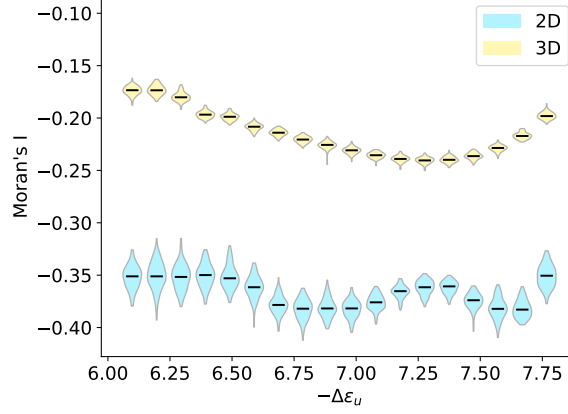


Figure 32: Combined violin plots for Moran's  $I$  in 2D and 3D. A total of 18 different values of  $\Delta\epsilon_u$  were considered. Each violin plot consists of 100 organoids simulated to  $T = 235$ . Respective median values are depicted as horizontal black lines.

exact stability interval first needs to be calculated using the NN signal (163). As demonstrated in (230), the steady state values for  $u$  are either 0 or approximately  $r_u/\gamma_u$ . Hence, the NN signal (163) yields

$$s_i = \frac{r_u}{\gamma_u} \frac{K}{|N_G(i)|}, \quad K \in \mathbb{N}, K \leq |N_G(i)|. \quad (232)$$

This means the only decisive factor is the proportion of  $u^+v^-$  cells adjacent to cell  $i$ . The theoretical minimum and maximum are therefore given by

$$\min_{i \in \{1, \dots, n\}} s_i = 0, \quad \max_{i \in \{1, \dots, n\}} s_i \approx \frac{r_u}{\gamma_u}. \quad (233)$$

Furthermore, the stability interval (160) can be replaced by the approximation (175). With  $r_u = r_v$  and  $\gamma_u = \gamma_v$ , two of the logarithmic terms vanish, leaving only

$$-\Delta\epsilon_v < -\Delta\epsilon_u < -\Delta\epsilon_v + \ln \left( 1 + e^{-\Delta\epsilon_v - \Delta\epsilon_{vs}} \frac{r_u}{\gamma_u} \right). \quad (234)$$

For the concrete parameters used in the simulations (Tab. 4), this leads to the intervals

$$-\Delta\epsilon_u \in (6, 7.87) \quad \Longleftrightarrow \quad \eta_u \in (403.43, 2606.08). \quad (235)$$

For simulations, this interval was divided into 20 equidistant values for  $\Delta\epsilon_u$ . For each of these values 100 organoids were grown in the simulation for a time of  $T = 235$ , resulting in  $315 \pm 45$  cells per organoid. From this point on, organoid growth was deactivated and the transcriptional regulation initiated. This results in a series of different organoids which show the transition from homogeneous  $u^-v^+$  to homogeneous  $u^+v^-$  states (Fig. 33). A nearly monotonous decrease of the mean of  $u^-v^+$  cell proportions is observed over the increase of  $-\Delta\epsilon_u$ . In contrast to this, the proportions of  $u^-v^+$  increase. At the left and right boundaries homogeneity is achieved. The standard deviation in cell type proportions for each value of  $-\Delta\epsilon_u$  was so low, that for the visualization of the error bars 10 times the standard deviation was used. The maximum of these standard deviations was found to be around 0.95%. The overall trend is therefore nearly identical in all organoids.

In 2D, the regions close to the boundaries of (235) suggest that proportions with about 73% of one cell type and 27% of the other are the maximum and minimum cell proportions achievable before reaching homogeneity. Likewise, a similar yet smaller jump is seen for 3D with 17% at the left and 20% at the right end. Analytically, this can be explained by the combination of neighbor proportion dependent signal (232) and the tipping point of the cell fates described via the critical signal value  $s^*$  (127). For  $s < s^*$  a cell adopts the  $u^+v^-$  fate. Likewise, values  $s > s^*$  lead to the  $u^-v^+$  fate. Replacing  $s^*$  in equation (127), by the discrete expression (232) yields

$$\frac{r_u}{\gamma_u} \frac{K}{|N_G(i)|} = \frac{r_u \gamma_v \eta_u - r_v \gamma_u \eta_v}{r_v \gamma_u \eta_v \eta_{vs}}. \quad (236)$$

This equation allows the identification of the maximum number  $K^{\max}$  of  $u^+v^-$  cells that can be supported in the neighborhood of a cell

$$K^{\max} := \left\lfloor |N_G(i)| \frac{\gamma_v \eta_u - \eta_v}{r_v \eta_v \eta_s \eta_{vs}} \right\rfloor. \quad (237)$$

Here,  $\lfloor x \rfloor$  describes the floor function, i.e. the nearest lower integer of a number  $x$ . For values of  $-\Delta\epsilon_u$  slightly above  $-\Delta\epsilon_v$ ,  $K^{\max}$  will still be 0. This means that  $u^+v^-$  cells are not allowed to have a single  $u^+v^-$  neighbor. On an ideal 2D hexagonal grid which is periodically connected on all ends, 1/3 of all cells could have  $u^+v^-$  fate and be arranged such that they are only neighboring  $u^-v^+$  cells. Cells in the simulated organoids vary in the number of their neighbors (Tab. 6), such that the exact height of the jump cannot exactly be determined beforehand.

		2D	3D
Neighbors	min	1	1
	mean	5.65	12.80
	max	10	23

Table 6: Minimum, mean and maximum number of neighbors appearing in the 2000 2D and 2000 3D organoids simulated for figure 33.

In conclusion, the spatial arrangement of the cells places a restriction on the possibly achievable cell type proportions.

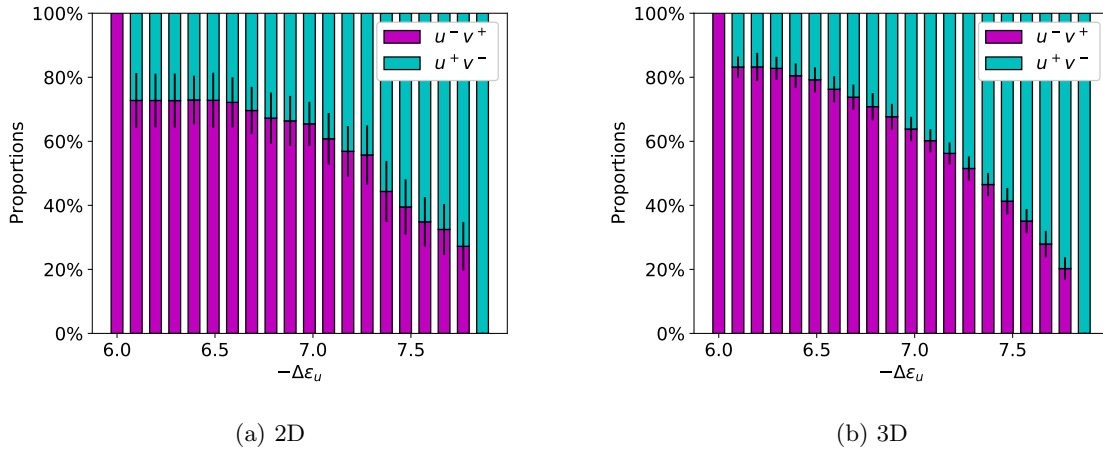


Figure 33: Range of cell type proportions for 2D (a) and 3D (b) simulations by varying  $-\Delta\epsilon_u$  over the stability interval (6,7.87). For each bar, the mean proportions of 100 different organoid simulations were used. Error bars indicate 10 times the standard deviation of the corresponding proportions. For 2D and 3D, the same evolution time of  $T = 235$  was used to generate organoids with  $315 \pm 45$  cells in 2D and  $1021 \pm 148$  cells in 3D.

#### 4.3.4 Conclusion

Nearest neighbor signals are widely used in simulations to include the effects of cell-cell communication in models about cell fate specifications [26, 27, 28]. For a signal that traverses mainly from one cell to its neighbors, this is a reasonable assumption that leads to interesting results. In the model presented here, the signal is synonymous to the main catalyst that drives pattern formation in the organoids. It was shown, that a signal that only influences nearest neighbor cells ultimately leads to a pattern which is best explained by cells of one type trying to avoid adjacency to cells of the same type. On a grid consisting of equal squares, this alternating appearance of cells resembles the patterning of a checkerboard/chessboard. Hence, the term "checkerboard pattern" is used to describe patterns of this type even on irregular grids like the 2D and 3D simulations of the organoids presented in this study. In total, the checkerboard patterns replicated in this section, resemble the findings of different studies using NN signaling [26, 27, 31]. The quantification of the patterns revealed large PCF values for direct neighbors as well as low values for Moran's  $I$ . Both are a sign for a strong negative auto correlation. The remaining distances in the



PCF highlight that there is no relation to cells of distances beyond 1, as desired from a model that only includes NN signaling. Moran's  $I$  shows a clear difference between an ideal  $8 \times 8$  pattern, simulated 2D organoids and 3D organoids. This implies, the lower bound of  $I$  is shifted upwards depending on the number of neighbors. This has to be taken into consideration for the remainder of this study.

The correct cell type proportions of PrE and Epi cells have been described as a crucial part in development of the mouse embryo with respective proportions of  $50.94\% \pm 13.42\%$  Epi cells and  $29.63\% \pm 18.06\%$  PrE cells [48, 30]. Luckily, the embryo is reliably able to generate the necessary cell type proportions. The proportions have been analyzed before during different stages in development [23]. Uncertainties in the experimental data, such as the N+G+ as well as N-G- cells that cannot be further classified, require the model to be flexibly changed to a different ratio of the two required cell types. This is where the extensive stability analysis of (M3) paid off. Varying only a single parameter of the system, enables a full range of different proportions of  $u^+v^-$  and  $u^-v^+$  cells. The proportions of  $u^+v^-$  increase monotonously with parameter  $-\Delta\varepsilon_u$ . At the same time, the proportions vary only little for organoids of slightly different shapes and sizes. The 2D and 3D simulations revealed that some proportions are impossible to attain for this model, such that the proportions of one cell type exhibit a jump of 27% at both ends of the parameter range in 2D and 17% to 20% in 3D. The subsequent analytical investigation revealed that this model dictates an exact number of  $u^+v^-$  cells that are allowed to be adjacent to another  $u^+v^-$  cell. This leads to the question of whether something similar can be demonstrated experimentally by surrounding single cells with cells in different proportions of respective precursor cells.

In summary, the model in this form provides an advantageous alternative to existing models, which is characterized by the fact that the proportions of the model are easily controlled by a single parameter.

## 4.4 Distance-based global signaling enables a transition from checkerboard to engulfing patterns

### 4.4.1 From checkerboard to engulfing patterns

Aside from checkerboard patterns, engulfing patterns are also prominent in biological systems. These are mostly believed to arise through cell sorting due to differential adhesion [69, 70]. In this section, it will be demonstrated that for the distance-based signal (165), the choice of the signal dispersion  $q$  decides between checkerboard and engulfing pattern. A systematic variation of simulations with different proportions, i.e.  $-\Delta\varepsilon_u$ , and signal dispersions  $q$  was performed (Fig. 34). For a visual comparison, the same 2D organoid as in section 4.3.2 is used here as well. For  $q = 0.1$ , the patterns mostly resemble the checkerboard patterns from the NN signaling with slightly more  $u^+v^-$  cells found at the boundary of the organoid. This occurs due to the change in normalization. In the NN case, the signal was averaged over the number of neighbors (cf. (163)). This was replaced by a normalization that is equal for every cell (cf. (165)). Cells at the boundary will usually have less direct neighbors, leading to possible less received signal values. This in turn increases the likelihood for cells to adopt the  $u^+v^-$  fate. For  $q = 0.5$ ,  $u^+v^-$  cells increasingly arise near the boundary. At the same time, smaller clusters of  $u^-v^+$  cells appear in the bulk area. For  $q = 0.9$ , the signal disperses strongly enough to lead to a signal accumulation in the center of the organoid. Therefore, the organoids become completely engulfed by  $u^+v^-$  cells with only a few of them appearing in the center. The increase of  $-\Delta\varepsilon_u$  tilts the cell type proportions in favor of  $u^+v^-$  cells, yet the global pattern remains the same for every  $q$ . For a more specific interpretation of the patterns, ten different 2D organoids with  $q \in 0, 0.1, \dots, 0.9$  were constructed specifically to have a cell ratio of 1 : 1. This was realized by using a bisection on the stability interval  $(\Delta\varepsilon_{\min}, \Delta\varepsilon_{\max}) = (6, 7.87)$ . This routine starts with a single simulation for

$$-\Delta\varepsilon_u = \frac{\Delta\varepsilon_{\min} + \Delta\varepsilon_{\max}}{2}. \quad (238)$$

If this yields a cell type ratio of 1, the bisection terminates. For a ratio  $|T^u|/|T^v| < 1$ , the lower bound  $\Delta\varepsilon_{\min}$  is replaced by  $-\Delta\varepsilon_u$  and (238) repeats. Likewise, if  $|T^u|/|T^v| > 1$ , the upper bound  $\Delta\varepsilon_{\max}$  is replaced by  $-\Delta\varepsilon_u$ . The resulting PCFs allow the quantification of the patterns without possible interference due to varying cell type proportions (Fig. 35). PCF  $\rho_u$  can be divided into three different parts. In the low distance region, low values of  $q$  show similarities to the PCFs of the NN signaling, i.e. low values for  $\rho_u(1)$  and slightly increased values for  $\rho_u(2)$ . This was already identified as a typical behavior of a checkerboard pattern. As  $q$  increases  $\rho_u(1)$  increases due to the formation of several layers of  $u^+v^-$  cells at the boundary of the organoid. Thus, adjacency can no longer be avoided. At intermediate distances, the increase of  $q$  leads to lower occurrences of  $u^+v^-$ . This shows that there is some kind of radial separation of the two cell types. At large cell distances, this separation of cells becomes very clear



as  $\rho_u$  tremendously increase with  $q$ . This means that the organoid consists of a few layers completely consisting of  $u^+v^-$  cells surrounding the remaining cells. In contrast to the averaged NN case, the PCFs even for lower values of  $q$  consistently show signs of  $u^+v^-$  accumulating at the boundary. Thus, even for the low number of cell pairs found for long distances, the results are robust with respect to changes in  $q$ . The drops observed at distance 19 for  $q$  ranging from 0.1 and 0.6, indicate that the model did not form a full second layer of  $u^+v^-$  cells.

The PCFs  $\rho_v$  show the transition towards the engulfment from the perspective of  $u^-v^+$  cells. The low distance regions exhibit an increase of  $\rho_v$ , while  $q$  increases, which is indicative for the formation of a larger cluster of  $u^-v^+$  cells. At the same time,  $\rho_v$  decreases at larger distances with  $q$  which again describes the radial separation of  $u^+v^-$  and  $u^-v^+$  cells, with  $u^-v^+$  residing in the center of the organoid.

