Modeling Human Epiblast Morphogenesis

by

Agnes M. Resto

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Mechanical Engineering)
in the University of Michigan
2023

Doctoral Committee:

       Professor Jianping Fu, Chair
       Professor Jennifer Linderman
       Associate Professor Allen Liu
       Associate Professor Luis Zaman

Agnes M. Resto-Irizarry

amresto@umich.edu

ORCID iD: 0000-0002-2292-6029

Acknowledgements

I would like to first express my deepest gratitude to my advisor Prof. Jianping Fu for his mentoring and guidance during my entire PhD study. I am always inspired by Prof. Fu's passion, his vision, and his rigorous attitude towards research. I thank him for allowing me the freedom to be creative and incorporate tools and ideas from unfamiliar fields, allowing me to grow and become confident as an independent researcher. I truly appreciate his genuine care towards my career path, which he showed by consistently dedicating his time to craft my professional image and build connections in the field.

Besides my advisor, I am truly grateful for all my mentors and collaborators. A great appreciation is paid to Prof. Jennifer Linderman, who helped me delve into the world of agent-based modeling, which is now a primary element of my PhD work. A great appreciation is also paid to Prof. Luis Zaman, who first introduced me to the concept of genetic algorithms. Dr. Zaman and his former post-doctoral student, Dr. Alexander Lalejini, guided me in my efforts to incorporate this into my research to the point that it is now one of the main components. I also want to thank Dr. Allen Liu, for being interested in my work, for taking the time to provide valuable feedback on my research.

I also sincerely thank my former and current lab mates: Dr. Xufeng Xue, Dr. Sajedeh Nasr Esfahani, Dr. Yi Zheng, Dr. Yue Liu, Dr. Norio Kobayashi, Dr. Yung Su Kim, Jeyoon Bok, Shiyu Sun, Zhuowei Zhou, Zhaoyi Xu, Aidan Terhune, Robin Zhexuan Yan, and Samuel B. Lee.

I want to thank them for freely sharing their knowledge and skills with me, for encouraging me, and for their insightful suggestions for my research.

Last but certainly not least, I am sincerely grateful to my family and my friends for their unbelievable love and emotional support.

Table of Contents

List of Figures

Abstract

      The development of the human embryo is arguably the most complex process that we could care to study. In this process, the developing embryo must undergo proliferation, reorganization, lineage diversification, and dozens of cell fate specification events. During this time, a myriad of events are happening in parallel at the cell level, each one setting the foundation for the emergence of increasingly complex tissues of increasingly complex function. Understanding the mechanisms guiding these processes is pivotal not only for embryogenesis-related applications in fertility and development, but also for regenerative medicine applications such as the development of organ replacements.

      In this dissertation, I propose an integrative approach to the study of morphogenesis and patterning, specifically in the context of stem cell-based models of human development. Firstly, I present a novel machine learning-assisted imaging pipeline that permits the careful characterization of cell-level events occurring in our *in vitro* model of epiblast cyst morphogenesis. Secondly, I present a novel agent-based model (ABM)-genetic algorithm (GA) framework for the generation of models of morphogenesis. The framework was first tested to determine its ability to generate structures of desired patterns. It was then applied for the generation of models that plausibly capture mechanisms at work during epiblast cyst morphogenesis and symmetry breaking. With preliminary *in silico* experiments, I showed that the framework was able to output models that partially captured the effect of initial cell number on final cyst composition. I further showed that correct structure formation was heavily impacted

by just a few model parameters. Combined with *in vitro* experimentation, these tools have the potential to shed light into the mechanisms guiding growth, movement, and cell fate specification in *in vitro* models of human development.