Moran's  $I$  was used to compare the spatial auto-correlations for different  $q$ . For  $q = 0.1$ , values close to those of the NN signaling can be found in 2D and 3D. The increase of  $q$  leads to an overall increase in  $I$ . At  $q = 0.5$ ,  $I$  adopts values close to 0 in 2D, which would resemble that of randomly distributed cells. In 3D, cell type proportions of  $u^+v^-$  have a greater influence on  $I$  than in 2D, such that an increase is observed up to a maximum of around 0.4, before it decreases again. The same behavior can be seen for  $q = 0.9$  with larger values than for  $q = 0.5$ , in both 2D and 3D. The increase of  $I$  can be attributed to the radial separation of  $u^+v^-$  and  $u^-v^+$  cells. The better the separation, the less contact cells of different types will have. At the same time, increasing the proportions of  $u^+v^-$  decreases the radius of the ring that separates the two cell types, thus leading to a reduced number of contacts.

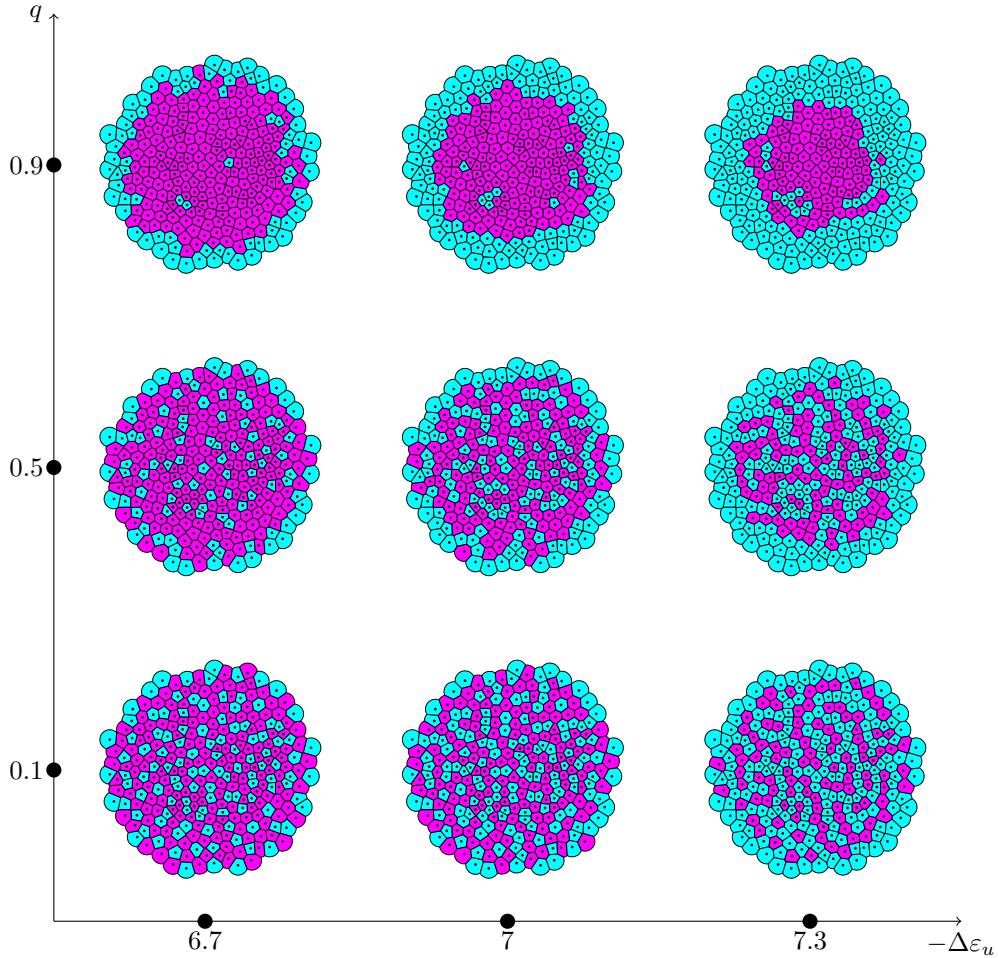


Figure 34: Simulated patterns on a 2D organoid with 302 cells. Colors depict the values of  $v_i$  in steady state between 0 and 0.1. Large values of  $v_i$  correspond to low values in  $u_i$  and vice-versa. Thus, cyan and magenta represent  $u^+v^-$  and  $u^-v^+$  cells, respectively. From left to right,  $-\Delta\epsilon_u$  increases. From bottom to top, the dispersion  $q$  increases.

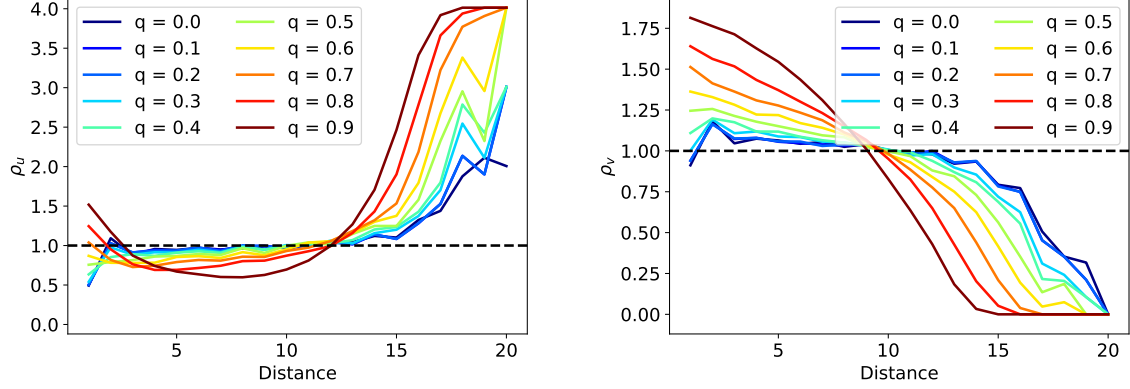


Figure 35: PCFs for  $u^+v$  cells (left) and  $u^v+$  cells (right) for dispersion parameter values  $q$  ranging from 0 to 0.9. Any PCF represents the same organoid geometry with a fixed cell type ratio  $|T^u|/|T^v| = 1$ . The horizontal and dashed black line at 1 resembles the PCF values of an ideal uniform distribution of two different cell types. Values above 1 mean that more pairs are found at the respective distance, values below mean less.

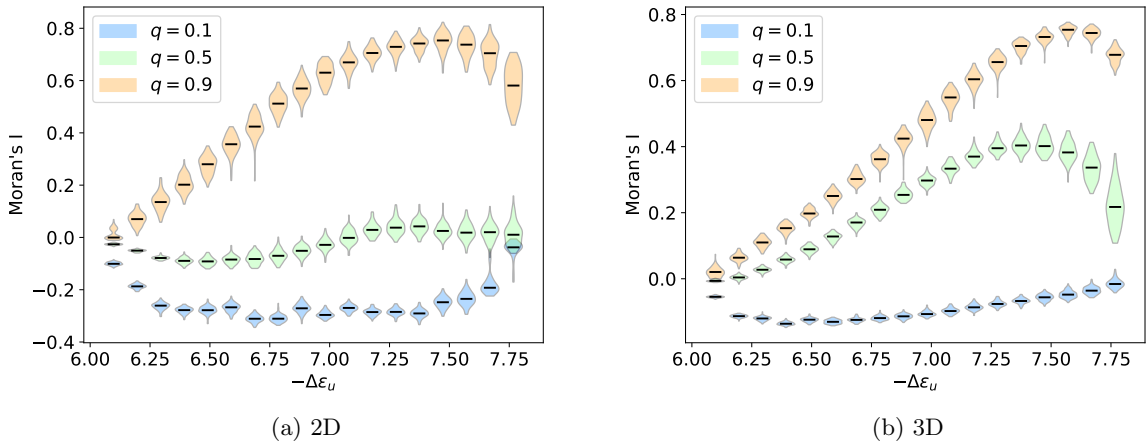


Figure 36: Combined violin plots for Moran's  $I$ . A total of 18 different values of  $\Delta\epsilon_u$  were considered with  $q \in \{0.1, 0.5, 0.9\}$ . Each violin plot consists of 100 organoids simulated to  $T = 235$  in both 2D and 3D. Respective median values are depicted as horizontal black lines.

#### 4.4.2 Cell type proportions

The ability of model (*M3*) to generate a range of different cell type proportions was already demonstrated for the NN signaling. In the case of distance-based signaling, the proportions exhibit a smoother transition from one homogeneous state to the other (Fig. 37). For  $q = 0.1$ , individual jumps in proportions can still be observed, especially in 2D. At the same time, the largest standard deviations are observed in this case, which are nevertheless still relatively small with a maximum of 2.35%. As  $q$  increases, the standard deviations decrease and the proportions approach a curve  $1 - f_u$ . The procedure, in which this function arises, is similar to the investigation of the critical signal  $s^*$  for the NN signal. In steady state, expressions of  $u_j$  are either 0 or  $r_u/\gamma_u$ . Hence, the absorbed signal (165) is only influenced by the cells in  $T^u$

$$s_i = \left( \sum_{j \in T^u} q^{d_{ij}-1} \frac{r_u}{\gamma_u} \right) / \left( \max_k \sum_{j \neq k} q^{d_{kj}-1} \right). \quad (239)$$

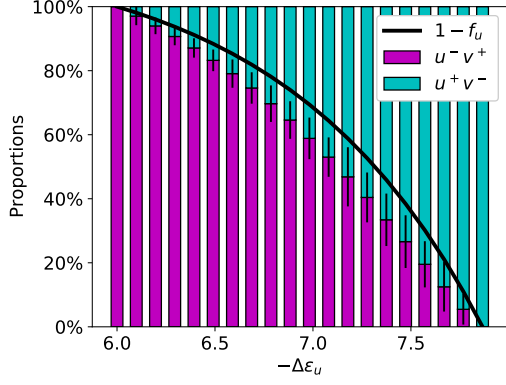
Therefore, equation (127) leads to

$$\frac{\gamma_u}{r_u} s_i = \left( \sum_{j \in T^u} q^{d_{ij}-1} \right) / \left( \max_k \sum_{j \neq k} q^{d_{kj}-1} \right) \stackrel{!}{=} \frac{\gamma_v}{r_v} \frac{\eta_u - \eta_v}{\eta_v \eta_s \eta_{vs}} =: \hat{f}_u(\eta_u), \quad \text{for } i = 1, \dots, n. \quad (240)$$

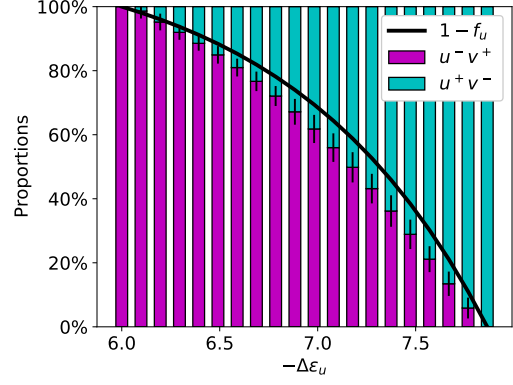
The collection of all  $u^+v^-$  cells  $T^u$  emerges from the fact, that only these cells play a role in the formation of the received signal. Although it could not be proven in this study, it is hypothesized, that cells  $T^u$  will always arrange in such a way, that (240) will approximately be fulfilled. Under this assumption,  $f_u$  describes the proportions of  $u^+v^-$  if the signal throughout the organoid was able to occupy a continuous region, rather than individual cells. For larger signal dispersions, more cells are relevant for the cell fate decision of a single cell. Hence, more cells are involved in trying to match  $s_i = \hat{f}_u$ . The function  $f_u$  used in figure 37 is then defined as  $\hat{f}_u(\eta_u) = f_u(-\Delta\varepsilon_u)$ .

#### 4.4.3 Patterns match experimental data at long distances

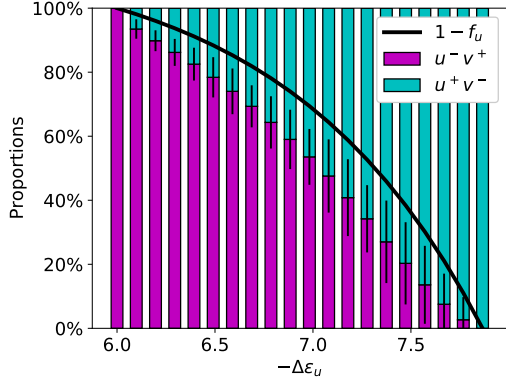
In figure 35, it was already shown that  $u^+v^-$  cells are the ones engulfing  $u^-v^+$  cells. However, the model itself was designed to resemble the NANOG expression levels via  $u$  and GATA6 expression levels via  $v$ . Thus, the cell fates in simulation and experiment are reversed. If the GRN is to be assumed a valid representation of the underlying mechanisms regarding the regulation of NANOG and GATA6, then the initial hypothesis that a global signal might be responsible for the sorting in the ICM organoid has been disproven. Nevertheless, the model is still capable of generating radially distributed patterns of two cell types. Hence, it is possible to match the patterns of the experiments with those of the simulations. For this purpose, simulations were performed on the geometry of the experimental data. A bisection was used on the stability interval (175) to find  $-\Delta\varepsilon_u$  such that the cell type ratio of  $u^+v^-$  and  $u^-v^+$  reflects the ratio of N-G+ and G+N- from the experimental data, up to a tolerance of 10%. For this comparison, six different organoids have been chosen to represent a variety of different cases. Organoids with ID 2, 32 and 36 represent three cases of partly successful matches between simulation and experiment (Fig. 38). For  $\rho_u$ , large values like  $q = 0.9$  provides the best match at low distances, regardless of the organoid. At long distances however, organoid 2 is best matched by lower values  $q \in (0.2, 0.4)$  while organoid 32 best fits with intermediate values  $q \in (0.4, 0.6)$ . Finally, large values  $q = 0.8$  provide a surprisingly good fit for organoid 36 on both ends. Simultaneously, the same values of  $q$  lead to  $\rho_v$  that best describe the trends of N+G- cells. For this organoid, there is no distinction to be made between low and long distances. One reason why the characterization of these three organoids works so well, are the cell type proportions of N+G- and N-G+ cells. They were chosen specifically, because the ratio of these cells lies between 0.5 and 2. Furthermore, the proportions of DN and DP combined were below 25%. If these restriction become severely disturbed, the quantification becomes unreliable (Fig. 39). Organoid 1 exhibits a tremendous amount of DN or DP cells. Therefore, the PCF envelope is extending both below and above 1. Extreme case scenarios such as an engulfing pattern, are highly unlikely to be represented with just 1000 samples. Therefore, neither  $\rho_u$  nor  $\rho_v$  seem to fit for any  $q$ . A different scenario is reached in organoid 12, where there are nine times more N+G- cells than N-G+ cells. Small changes in cell fates of certain distances have a tremendous impact on the PCFs for N-G+. This leads to a massive overestimation by the model and therefore  $\rho_u$  which is designed to lead to an engulfing pattern. Lastly, organoid 39 is another example of an organoid that is hard to match with the computational model. This time, proportions are not as exaggerated as in the previous two examples. For  $q = 0.2$ , a reasonable match was found, yet the PCF envelopes extending both below and above 1 make accepting this match a decision that was abstained from. The same organoids were used to calculate the range of possible values for Moran's  $I$ . These values are compared to the simulated organoids with different  $q$  (Fig. 40). This reveals that in two of the three



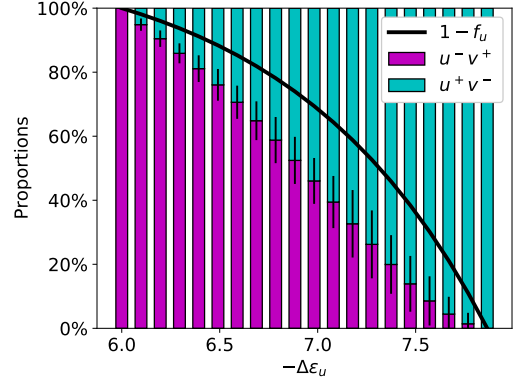
(a) 2D,  $q = 0.9$



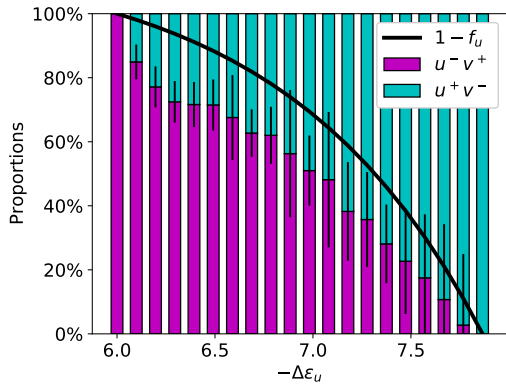
(b) 3D,  $q = 0.9$



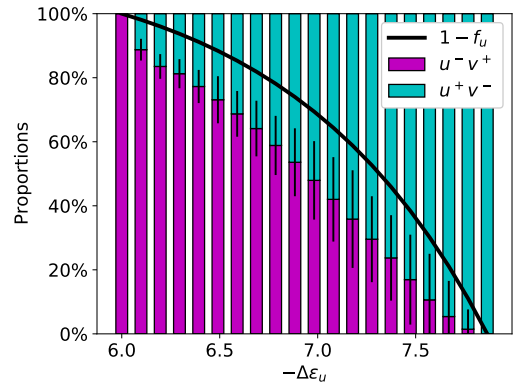
(c) 2D,  $q = 0.5$



(d) 3D,  $q = 0.5$



(e) 2D,  $q = 0.1$



(f) 3D,  $q = 0.1$

Figure 37: Cell type proportions for 2D (left) and 3D (right) simulations given the distance-based signaling. In each graph,  $-\Delta\epsilon_u$  varies over the stability interval  $[6, 7.87]$ . Bars represent the mean proportions of 100 different organoids with  $T = 235$ . Error bars indicate 10 times the standard deviation of the corresponding proportions. From bottom to top,  $q$  increases from 0.1 over 0.5 to 0.9.

positive examples, there is a major mismatch between the values of  $q$  that best fit the PCFs and the ones that best fit for  $I$ . The experimental values tend to be greater, suggesting that the model is incapable of providing this amount of spatial auto-correlation.

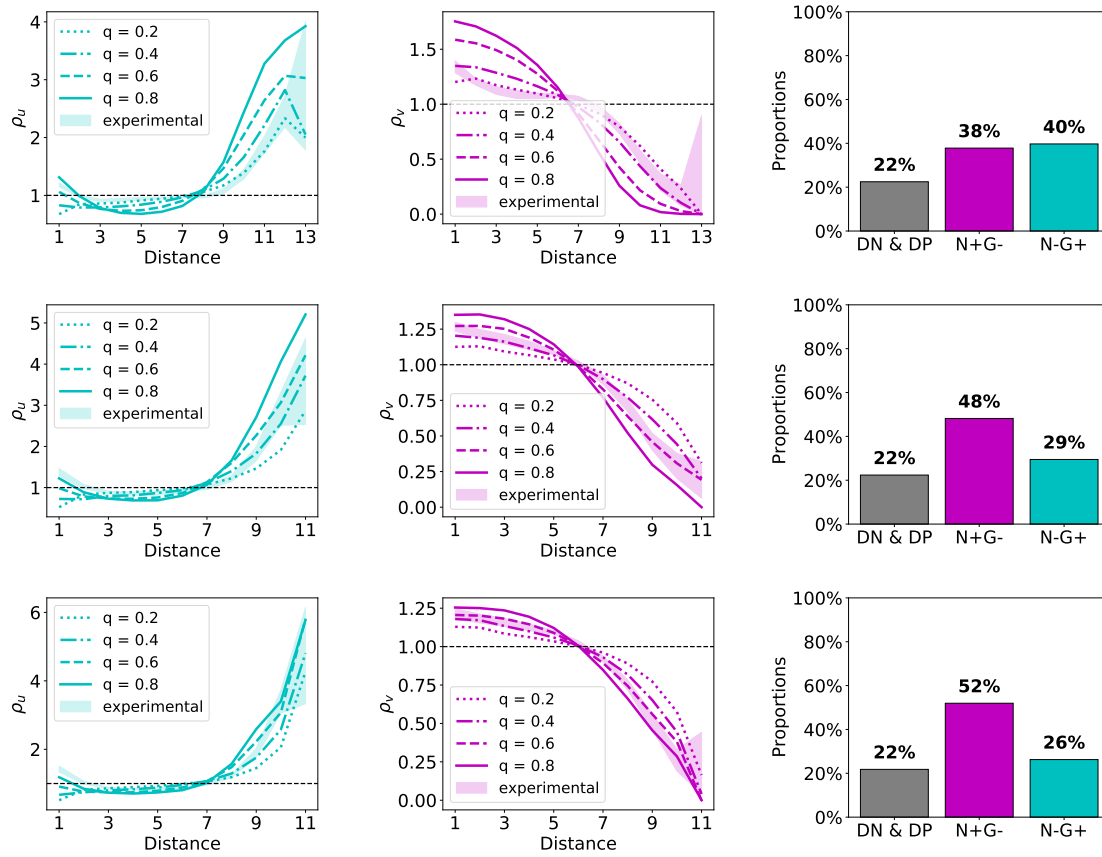


Figure 38: Three positive examples of the connection between simulation and experimental data. In each row, we observe the data of a single ICM organoid from [23]. From left to right, the PCFs  $\rho_u$  and  $\rho_v$  and the cell type proportions are visualized. The PCF plots include PCFs for different dispersion parameters  $q$  as well as the PCF envelope generated by 1000 different samples of randomly choosing cell fates for DN and DP cells. Corresponding organoid IDs from top to bottom are 2, 32 and 36.

#### 4.4.4 Conclusion

One possibility of global cell-cell communication is realized by using a signal that disperses throughout the tissue based on the distance traveled. This enabled a range of different patterns which is best described by a transition from a checkerboard to an engulfing pattern. This transition was found by increasing the signal dispersion parameter  $q$  in the interval  $[0, 1)$ . The PCFs made it possible to capture this transition quantitatively showing large increases and decreases for  $\rho_u$  and  $\rho_v$  respectively in regions of low and high distances. In addition to that, Moran's  $I$  shows that for the distance based-signaling, the spatial auto-correlation of the cells changes with respect to the cell type proportions for larger values of  $q$ . This is yet another sign, that the global patterns, generated with the use of the distance-based signal, cannot easily be characterized by their local cell neighborhood.

It could be shown that the use of a signal that incorporates more than just the neighboring cells, yields an improved coverage of all possible cell type proportions. This coverage is further improved by the increase of  $q$ . Hints were found, that for increasing  $q$ , cell type proportions approach the theoretic cell type proportions of what is hypothesized to be the result of a continuous distribution of  $u$  and  $v$ . While not relevant for the remainder of this study, this hypothesis is still an open question and provides space for further research.

The comparison with the experimental data has revealed, that the initial hypothesis of this thesis must be false. Differently dispersing signals are not responsible for the formation of engulfing patterns in mouse ICM organoids. However, the model has still proven to generate patterns of this type, with the

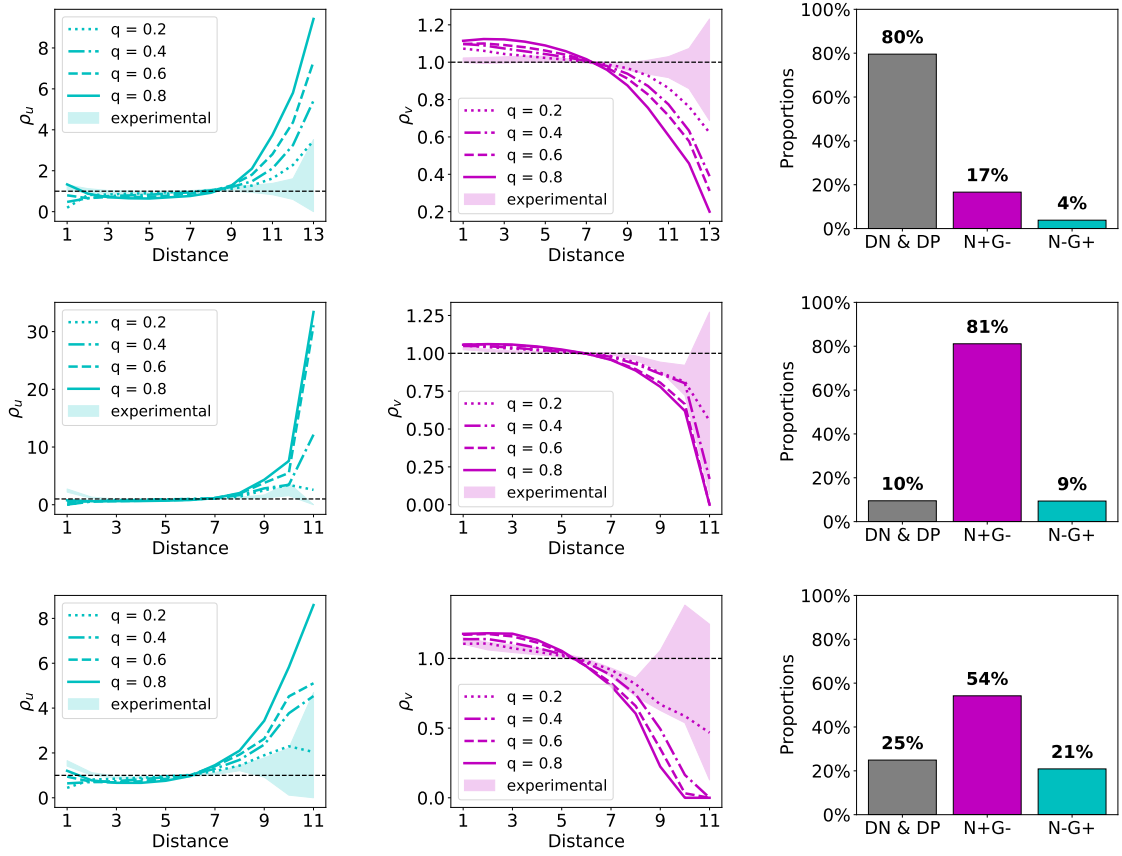


Figure 39: Three negative examples of the connection between simulation and experimental data. In each row, we observe the data of a single ICM organoid from [23]. From left to right, the PCFs  $\rho_u$  and  $\rho_v$  and the cell type proportions are visualized. The PCF plots include PCFs for different dispersion parameters  $q$  as well as the PCF envelope generated by 1000 different samples of randomly choosing cell fates for DN and DP cells. Corresponding organoid IDs from top to bottom are 1, 12 and 39.

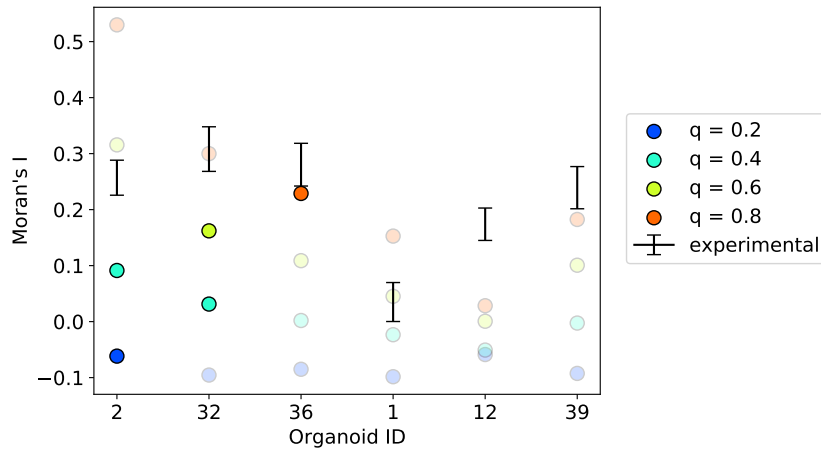


Figure 40: Moran's  $I$  resulting from the simulations compared to the range of experimental values gained by 1000 samples of randomly choosing cell fates for DN and DP cells. The three examples on the left represent the positive examples, where the PCF looked promising, whereas the three on the right are the negative ones. Markers with higher opacity highlight the values of  $q$  that were the best matches for the PCFs.

cell fates being reversed. As such it could be shown, that a single value for  $q$  is not able to describe all of the patterns in the data. If  $q$  is taken as a measure for the radial distribution of the cells, this means that some organoids show signs of a near perfect engulfing, while others tend more towards checkerboard with large amount of N-G+ cells at the boundary. Moreover, the PCFs for distances 1 as well as Moran's  $I$  have shown, that the computational model underestimates the experimental findings. The presence of clusters of one cell type would increase  $I$ , leading to the conclusion, that the model lacks the capability of cluster formation. This gives rise to a new hypothesis that cell division is needed to generate accurate representations of such patterns.

In total, the results show that the formation of checkerboard and engulfing pattern could be unified under the notion of a differently dispersing signal. At the same time, nearly any cell type proportions can be realized with the use of this signal. While the original hypothesis was proven to be false, the model was found to be lacking the ability to form clusters. This motivates extending the model by incorporating cell division.

## 4.5 Cell division leads to clustering of equal cell types

The PCF values  $\rho_u(1)$  as well as  $\rho_v(1)$  of the static systems are much lower than those observed in the experimental data. This suggests that in reality, cells of equal type are much more likely to cluster together. Cell division has been demonstrated to lead to a clustering of this type [43]. Therefore, it is only natural to examine all model parts combined together in parallel. This means that the organoid will no longer be simulated separately from the transcriptional regulation. In the model, this is realized by symmetric division. This means that after cell division,  $u$  and  $v$  of the mother cell are split evenly among the daughter cells. This section first addresses handling the proper connection of the two timescales (transcriptional regulation model and organoid growth model) involved in our system. This is followed by a direct comparison of the static and dynamic system, before quantifying the effects of cell division.