Chapter 1 Introduction

1.1 Introduction

The study of the emergence of form and function in the context of human development is a huge undertaking. The high degree of complexity of morphogenetic events occurring at different temporal and spatial scales as well as the difficulties surrounding the study of early embryo development have both contributed to a lack in our knowledge regarding embryogenesis and organogenesis. This dissertation aims to provide novel computational tools that can be used in conjunction with *in vitro* stem cell-based models of development in order to overcome this gap in knowledge (Figure 1.1). In this chapter, I will give a brief background and outline of the dissertation.
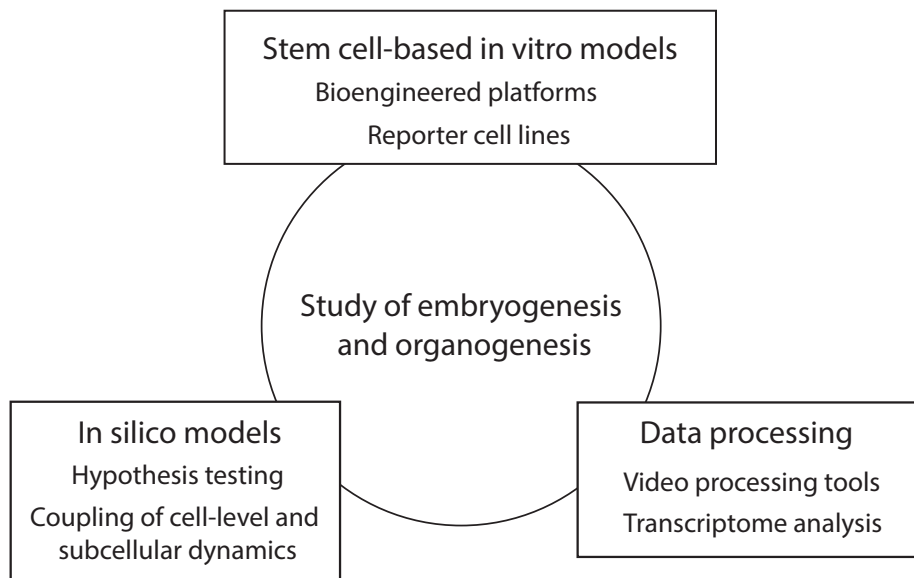
Figure 1.1 Integrated approach to the study of development

1.2 Beginning of embryogenesis – the first three weeks

The formation of a human body is a forty-week process that starts at conception and culminates at birth. Starting from a single fertilized egg, the embryo undergoes a myriad of processes involving divisions, migration, and cell fate specification, all of which must occur in a precise manner and at a precise time. Of particular focus in this work will be the first three weeks of development, a period of development that is understandably difficult to study but in which many pivotal events for human development take place. Firstly, the fertilized egg or zygote must undergo a series of cleavage events. At this point, this seemingly homogenous cluster of cells (morula) undergoes the first cell fate bifurcation in development to give rise to the blastocyst, a hollow structure containing an inner cell mass (ICM) and lined by trophectoderm cells, which will later develop into placental tissues and other extraembryonic tissues (Figure 1.2). Shortly after, the cells in the ICM undergo a second cell fate bifurcation, in which two populations come about, the epiblast population will give rise to the embryo proper, and the hypoblast cells will give rise to the primary and secondary yolk sacs.
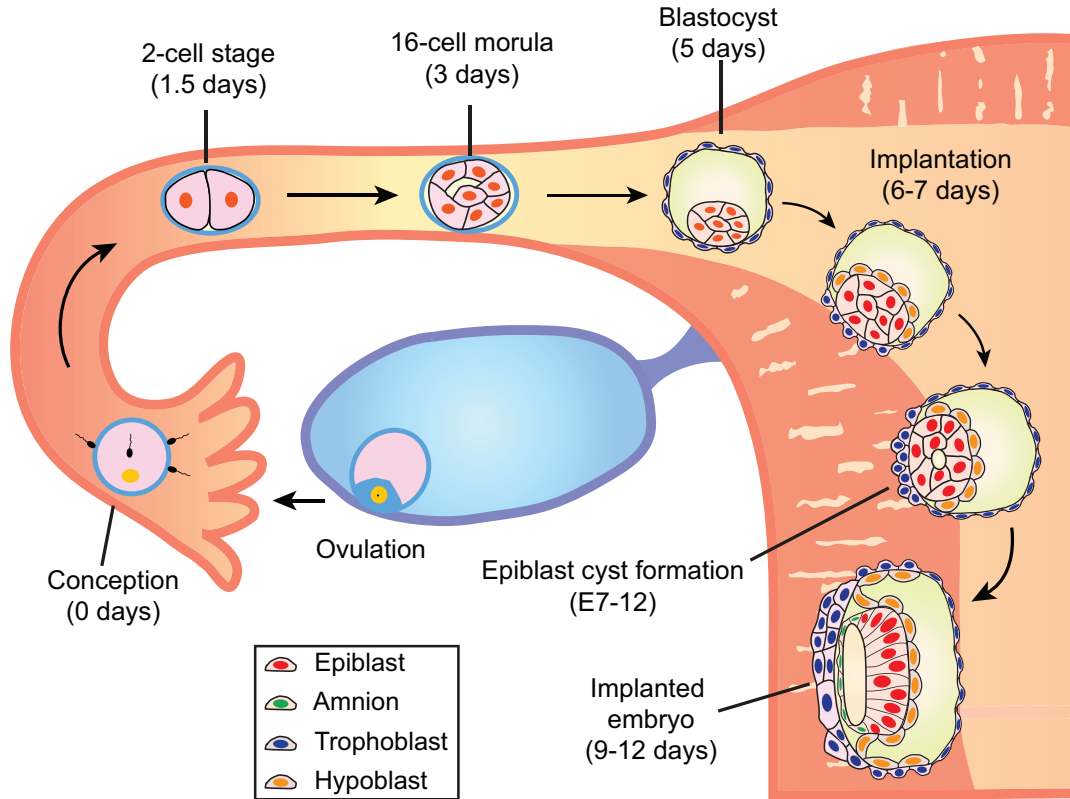
Figure 1.2 Schematic of first two weeks of human development.