### 4.5.1 Differentiation time has little influence on the resulting pattern

So far, the time needed for differentiation has been chosen separately from the organoid growth time. Now that the two parts of the model are combined into one overall model that is running in parallel, the time scale of one model must be adjusted to that of the other. The organoid growth model was already fitted to the cell numbers found in the experimental data from [23]. Therefore, it suffices to adjust the time scale of the transcriptional regulation model accordingly. To this end, the production and decay parameters will simply be multiplied by a factor  $\tau$

$$r_u \rightarrow \tau r_u \qquad \gamma_u \rightarrow \tau \gamma_u, \qquad (241)$$

$$r_v \rightarrow \tau r_v \qquad \gamma_v \rightarrow \tau \gamma_v. \qquad (242)$$

This way,  $\tau$  is used to change the time it takes for the transcriptional regulation model to reach a certain point, while the organoid growth remains the same. More precisely, for  $\tau > 1$ , the model gets faster, whereas for  $\tau < 1$ , it gets slower. Simulation results using NN signaling (Fig. 41) reveal that for extremely low values  $\tau = 0.01$ , no mixture of two cell types is achieved. The reason for this is that the cells divide faster than they are able to differentiate. Every cell division leads to the amount of  $u$  and  $v$  being split in half, while not being able to grow fast enough to actually show any signs of differential expression. For  $\tau = 0.05$ , a mixture of different cell types  $u^+v^-$  and  $u^-v^+$  is found, where a few of the cells are not yet fully committed to a specific cell fate. As  $\tau$  increases further, the number of undecided cells decreases. Qualitatively, the patterns resemble the checkerboard pattern, with some clusters of identical cells. A quantitative analysis including the comparison to the static case will be discussed in the following chapter. For the distance-based signaling (165), a similar trend is observed (Fig. 42). A signal dispersion value  $q = 0.5$  was chosen as a representative value for distance-based simulations, with other values differing in principle only in the nature of the pattern. For  $\tau = 0.01$ , large amounts of cells are not committed to a specific cell fate. This time however, there are signs of  $v$  decreasing as the distance towards the center of the organoid increases. The increase of  $\tau$  again leads to the decrease of not yet fully committed cells up to  $\tau = 1$ , where almost every cell adopts a specific fate. The resulting patterns show a radial separation of  $u^+v^-$  and  $u^-v^+$  cells, which from our qualitative observation differ again in the amount of cluster formation from the static case. In the cases of both NN and distance-based signaling, there is no evidence that the patterns differ in their respective nature with regards to  $\tau$ , as long as signs of differentiation can be observed. Hence, a value of  $\tau = 1$  was perceived as a justified choice for further simulations.

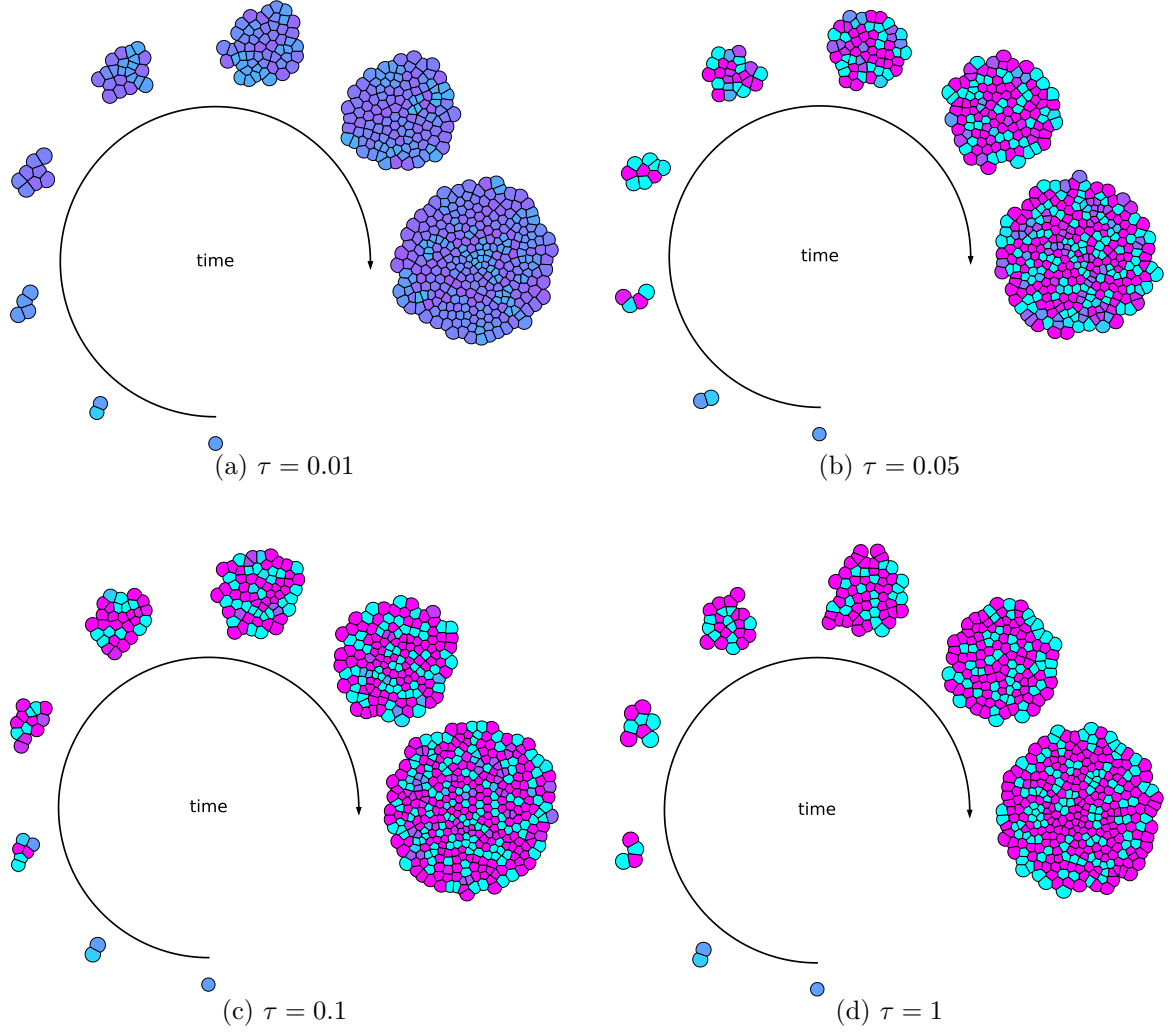


Figure 41: Temporal evolution of four simulated organoids using NN signaling. From (a) to (d), the values of  $\tau$  increase. Individual images show snapshots of the organoids at eight different points in time, that are equidistantly distributed until the final time  $T = 235$  is reached. The proportions are the result of  $-\Delta\varepsilon_u = 7$ . Color gradient from cyan to magenta is fixed to values between 0 and 0.1, where cyan represents low values of  $v$ , whereas magenta represents large values of  $v$ .



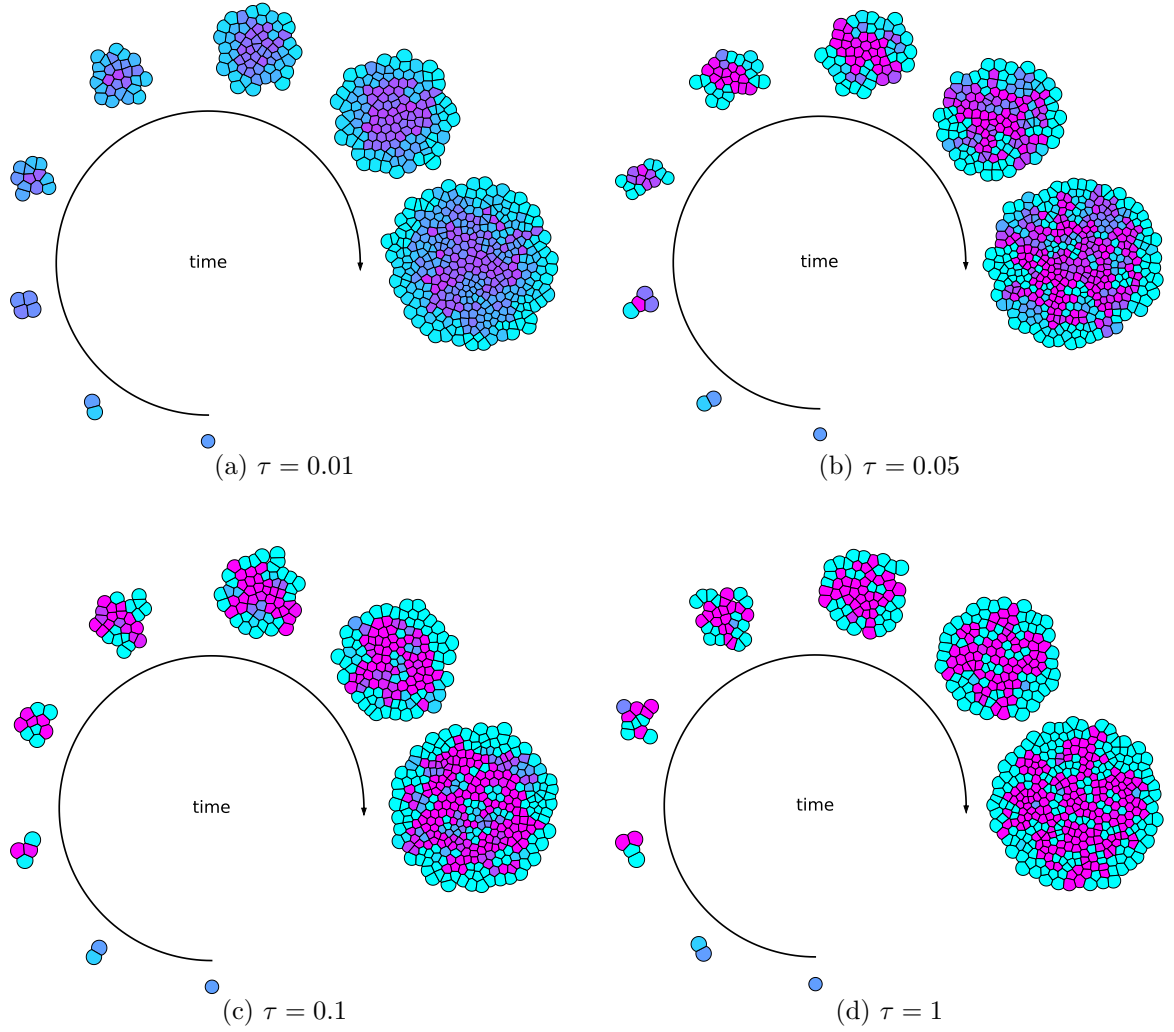


Figure 42: Temporal evolution of four simulated organoids using distance-based signaling. From (a) to (d), the values of  $\tau$  increase. Individual images show snapshots of the organoids at eight different points in time, that are equidistantly distributed until the final time  $T = 235$  is reached. The proportions are the result of  $-\Delta\varepsilon_u = 7$ . Color gradient from cyan to magenta is fixed to values between 0 and 0.1, where cyan represents low values of  $v$ , whereas magenta represents large values of  $v$ .

#### 4.5.2 Dynamic vs. static – Cell division leads to increased cluster formation

In this section, the patterns formed via dynamic and static differentiation are quantitatively compared. Four exemplary organoids were simulated dynamically for different types of signaling. In addition, the positions and radii from the final time step were used to run static simulations for cell differentiation (Fig. 43). At first glance, it is apparent that the dynamic and static differentiation patterns can be visually distinguished from each other. The averaged NN signal results again in its characteristic checkerboard pattern in the static case. In the dynamic case, instead of single  $u^+v^-$  avoiding adjacency to cells of the same type, smaller clusters of  $u^+v^-$  cells become more prominent throughout the organoid. This is reflected by  $\rho_u(1)$  values much closer to 1. Likewise,  $\rho_u(2)$  is decreased in the dynamic case. For long distances,  $\rho_u$  and  $\rho_v$  are mostly close to 1 with some outliers at extreme distances. In total, this means that the perfect checkerboard pattern gets lost in the dynamic case, resulting in a more clustered or random looking pattern. Similar behaviors are found for the dispersive signals. Values  $\rho_u(1)$  consistently increase for  $q \in \{0.1, 0.5, 0.9\}$ . For  $q = 0.1$  and  $q = 0.5$ , they even switch from below 1 to above 1 hinting at increased cluster formation. For  $\rho_v(1)$  however, there is no consistency to be found making it difficult to quantify the extent of the clustering using the PCFs alone. The overall pattern of radially distributed cells is conserved to some extent. The PCFs  $\rho_u$  and  $\rho_v$  still adopt large and low values respectively at long distances. However, by visual observation, it becomes obvious, that an ideal radial segregation of  $u^+v^-$  and  $u^-v^+$  cells is no longer guaranteed.

At the level of cluster formation, the exact difference between static and dynamic simulations is quantified using Moran's  $I$  (Fig. 44). In 2D, each dynamically simulated organoid exhibits a value of  $I$  that is close to or above 0. What initially could be classified as checkerboard, i.e. NN or  $q = 0.1$ , is now resembling something that is closer to a random distribution of cells. This trend continues for  $q = 0.5$ , meaning that  $I$  increases even further. Only for  $q = 0.9$  there is a point for the energy difference  $-\Delta\epsilon_u$ , where the static case shows larger values of  $I$  compared to the dynamic one. For the radially segregated pattern, increasing  $-\Delta\epsilon_u$  increases the amounts of  $u^+v^-$  layers surrounding the  $u^-v^+$  cluster in the center.  $I$  adopts the largest values for two clusters of different cell types with as few as possible points of contact. Therefore, the static case has an advantage over the dynamic one, where the stochastic nature of cell division is able to separate large clusters into two or more smaller clusters. Still the values of  $I$  are found to be very large, meaning that cluster formation is working as expected. Compared to the static simulation, the variability of  $I$  has also increased. This again is attributed to the stochasticity of the cell division.

In total, the formation of clusters is observed at every level. Qualitative observations show that dynamic simulations form larger clusters of equal cell types compared to the static simulations. The quantitative analysis of Moran's  $I$  confirms this observation. The PCFs show that globally, the patterns remain mostly similar. However, the perfect formation of a checkerboard or engulfing is no longer guaranteed.

#### 4.5.3 Number of cell divisions increase cluster sizes

The experimental data were obtained from organoids grown from 200 cells [23]. Assuming these cells to be identical in their potential to differentiate, it makes sense to investigate the cluster formation based on varying initial cell numbers. Therefore, transcriptional regulation is first initiated at specific points in time to resemble a certain number of cell divisions. The estimated number of cells follow equation (191)

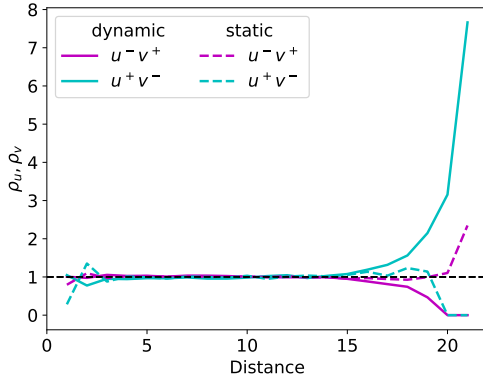
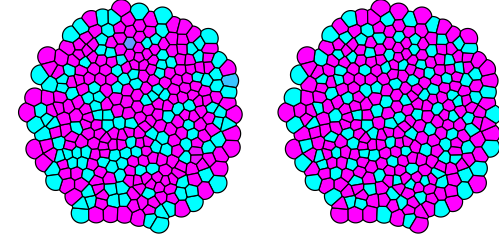
$$n(t) = e^{\Lambda t} n(0), \quad \Lambda \in \{\Lambda_{2D}, \Lambda_{3D}\} \quad (243)$$

i.e. the time it takes to double the number of cells and thus the approximate time for one cell division is

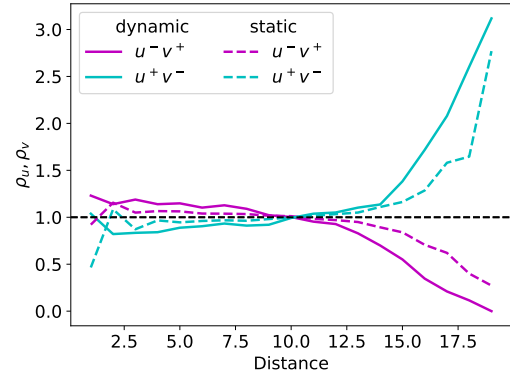
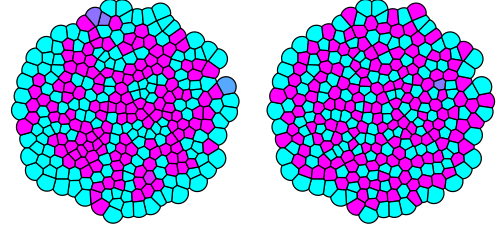
$$t_{\text{div}} = \frac{\ln(2)}{\Lambda}. \quad (244)$$

The different times chosen for the transcription to start are  $T - k_{\text{div}} t_{\text{div}}$  with  $T = 235$ . This way,  $k_{\text{div}}$  denotes the average number of cell divisions after the onset of transcription. For visual comparison, four different organoids were simulated using NN signaling and  $k_{\text{div}} \in \{1, 2, 3, 4\}$  (Fig. 45). Each of the respective organoids ends up exhibiting a checkerboard pattern with several smaller clusters of  $u^+v^-$  cells distributed evenly. Visually, differences between the patterns are impossible to discern. Although one would expect  $k_{\text{div}}$  to lead to increased clustering in the organoid, this cannot be observed in this tiny sample of four organoids. This is also seen by Moran's  $I$  showing no unique characteristics.

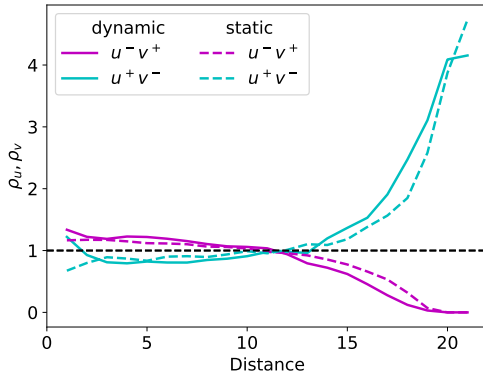
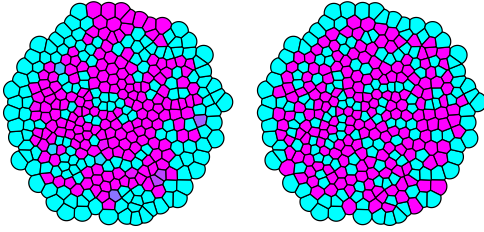
Repeating the same procedure using a distance-based signal with  $q = 0.5$ , the difference between patterns of increasing  $k_{\text{div}}$  is again difficult to observe visually (Fig. 46). All four organoids show the



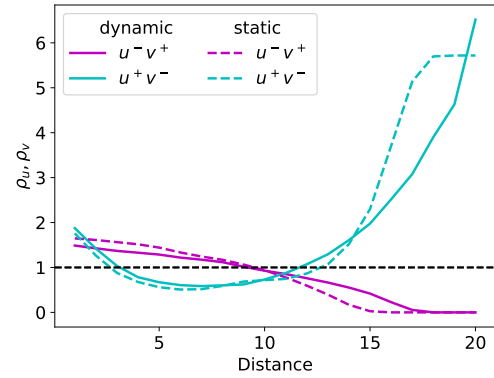
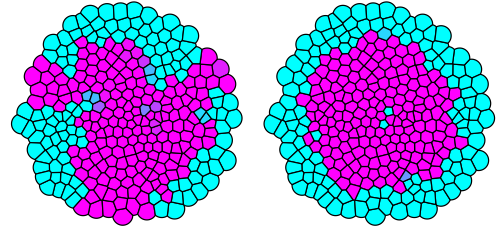
(a) Nearest neighbor signal



(b) Distance-based signal  $q = 0.1$

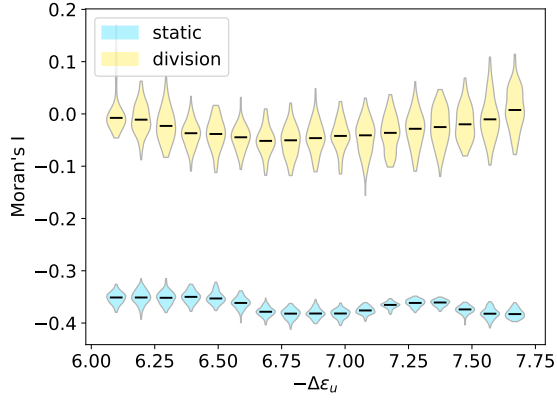


(c) Distance-based signal  $q = 0.5$

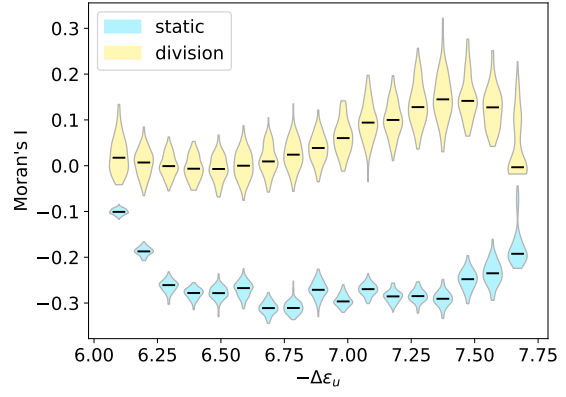


(d) Distance-based signal  $q = 0.9$

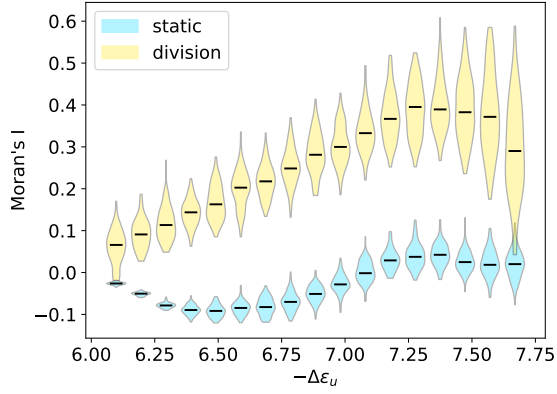
Figure 43: Comparison of pattern formation between the dynamic and static simulations. Four exemplary 2D organoids with different signaling were simulated dynamically until  $T = 235$ . Their final geometries were used for the static simulations. In each sub-figure, the left organoid represents the dynamic case, whereas the right one is the static one. Below the organoids, the corresponding PCFs are depicted.



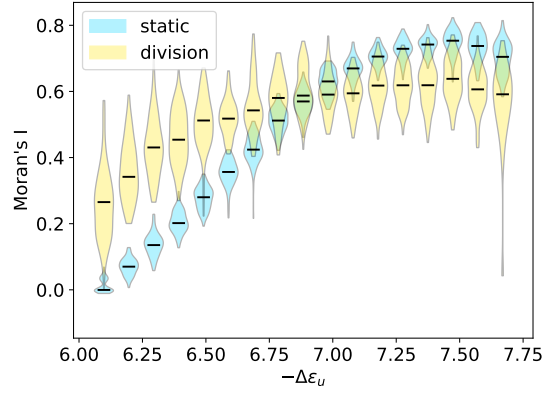
(a) NN signal



(b) Distance-based signal  $q = 0.1$



(c) Distance-based signal  $q = 0.5$



(d) Distance-based signal  $q = 0.9$

Figure 44: Violin plots showing the influence of cell division during differentiation on Moran's  $I$  compared to cell differentiation on a static organoid. Each violin represents 100 different 2D organoids, developed over a time of  $T = 235$ . Respective median values are given by black bars. Individual plots depict the different types of signaling.

engulfing of  $u^-v^+$  cells by  $u^+v^-$  cells. In addition to that, clusters of  $u^+v^-$  cells are found distributed throughout the organoid. For  $k_{\text{div}} = 8$  and  $k_{\text{div}} = 6$ , these clusters look more clumped together compared to  $k_{\text{div}} = 4$  and  $k_{\text{div}} = 2$ . Quantitatively, this exact behavior is captured by the decrease of  $I$  with decreasing  $k_{\text{div}}$ .

In order to confirm the assumption that the number of cell divisions after the onset of transcription  $k_{\text{div}}$  influences  $I$ , a broader analysis was performed. A total of 1600 organoids in 2D and 4000 organoids in 3D, evenly distributed to different  $k_{\text{div}}$  and signaling types (NN or  $q \in \{0.1, 0.5, 0.9\}$ ), were simulated. For the averaged NN as well as distance-based signaling with  $q = 0.1$ , an increase of  $I$  is observed with increasing  $k_{\text{div}}$ . Its slope decreases as  $k_{\text{div}}$  increases, suggesting that there will be some kind of maximum value for the respective type of signaling. For  $q = 0.5$ , a difference between 2D and 3D is observed. In 2D, the trend is similar to the low dispersion signals, with an increase of  $I$  with respect to  $k_{\text{div}}$ . In 3D, there are only slight changes in  $I$  suggesting that the number of cell divisions after the onset of transcription has a comparatively low impact on the size of the clusters. For  $q = 0.9$ , both 2D and 3D follow the opposite trend of a decrease in  $I$  with  $k_{\text{div}}$ . This likely occurs due to the aforementioned disruption of the near perfect engulfing pattern. It can only be assumed that  $I$  in this case also approaches some bounding value, which in this case would be a minimum. Comparing the range of  $I$  for 2D and 3D, it is seen that the low dispersion signals admit to larger values of  $I$  in 3D. At the same time, the large dispersion signals in 3D admit to lower values. Overall, this means that the total range of values covered by  $I$  is smaller in 3D compared to 2D. This is likely due to the difference in the average number of adjacent cells.

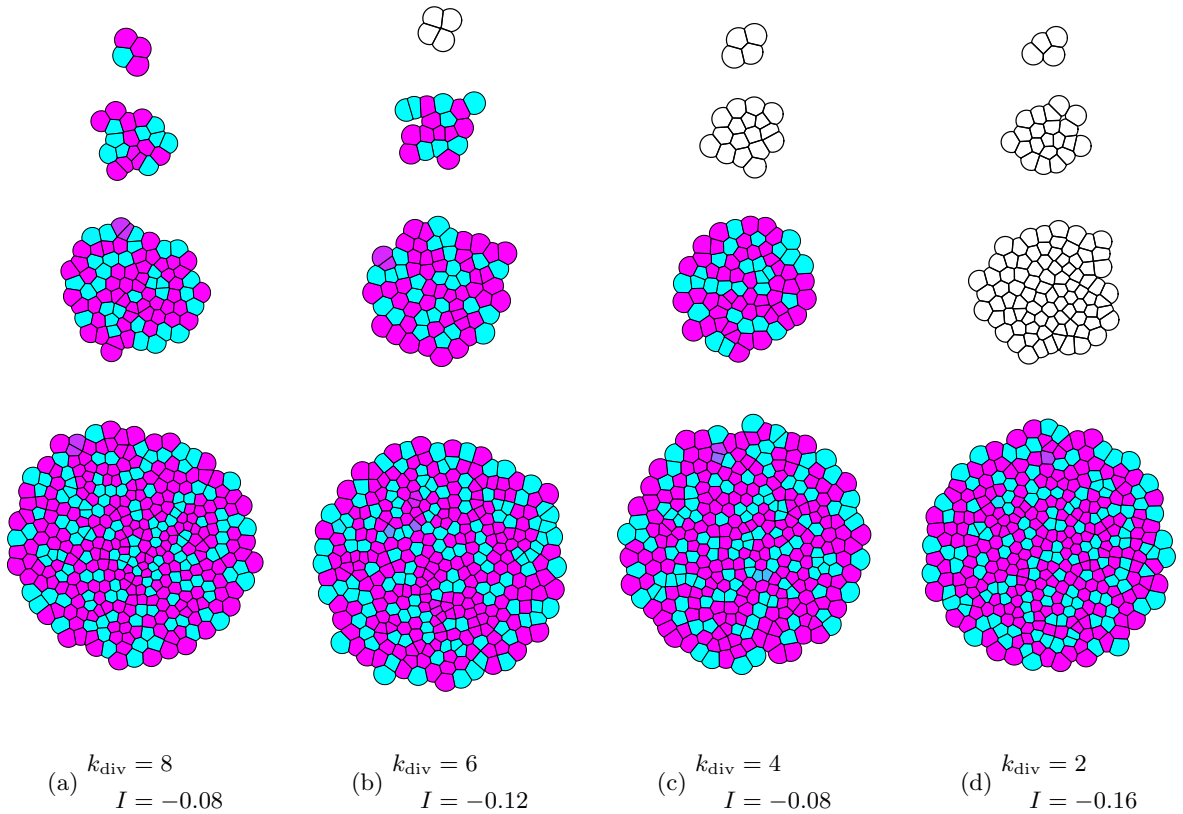


Figure 45: Evolution of four different organoids with NN signaling. From left to right, the starting time of differentiation increases with  $k_{\text{div}}$ . From top to bottom, development of the respective organoids is depicted over time. The time passed from one stage to the following stage below is  $2t_{\text{div}}$ , which represents an estimated number of two cell divisions. Blank organoids represent organoids where the transcriptional regulation has not yet been started. Corresponding values of Moran's  $I$  are given in the subcaptions.

#### 4.5.4 Conclusion

By executing organoid growth and transcriptional regulation in parallel, a whole series of investigations was motivated and performed throughout this section. These include the study of the different time scales provided by the transcriptional regulation and organoid growth models, the systematic study of

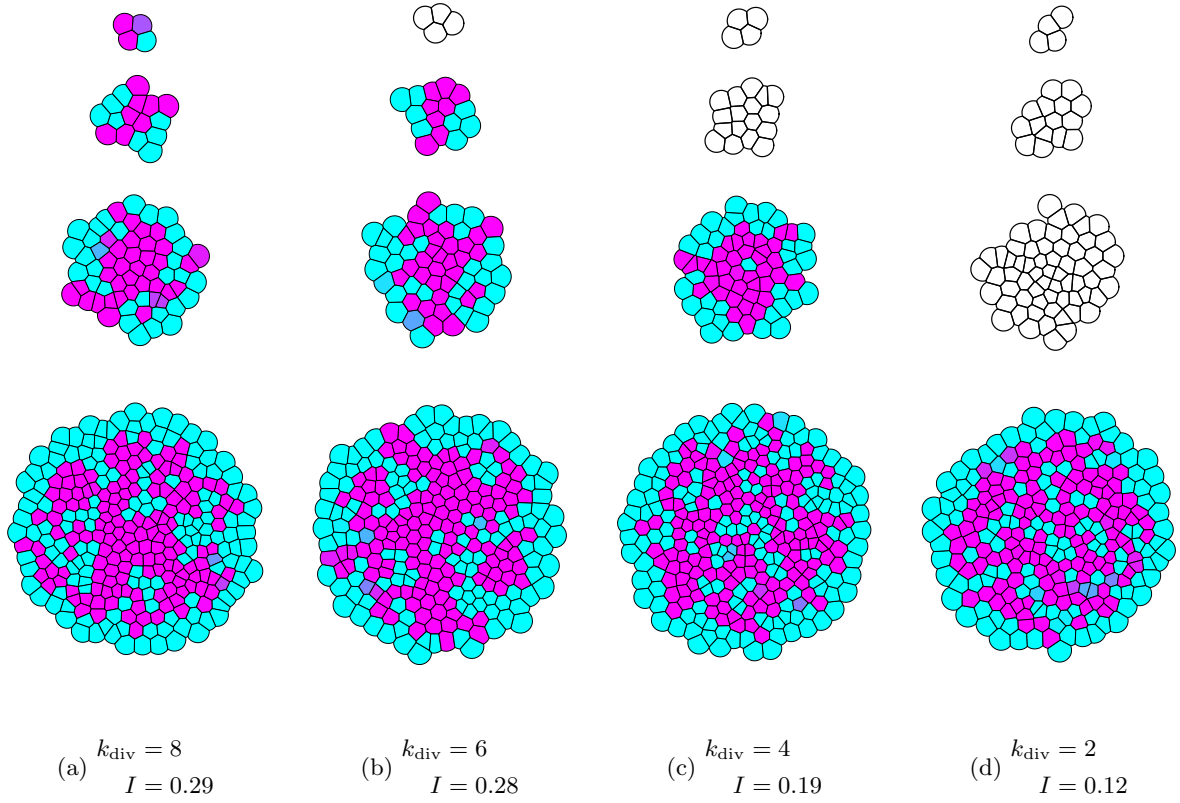


Figure 46: Evolution of four different organoids with distance-based signaling and  $q = 0.5$ . From left to right, the starting time of differentiation increases with  $k_{\text{div}}$ . From top to bottom, development of the respective organoids is depicted over time. The time passed from one stage to the following stage below is  $2t_{\text{div}}$ , which represents an estimated number of two cell divisions. Blank organoids represent organoids where the transcriptional regulation has not yet been started. Corresponding values of Moran's  $I$  are given in the subcaptions.

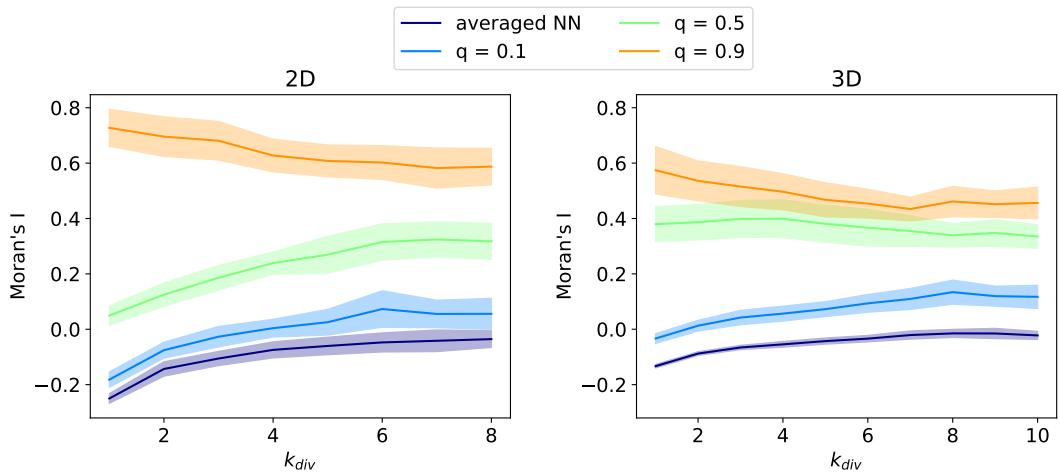


Figure 47: Change of Moran's  $I$  with respect to the estimated number of divisions the 2D (left) and 3D (right) organoids went through after transcription was started. For each number of divisions, 100 organoids were simulated with randomly chosen  $-\Delta\varepsilon_u$  between 6.5 and 7.5. Resulting organoids were afterwards filtered for ratios of cell types between 0.5 and 2 in order to reduce the variation in values of  $I$ . In 2D, this resulted in 37-51% of the organoids being discarded. In 3D, the numbers were similar with 45-53%. Solid lines mark their respective mean values for  $I$  and shaded regions the corresponding standard deviations.

the effect of cell division on the clustering, and the matching of in vitro and in silico data.

Both the transcriptional regulation model as well as the organoid growth model were derived independent of each other. Thus, the time it takes for a cell to reach steady state might be completely different than the time it takes for an organoid to grow up to hundreds of cells. Therefore, the different time scales of both models had to be matched such that towards the end of the development, the organoids have values  $u$  and  $v$  close to their steady states. It has been shown that as long as cells are still able to differentiate before division, the patterns generated by the model are qualitatively similar. Hence, fast differentiation times could be used to generate patterns of mostly differentiated cells without possibly influencing the pattern formation.

It has been shown that cell divisions can lead to the formation of clusters of equal cell types in organoids consisting of cells with predefined fates [43]. In this section, a detailed comparison of static and dynamic differentiation was performed. With this, it could be demonstrated that cell division in combination with cell differentiation also leads to an increased formation of clusters. Furthermore, it was shown that the cluster formation is independent of the type of cell-cell communication. Even better, the general trends provided by differently dispersing signals is still preserved to some extent. However, there is no longer a guarantee, that ideal patterns like checkerboard or engulfing will emerge from the simulation. Seeing that the experimental data matches neither of those patterns perfectly, this might be the solution to generate patterns that resemble the experimental data more accurately.

The quantification of the number of divisions has provided more insight in how cluster formation is affected by consecutive cell divisions. For NN and low values of  $q$ , clusters tend to grow with each division. For large values of  $q$ , the exact opposite was found. The cell division disrupts the perfectly engulfing pattern of  $u^+v^-$  and  $u^-v^+$  cells, leading to a less effective segregation of these cell types and in turn a decrease of  $I$  with respect to the number of cell divisions after transcription is initiated. For intermediate values of  $q$ , a change in the trend from increasing to decreasing is therefore suggested by the results. In 2D and 3D, the value of  $q$  at which the change occurs differs.

In conclusion, it was proven that cell division indeed promotes cluster formation depending on the type of cell-cell communication. However, it remains unclear which dispersion parameter  $q$  is able to generate patterns like the experimental data. Provided the low values of  $I$  generated by an averaged NN signal or low  $q$ , it can be narrowed down to intermediate or large values of  $q$ .

## 4.6 Matching simulations with experimental data

If cell division is activated, then the geometry of the resulting organoids will never exactly match that of the experimental data due to the stochastic nature of the organoid growth model. Therefore, a direct quantitative comparison is no longer possible. Instead, the assignment of a realistic range of  $q$  that best matches the values of Moran's  $I$  was pursued. To this end, the conditions of the experimental data were replicated by simulating 3D organoids up to 200 cells before transcription is initiated. This was performed for  $T = 24$  and  $T = 48$  (Fig. 48). The simulation results again demonstrate the increase of  $I$  with both the increase of  $q$  and the increase of the cell type ratio of  $u^+v^-$  and  $u^-v^+$  cells, which is directly influenced by  $-\Delta\epsilon_u$ . In contrast to this, the 24 h-organoids exhibit only slight variations with respect to the cell type ratio with no clear trend. For low ratios, the simulation results are so close together that most organoids in this region match several values of  $q$  at once, making it difficult to only find a single value. As the ratio increases, the region covered by the values of  $I$  widens, making it easier to find good matches between experiment and simulation. A large amount of organoids fit into a range of  $q$  between 0.1 and 0.3. The 48 h-organoids already exhibit an increasing trend with respect to the cell type ratio, similar to the increase observed in the simulations. This time, the best matches are found for values of  $q$  between 0.3 and 0.4 with larger values 0.5 found frequently for ratios below 1. In total, this provides a rough classification of the cell type patterns found in the experimental data via the distance based signaling used in the computational model and the corresponding dispersion parameter  $q$ .

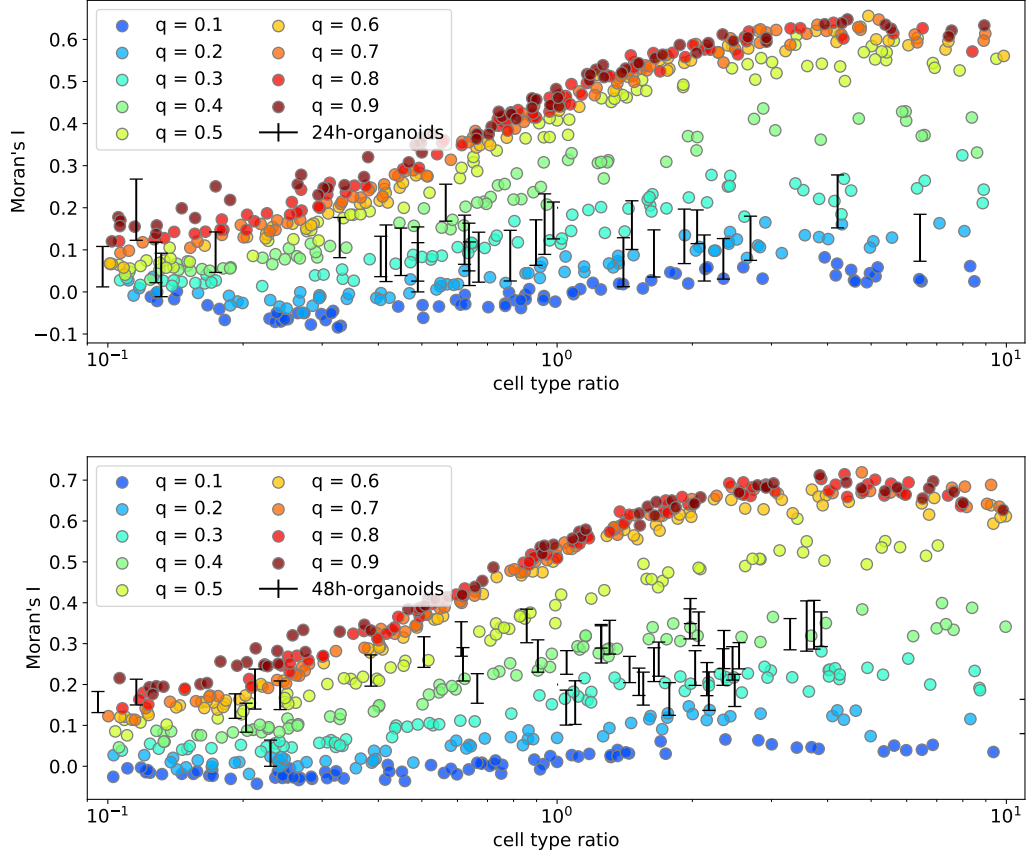


Figure 48: Moran's  $I$  with respect to the cell type ratio. Ratios are given as number of  $u^+v^-$  cells divided by the number of  $u^-v^+$  cells for the simulation data, i.e.  $|T^u|/|T^v|$ . For the experimental data, this corresponds to the number of N-G+ cells divided by the number of N+G- cells. For each  $q$  a total of 100 simulations were performed with  $-\Delta\varepsilon_u$  chosen randomly between (6, 7.86). In each simulation, the organoid was grown without transcription until 200 cells were reached. Afterwards, for  $T = 24$  (top) or  $T = 48$  (bottom) the organoid further developed with transcriptional regulation activated. Organoid ratios were filtered to ratios between 0.1 and 10. After the filtering step 63-79% of the organoids remained. Ranges of the experimental data were again obtained by 1000 samples with randomly chosen cell fates for DN and DP cells.



## 5 Discussion and Outlook

This thesis answers the question of how different cell differentiation patterns in ICM organoids can be simulated. This was achieved by the combination of a transcriptional regulation model and an organoid growth model. Furthermore, it was demonstrated how two cell types can organize themselves on a global scale using a signal that reaches beyond their nearest neighbors. Finally, the fully combined model was able to characterize the cell differentiation pattern in the ICM organoids.

### 5.1 Analytical results for the stability of heterogeneous steady states could be extended

The well known stability condition for heterogeneity of the Delta-Notch system in *Drosophila* [26] was further extended and generalized in this thesis. This generalization includes the addition of auto-activation, mutual inhibition and adjacency-based inhibition. The results is valid for a system of two cells in contact with each other. However, it remains an open question whether the stability condition can be can be further generalized for 2D or 3D tissues with  $n$  cells. For a simpler GRN, the stability condition for the two cell system has been shown to be sufficient for stability of a one-dimensional  $n$  cell system. In total, the resulting stability conditions are rather complicated in nature, making it quite difficult to handle. Luckily, none of these were needed as a different condition could be derived later on. Still it was proven, these types of analytical investigations still have room for improvement, to which the first step was initiated in this thesis.

### 5.2 Comparison of transcriptional regulation models results in one model overshadowing the rest.

Models that are subject to consecutive applications of the Hill equation are widely used in the modeling community of embryo development [27, 28, 31]. They provide a phenomenological explanation of the transcriptional regulation of NANOG and GATA6. Contrary to this, the thermodynamic model derived in this thesis aims to explain the transcriptional regulation mechanistically. This was based on the theory of transcription factor binding probabilities derived from statistical thermodynamics [37, 38, 39]. It provides a completely new perspective on the notion of inhibition and activation by the deviation from the basal activity of the respective promoter. To the best of our knowledge, this concept has not been investigated so far and was thus neither proven true or false. Yet, it opens the possibility for exciting new research. The model that was mainly chosen for the cell fate specifications in this thesis is based on a special case of the thermodynamic model. If the auto-activation is assumed to be strong enough such that the binding of the TFs NANOG and GATA6 to their own promoter becomes synonymous with transcription itself, a new model arises.

In terms of the cell differentiation itself, it was shown that the transcription factors  $u$  and  $v$  in each of the three models are directly influenced by the amount of the absorbed signal. This reflects the combined findings of [13, 14] that low amounts of FGF4 promote NANOG expression of NANOG whereas large amounts promote GATA6 expression. The inclusion of the FGF/ERK pathway as separate equations in [27, 28] leads to the existence of a third steady state, given the correct parameter values. This steady state is supposed to reflect a state of co-expression of NANOG and GATA6. In [22] however, it was demonstrated that after NANOG and GATA6 were co-expressed, cells would eventually transition towards one of the mutually exclusive states N+G− and N−G+. This reduces the problem of steady states to a binary decision. The auto-activation dominant model was shown to be a model with only two relevant steady states, which is therefore sufficient as a model to represent the cell differentiation in the mouse ICM organoid.

The auto-activation dominant model has shown that the stability conditions of the fully generalized model was not needed, as it was possible to derive a different stability condition directly for the two- and three-dimensional case. This revealed the possibility to control the heterogeneity of the steady state by adjusting the single parameter  $-\Delta\varepsilon_u$ . This was found to be a huge advantage over any other model, since large parameter sweeps and additional bifurcation analyses were not necessary to generate different cell types in an organoid. There is a trade-off in that the expression values of the steady states are only slightly affected by the signal, except for the point at which the cell fate switches. Thus, it can be argued that the auto-activation dominant model loses some of the information that would be included in the full thermodynamic and perhaps even in the phenomenological model. Nevertheless, this trade-off was decided to be more than justified in exchange for the control over pattern formation.

In conclusion, a novel model was introduced by using methods from statistical thermodynamics and the simplification by the assumption of auto-activation dominance. The model retains its ability to generate two different steady states that reflect the  $N+G-$  and  $N-G+$  states found in the ICM organoid [23]. Furthermore, its corresponding stability analysis has resulted in an a priori statement about the heterogeneity of the cell differentiation pattern. Therefore, the auto-activation dominant model has proven to be the best choice to conduct further research throughout this thesis.

### 5.3 The effect of cell-cell communication on the resulting differentiation patterns has been largely identified

Cell-cell communication via NN signaling was demonstrated to accurately describe the formation of checkerboard patterns. In contrast to similar studies with comparable results [26, 27, 31] the simulations presented in this thesis were carried out on an off-lattice organoid that is described by a graph. Thus, the checkerboard pattern was first formally characterized in this thesis with the use of PCFs as well as Moran's  $I$  for geometries that deviated from regular grids. Distance-based signaling has shown its capability to describe both checkerboard and engulfing pattern by simply increasing the dispersion parameter  $q$ . The resulting patterns closely resemble the patterns in the ICM organoids with one cell type engulfing the other. Contrary to the initial hypothesis however, the patterns were found to be reversed. That means that the cells in the computational model that were supposed to represent the  $N+G-$  cells of the ICM organoids are found to be the engulfing ones. In the ICM organoids, it was indicated that the  $N-G+$  are the engulfing ones [23], which could be confirmed on a quantitative level in this thesis by using the PCFs. Unless further research emerges that somehow justifies this reversal, long ranging cell-cell communication cannot be considered an appropriate candidate for the formation of the sorted patterns in ICM organoids. This also aligns with the recent findings of [32]. Comparing aggregates of FGF4-mutant mouse ESCs with aggregates of wild type mouse ESCs, they were able to demonstrate that after 40 h of development there was no significant difference in the cluster sizes of the two. This suggests cell-cell communication is not driving the spatial organization of the two cell types.

Despite the sobering result that the cell-cell communication via distance-based signaling is not responsible for the spatial segregation of  $N+G-$  and  $N-G+$  cells, the patterns created by the model closely resemble the ICM organoid data with the cell fates being interchanged. Motivated by this, the different resulting patterns could be classified based on the respective dispersion parameter  $q$  used for the simulation. The PCFs already revealed some matches concerning the engulfing nature of the models. However, from the low values of Moran's  $I$  it could be seen that the model lacks the capability to sufficiently form clusters that match the experimental data.

In conclusion, similar to [31] it was shown conceptually, that long range cell-cell communication is capable of influencing global cell differentiation pattern. In this thesis, the distance-based signaling obeys a simple mathematical description depending on the number of cells it passes in order to reach its destination. In reality, the absorbed signal is subject to signal emission and absorption rates as well as signal transport. In [27, 28, 31], the question of emission and absorption was already answered based on an extended GRN that includes the signal as a separate quantity. One way to tie into this would be to again replace the Hill equations with the thermodynamic description of binding events that were used throughout this thesis and introduced in [37, 38]. Apart from this, the model could benefit from a more detailed description of the signal transport mechanism, providing room for further research. One example could be a simple diffusion leading to a system similar to the well known Turing mechanism [71]. Considering the ICM organoid to be a porous medium could even open a completely new way of approaching the paracrine signaling involved [72]. Besides diffusion, the growing organoid could lead to a rise of internal pressure leading to an advection of the interstitial fluid similar to the problem of large interstitial fluid pressure found in tumors [73]. This in turn could make the signal dependent on the overall fluid dynamics within the organoid, providing a field of research that is both challenging and exciting.

### 5.4 Cell division provides the necessary clustering to describe the patterns in ICM organoids

The lacking amount of clustering in the static model was compensated via cell division. Cell division was already demonstrated to promote cluster formation [43] in organoids with fixed cell fates. A different approach to combine transcriptional regulation and cell division was attempted in [31]. However, one could argue that through the grid based arrangement of the cells, as well as the synchronous cell divisions, this approach lacks the capability to represent reality. In this thesis, the multi-scale model combined of

transcriptional regulation and an off-lattice organoid growth demonstrate the first attempts of truly replicating the mechanisms inside the mouse embryo, or rather the ICM organoid. In this combined model, cluster formation was indeed observed and was shown to be dependent on the number of cell divisions after the onset of transcriptional regulation. Through the simulations, it was possible to generate *in silico* organoids that closely resemble the *in vitro* ICM organoids from [23]. An extensive parameter study regarding the energy coefficient  $-\Delta\varepsilon_u$  that dictates the cell type proportions, as well as the dispersion parameter  $q$  was performed. This enabled the cell differentiation pattern of the ICM organoids to be classified like that of *in silico* organoids generated using a medium range of signal dispersion, i.e.  $q$  about 0.3-0.5.

In total, the analysis of the ICM organoids conducted in [23] could be extended by classifying the cell differentiation patterns more accurately. However, the exact mechanism that leads to the formation of these patterns is still missing and needs further research. Most prominently, the differential adhesion hypothesis is often mentioned as a potential cause for the cell sorting. Indeed, it could be shown in general that given large adhesive forces between cells of type A, intermediate forces between the cell types A and B and weaker forces between cells of type B, the energetically most favorable state is the one where type B cells are engulfing type A cells [70]. Simulations for the mouse embryo regarding differential adhesion have already shown the capability of cell sorting for fixed cell types [29]. However, it is unclear if the sorting mechanism alone suffices to yield a spatial segregation on the scale to the ICM organoid. This could motivate further research, by extending the model presented in this thesis with differential adhesion. One way this could be realized is by modification of the Morse potential based on the the respective cell types that are in contact with each other.

## 5.5 Outlook - From ICM organoid towards the embryo

During the course of this work, it became clear that a computational model for describing the pattern formation of ICM organoids already requires a great effort by combining the mathematical descriptions of transcriptional regulation and tissue growth. This effort increases even further for the mouse embryo. Although ICM organoids have much more cells than mid to late stage mouse blastocysts, they also present a simpler geometry. In the mouse embryo, the ICM cells are additionally surrounded by the trophectoderm and the blastocoel cavity is formed. In [23] it was already hypothesized that the trophectoderm provides a mechanical constraint, such that the N-G+ cells cannot fully engulf the N+G- cells but rather only form the layer separating the N+G- cells from the blastocoel cavity. It remains to be shown, how exactly this layer is realized. Differential adhesion has already shown promising results [29, 19] but still remains to be implemented in a model that additionally runs transcriptional regulation and tissue growth in parallel.

In a different study, it was shown that disturbance of the blastocoel expansion leads to a disruption of PrE cell maturation, hinting towards mechanical cues influencing the cell differentiation [74]. However, they also hypothesize that by expansion of the blastocoel, the accumulated FGF4 concentration inside the blastocoel increases, guiding the cell differentiation of ICM cells. This suggests again a signal based differentiation and spatial segregation of the Epi and PrE cells.

Both the mechanical effects and the chemical effects could be promising in answering the question of how the embryo is composed. For this purpose, it is advised to use a model like the one presented in this study. Although already mentioned several times throughout this thesis, its advantages should again be listed enumerated. The transcriptional regulation model allows the *a priori* control over the cell type proportions. The tissue growth model is off-lattice, so it is versatile for any tissue geometry. Both parts of the fully combined model run in parallel, reflecting a realistic biological scenario. If the work of [23] is considered to be the foundation of the research presented in this thesis, then it is hoped that this work will in turn provide the basic building block for potential extensions that will enable a transition from ICM organoids towards the embryo.

## Acknowledgements

I would like to take this opportunity to express my gratitude to some of the people who have accompanied and supported me on this amazing journey. Most of all Sabine Fischer, who always provided support at any given time for any given problem. In her mind it was always a given, but in mine it was anything but. I can only hope that she realizes how helpful she is to me and her entire working group. Also, her ability to translate crazy biology talk for dumb mathematicians really helped me out. I also want to thank Philip Kollmannsberger and Christian Schröter for the additional supervision of my PhD project. I have always appreciated their feedback in a way that makes me question why I haven't reached out more often. I am grateful for the whole CSI group especially Kerstin Schmid, Andreas Kuhn and Sascha Ollertz who accompany me on the same path. I am convinced we have created a lovely working environment that really helps in both joyful and desperate times. The CCTB as a whole, despite all the organizational issues, is full of interesting research topics and people. If there wasn't so much pressure to finish my own work, I would have loved to collaborate more with others and learn about ecological modeling, image analysis and whatever GWAS is. Nevertheless, the numerous conversations during lunch times, coffee breaks and other events were always a pleasure. However, I had to learn that opinions, despite being completely subjective, can still be objectively wrong, especially about food. A special thank you also goes out to Jana Blechschmidt who took the time to proofread this thesis. Finally, I want to thank my wife Janine, who provided me with the necessary comfort needed to stay sane throughout this whole journey, especially in the final weeks of finishing this thesis.

## References

- [1] Joel E Cohen. Mathematics Is Biology's Next Microscope, Only Better; Biology Is Mathematics' Next Physics, Only Better. PLoS Biology, 2(12):e439, December 2004.
- [2] Gavin E. Jarvis. Estimating limits for natural human embryo mortality. F1000Research, 5:2083, August 2016.
- [3] Allen J Wilcox, Quaker Harmon, Kevin Doody, Don P Wolf, and Eli Y Adashi. Preimplantation loss of fertilized human ova: estimating the unobservable. Human Reproduction, 35(4):743–750, April 2020.
- [4] Fredrik Lanner. Lineage specification in the early mouse embryo. Experimental Cell Research, 321(1):32 – 39, 2014.
- [5] Jaime A. Rivera-Pérez and Anna-Katerina Hadjantonakis. The dynamics of morphogenesis in the early mouse embryo. Cold Spring Harbor perspectives in biology, 7(11):a015867, 2014.
- [6] Tristan Frum and Amy Ralston. Cell signaling and transcription factors regulating cell fate during formation of the mouse blastocyst. Trends in Genetics, 31(7):402 – 410, 2015.
- [7] Claire Chazaud and Yojiro Yamanaka. Lineage specification in the mouse preimplantation embryo. Development, 143(7):1063–1074, 2016.
- [8] Kaoru Mitsui, Yoshimi Tokuzawa, Hiroaki Itoh, Kohichi Segawa, Mirei Murakami, Kazutoshi Takahashi, Masayoshi Maruyama, Mitsuyo Maeda, and Shinya Yamanaka. The homeoprotein nanog is required for maintenance of pluripotency in mouse epiblast and es cells. Cell, 113(5):631 – 642, 2003.
- [9] Nadine Schrode, Néstor Saiz, Stefano Di Talia, and Anna-Katerina Hadjantonakis. Gata6 levels modulate primitive endoderm cell fate choice and timing in the mouse blastocyst. Developmental cell, 29(4):454–467, 2014.
- [10] Amar M. Singh, Takashi Hamazaki, Katherine E. Hankowski, and Naohiro Terada. A heterogeneous expression pattern for nanog in embryonic stem cells. STEM CELLS, 25(10):2534–2542, 2007.
- [11] Sissy E. Wamaitha, Ignacio del Valle, Lily T. Y. Cho, Yingying Wei, Norah M. E. Fogarty, Paul Blakeley, Richard I. Sherwood, Hongkai Ji, and Kathy K. Niakan. Gata6 potently initiates reprogramming of pluripotent and differentiated cells to extraembryonic endoderm stem cells. Genes & Development, 29(12), June 2015.
- [12] Laurie A. Boyer, Tong Ihn Lee, Megan F. Cole, Sarah E. Johnstone, Stuart S. Levine, Jacob P. Zucker, Matthew G. Guenther, Roshan M. Kumar, Heather L. Murray, Richard G. Jenner, David K.

- Gifford, Douglas A. Melton, Rudolf Jaenisch, and Richard A. Young. Core Transcriptional Regulatory Circuitry in Human Embryonic Stem Cells. *Cell*, 122(6):947–956, September 2005. Publisher: Elsevier.
- [13] Jennifer Nichols, Jose Silva, Mila Roode, and Austin Smith. Suppression of Erk signalling promotes ground state pluripotency in the mouse embryo. *Development*, 136(19):3215–3222, 2009.
  - [14] Yojiro Yamanaka, Fredrik Lanner, and Janet Rossant. FGF signal-dependent segregation of primitive endoderm and epiblast in the mouse blastocyst. *Development*, 137(5):715–724, 2010.
  - [15] Yusuke Ohnishi, Wolfgang Huber, Akiko Tsumura, Minjung Kang, Panagiotis Xenopoulos, Kazuki Kurimoto, Andrzej K. Oleś, Marcos J. Araújo-Bravo, Mitinori Saitou, Anna-Katerina Hadjantonakis, and Takashi Hiragi. Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages. *Nature Cell Biology*, 16(1):27–37, January 2014.
  - [16] Katherine E. Santostefano, Takashi Hamazaki, Carolina E. Pardo, Michael P. Kladde, and Naohiro Terada. Fibroblast Growth Factor Receptor 2 Homodimerization Rapidly Reduces Transcription of the Pluripotency Gene Nanog without Dissociation of Activating Transcription Factors \*. *Journal of Biological Chemistry*, 287(36):30507–30517, August 2012. Publisher: Elsevier.
  - [17] Stephen Frankenberg, François Gerbe, Sylvain Bessonard, Corinne Belville, Pierre Pouchin, Olivier Bardot, and Claire Chazaud. Primitive Endoderm Differentiates via a Three-Step Mechanism Involving Nanog and RTK Signaling. *Developmental Cell*, 21(6):1005–1013, December 2011. Publisher: Elsevier.
  - [18] Sabine C. Fischer, Elena Corujo-Simon, Joaquin Lilao-Garzon, Ernst H. K. Stelzer, and Silvia Muñoz-Descalzo. The transition from local to global patterns governs the differentiation of mouse blastocysts. *PLOS ONE*, 15(5):1–29, 2020.
  - [19] Silas Boye Nissen, Marta Perera, Javier Martin Gonzalez, Sophie M. Morgani, Mogens H. Jensen, Kim Sneppen, Joshua M. Brickman, and Ala Trusina. Four simple rules that are sufficient to generate the mammalian blastocyst. *PLOS Biology*, 15(7):1–30, 2017.
  - [20] Katarzyna Filimonow, Nestor Saiz, Aneta Suwinska, Tomasz Wyszomirski, Joanna B. Grabarek, Elisabetta Ferretti, Anna Piliszek, Berenika Plusa, and Marek Maleszewski. No evidence of involvement of e-cadherin in cell fate specification or the segregation of epi and pre in mouse blastocysts. *PloS one*, 14(2):e0212109–e0212109, 2019.
  - [21] Zixuan Cang, Yangyang Wang, Qixuan Wang, Ken W. Y. Cho, William Holmes, and Qing Nie. A multiscale model via single-cell transcriptomics reveals robust patterning mechanisms during early mammalian embryo development. *PLOS Computational Biology*, 17(3):1–20, 2021.
  - [22] Christian Schröter, Pau Rué, Jonathan Peter Mackenzie, and Alfonso Martinez Arias. FGF/MAPK signaling sets the switching threshold of a bistable circuit controlling cell fate decisions in embryonic stem cells. *Development*, 142(24):4205–4216, 2015.
  - [23] Biena Mathew, Silvia Muñoz-Descalzo, Elena Corujo-Simon, Christian Schröter, Ernst H.K. Stelzer, and Sabine C. Fischer. Mouse ICM organoids reveal three-dimensional cell fate clustering. *Biophysical Journal*, 116(1):127 – 141, 2019.
  - [24] Sui Huang, Yan-Ping Guo, Gillian May, and Tariq Enver. Bifurcation dynamics in lineage-commitment in bipotent progenitor cells. *Developmental Biology*, 305(2):695 – 713, 2007.
  - [25] Joshua L Cherry and Frederick R Adler. How to make a Biological Switch. *Journal of Theoretical Biology*, 203(2):117–133, March 2000.
  - [26] Joanne R. Collier, Nicholas A. M. Monk, Philip K. Maini, and Julian H. Lewis. Pattern Formation by Lateral Inhibition with Feedback: a Mathematical Model of Delta-Notch Intercellular Signalling. *Journal of Theoretical Biology*, 183(4):429–446, December 1996.
  - [27] Sylvain Bessonard, Laurane De Mot, Didier Gonze, Manon Barriol, Cynthia Dennis, Albert Goldbeter, Geneviève Dupont, and Claire Chazaud. Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network. *Development*, 141(19):3637–3648, October 2014.

- [28] Alen Tosenberger, Didier Gonze, Sylvain Bessonard, Michel Cohen-Tannoudji, Claire Chazaud, and Geneviève Dupont. A multiscale model of early cell lineage specification including cell division. npj Systems Biology and Applications, 3(1):16, 2017.
- [29] Pawel Krupinski, Vijay Chickarmane, and Carsten Peterson. Simulating the mammalian blastocyst - molecular and mechanical interactions pattern the embryo. PLOS Computational Biology, 7(5):1–11, 2011.
- [30] Néstor Saiz, Laura Mora-Bitria, Shahadat Rahman, Hannah George, Jeremy P Herder, Jordi Garcia-Ojalvo, and Anna-Katerina Hadjantonakis. Growth-factor-mediated coupling between lineage size and cell fate choice underlies robustness of mammalian development. eLife, 9:e56079, 2020.
- [31] Angel Stanoev, Christian Schröter, and Aneta Koseska. Robustness and timing of cellular differentiation through population-based symmetry breaking. Development, 148(3), 2021.
- [32] Dhruv Raina, Azra Bahadori, Angel Stanoev, Michelle Protzek, Aneta Koseska, and Christian Schröter. Cell-cell communication through FGF4 generates and maintains robust proportions of differentiated cell types in embryonic stem cells. Development, (dev.199926), October 2021.
- [33] P. A. P. Moran. Notes on Continuous Stochastic Phenomena. Biometrika, 37(1/2):17–23, 1950. Publisher: [Oxford University Press, Biometrika Trust].
- [34] Benjamin J. Binder and Matthew J. Simpson. Quantifying spatial structure in experimental observations and agent-based simulations using pair-correlation functions. Phys. Rev. E, 88:022705, 2013.
- [35] Enrico Gavagnin, Jennifer P. Owen, and Christian A. Yates. Pair correlation functions for identifying spatial correlation in discrete domains. Phys. Rev. E, 97:062104, Jun 2018.
- [36] A. V. Hill. The possible effects of the aggregation of the molecules of haemoglobin on its dissociation curves. The Journal of Physiology, 40:4–7, 1910.
- [37] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, and Rob Phillips. Transcriptional regulation by the numbers: models. Current Opinion in Genetics & Development, 15(2):116–124, 2005. Chromosomes and expression mechanisms.
- [38] Lacramioara Bintu, Nicolas E Buchler, Hernan G Garcia, Ulrich Gerland, Terence Hwa, Jané Kondev, Thomas Kuhlman, and Rob Phillips. Transcriptional regulation by the numbers: applications. Current Opinion in Genetics & Development, 15(2):125–135, 2005. Chromosomes and expression mechanisms.
- [39] Hernan G. Garcia, Jane Kondev, Nigel Orme, Julie A. Theriot, and Rob Phillips. Thermodynamics of biological processes. Methods in enzymology, 492:27–59, 2011.
- [40] Alistair M. Middleton, Christian Fleck, and Ramon Grima. A continuum approximation to an off-lattice individual-cell based model of cell migration and adhesion. Journal of Theoretical Biology, 359:220 – 232, 2014.
- [41] Damian Stichel, Alistair M. Middleton, Benedikt F. Müller, Sofia Depner, Ursula Klingmüller, Kai Breuhahn, and Franziska Matthäus. An individual-based model for collective cancer cell migration explains speed dynamics and phenotype variability in response to growth factors. npj Systems Biology and Applications, 3(1):5, 2017.
- [42] Philip M. Morse. Diatomic Molecules According to the Wave Mechanics. II. Vibrational Levels. Physical Review, 34(1):57–64, July 1929. Publisher: American Physical Society.
- [43] Tim Liebisch, Armin Drusko, Biana Mathew, Ernst H. K. Stelzer, Sabine C. Fischer, and Franziska Matthäus. Cell fate clusters in icm organoids arise from cell fate heredity and division: a modelling approach. Scientific Reports, 10(1):22405, 2020.
- [44] Alexander Schmitz, Sabine C. Fischer, Christian Mattheyer, Francesco Pampaloni, and Ernst H. K. Stelzer. Multiscale image analysis reveals structural heterogeneity of the cell microenvironment in homotypic spheroids. Scientific Reports, 7(1):43693, 2017.
- [45] Edward F. Moore. The shortest path through a maze. Proceedings of the international Symposium on the Theory of Switching, pages 285–292, 1959.

- [46] J. A. Barker and D. Henderson. What is "liquid"? Understanding the states of matter. Reviews of Modern Physics, 48(4):587–671, October 1976.
- [47] I. Smyrek and E. H. K. Stelzer. Quantitative three-dimensional evaluation of immunofluorescence staining for large whole mount spheroids with light sheet microscopy. Biomedical Optics Express, 8(2):484–499, January 2017.
- [48] Néstor Saiz, Kiah M. Williams, Venkatraman E. Seshan, and Anna-Katerina Hadjantonakis. Asynchronous fate decisions by single cells collectively ensure consistent lineage composition in the mouse blastocyst. Nature Communications, 7(1):13463, 2016.
- [49] Ian Chambers, Douglas Colby, Morag Robertson, Jennifer Nichols, Sonia Lee, Susan Tweedie, and Austin Smith. Functional Expression Cloning of Nanog, a Pluripotency Sustaining Factor in Embryonic Stem Cells. Cell, 113(5):643–655, May 2003.
- [50] Kathy Q. Cai, Callinice D. Capo-Chichi, Malgorzata E. Rula, Dong-Hua Yang, and Xiang-Xi Xu. Dynamic GATA6 Expression in Primitive Endoderm Formation and Maturation in Early Mouse Embryogenesis. Developmental dynamics : an official publication of the American Association of Anatomists, 237(10):2820–2829, October 2008.
- [51] Plotly Technologies Inc. Collaborative data science. Montreal, QC.
- [52] Pauli Virtanen, Ralf Gommers, Travis E. Oliphant, Matt Haberland, Tyler Reddy, David Cournapeau, Evgeni Burovski, Pearu Peterson, Warren Weckesser, Jonathan Bright, Stéfan J. van der Walt, Matthew Brett, Joshua Wilson, K. Jarrod Millman, Nikolay Mayorov, Andrew R. J. Nelson, Eric Jones, Robert Kern, Eric Larson, C J Carey, İlhan Polat, Yu Feng, Eric W. Moore, Jake VanderPlas, Denis Laxalde, Josef Perktold, Robert Cimrman, Ian Henriksen, E. A. Quintero, Charles R. Harris, Anne M. Archibald, Antônio H. Ribeiro, Fabian Pedregosa, Paul van Mulbregt, and SciPy 1.0 Contributors. SciPy 1.0: Fundamental Algorithms for Scientific Computing in Python. Nature Methods, 17:261–272, 2020.
- [53] Charles R. Harris, K. Jarrod Millman, Stéfan J. van der Walt, Ralf Gommers, Pauli Virtanen, David Cournapeau, Eric Wieser, Julian Taylor, Sebastian Berg, Nathaniel J. Smith, Robert Kern, Matti Picus, Stephan Hoyer, Marten H. van Kerkwijk, Matthew Brett, Allan Haldane, Jaime Fernández del Río, Mark Wiebe, Pearu Peterson, Pierre Gérard-Marchant, Kevin Sheppard, Tyler Reddy, Warren Weckesser, Hameer Abbasi, Christoph Gohlke, and Travis E. Oliphant. Array programming with NumPy. Nature, 585(7825):357–362, September 2020.
- [54] J. D. Hunter. Matplotlib: A 2d graphics environment. Computing in Science & Engineering, 9(3):90–95, 2007.
- [55] Till Tantau. The TikZ and PGF Packages.
- [56] WEN-CHYUAN YUEH and SUI SUN CHENG. Explicit eigenvalues and inverses of tridiagonal toeplitz matrices with four perturbed corners. The ANZIAM Journal, 49(3):361–387, 2008.
- [57] Jan-Hendrik S. Hofmeyr and Hofmeyr Cornish-Bowden. The reversible Hill equation: how to incorporate cooperative enzymes into metabolic models. Bioinformatics, 13(4):377–385, August 1997.
- [58] Leonor Michaelis and Maud Leonora Menten. Die kinetik der invertinwirkung. Biochemische Zeitschrift, 49:333 – 369, 1913.
- [59] G.S. Adair, A.V. Bock, and H. Field. The hemoglobin system: VI. the oxygen dissociation curve of hemoglobin. Journal of Biological Chemistry, 63(2):529–545, 1925.
- [60] Jacques Monod, Jeffries Wyman, and Jean-Pierre Changeux. On the nature of allosteric transitions: A plausible model. Journal of Molecular Biology, 12(1):88–118, 1965.
- [61] Alison C. Lloyd. The Regulation of Cell Size. Cell, 154(6):1194–1205, September 2013.
- [62] Ian Conlon and Martin Raff. Differences in the way a mammalian cell and yeast cells coordinate cell growth and cell-cycle progression. Journal of Biology, 2(1):7, April 2003.
- [63] Mustafa Mir, Zhuo Wang, Zhen Shen, Michael Bednarz, Rashid Bashir, Ido Golding, Supriya G. Prasanth, and Gabriel Popescu. Optical measurement of cycle-dependent cell growth. Proceedings of the National Academy of Sciences, 108(32):13124–13129, August 2011. Publisher: Proceedings of the National Academy of Sciences.

- [64] Helmut Dolznig, Florian Grebien, Thomas Sauer, Hartmut Beug, and Ernst W. Müllner. Evidence for a size-sensing mechanism in animal cells. Nature Cell Biology, 6(9):899–905, September 2004. Number: 9 Publisher: Nature Publishing Group.
- [65] Ran Kafri, Jason Levy, Miriam B. Ginzberg, Seungeun Oh, Galit Lahav, and Marc W. Kirschner. Dynamics extracted from fixed cells reveal feedback linking cell growth to cell cycle. Nature, 494(7438):480–483, February 2013. Number: 7438 Publisher: Nature Publishing Group.
- [66] Jie Lin and Ariel Amir. From single-cell variability to population growth. Physical Review E, 101(1):012401, January 2020. Publisher: American Physical Society.
- [67] R. Milo and R. Phillips. Cell biology by the numbers (1st ed.). Garland Science, 2015.
- [68] J.-L. Doumont. The three laws of professional communication. IEEE Transactions on Professional Communication, 45(4):291–296, December 2002. Conference Name: IEEE Transactions on Professional Communication.
- [69] Christopher Revell, Raphael Blumenfeld, and Kevin J. Chalut. Force-based three-dimensional model predicts mechanical drivers of cell sorting. Proceedings of the Royal Society B: Biological Sciences, 286(1895):20182495, January 2019. Publisher: Royal Society.
- [70] Mathieu Emily and Olivier François. A statistical approach to estimating the strength of cell-cell interactions under the differential adhesion hypothesis. Theoretical biology & medical modelling, 4:37–37, 2007.
- [71] Alan Turing. The chemical basis of morphogenesis. Philosophical Transactions of the Royal Society B, 237:37–72, 1952.
- [72] Daniel M. Tartakovsky and Marco Dentz. Diffusion in Porous Media: Phenomena and Mechanisms. Transport in Porous Media, 130(1):105–127, October 2019.
- [73] Michael Welter and Heiko Rieger. Interstitial Fluid Flow and Drug Delivery in Vascularized Tumors: A Computational Model. PLOS ONE, 8(8):e70395, August 2013. Publisher: Public Library of Science.
- [74] Allyson Quinn Ryan, Chii Jou Chan, François Graner, and Takashi Hiiragi. Lumen Expansion Facilitates Epiblast-Primitive Endoderm Fate Specification during Mouse Blastocyst Formation. Developmental Cell, 51(6):684–697.e4, December 2019.