Upon 5 to 6 days after fertilization (d.p.f.), the newly formed blastocyst begins the process of implantation on the uterine wall. During this time, the epiblast cyst undergoes polarization to form the pro-amniotic cavity (Figure 1.3). As implantation proceeds, this cavity expands and the cells facing the maternal tissue undergo differentiation into presumptive amniotic cells, giving rise to the amniotic sac. At 15-17 d.p.f., cells in the epiblast's projected posterior end start the process of gastrulation, which will give rise to the three germ layers (ectoderm, endoderm, and mesoderm). Gastrulation is one of the most crucial steps in development as it primes the system for downstream organogenesis.
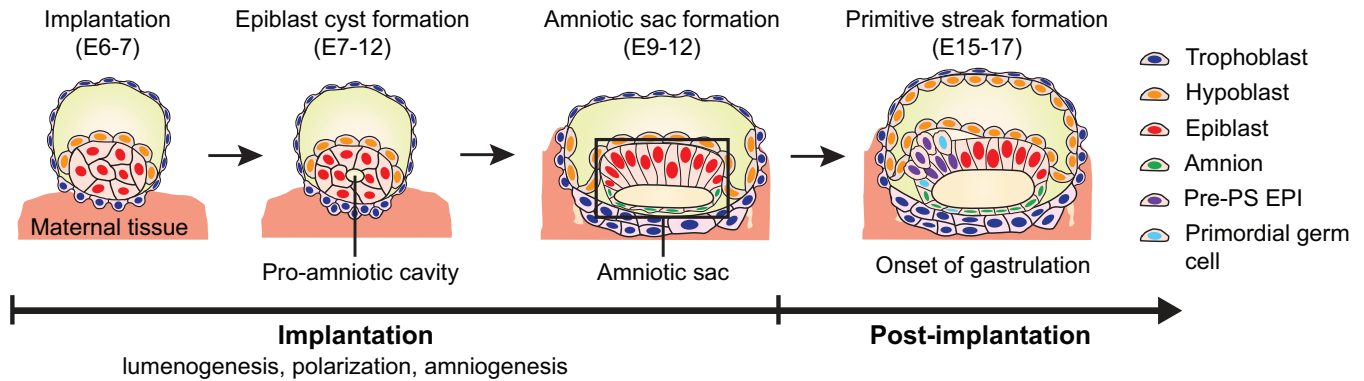
Figure 1.3 Schematic of embryo development during and shortly after implantation.

1.3 Human pluripotent stem cell-based models of development

As one of the key areas in modern biology, human development is a complex process that involves cell differentiation, self-organization, and growth. It takes place across multiple spatial and temporal scales, with processes at the subcellular level affecting the behavior of individual cells and leading to complex emergent multicellular phenomena. The difficulty of studying human embryo development *in vivo* has led researchers to turn to the use of animal models. Monkey embryos have been used to gain insights into primate development[1–4]. However, working with monkey embryos is expensive, time-consuming, and may raise its own ethical concerns[5]. Research with the mouse is more ubiquitous, and most of our current knowledge of mammalian development is obtained from the mouse embryo. Importantly, great strides have been made in recent years with the creation of *in vitro* embryo models (or embryoids) using mouse stem cells[6–10]. However, we've come to discover that there are key differences between the mouse and the human even at the early days of the embryo development[11–14]. Another approach to study human development has been the culture of human embryos obtained from *in vitro* fertilization[15–17]. This approach has yielded insights into the autonomy and self-organizing properties of human embryo development even in the absence of the maternal environment.

However, there is limited accessibility to human embryos, and studies of human embryos are further limited by the small window of research afforded by the 14-day rule [18,19]. The limitations of the aforementioned approaches have led researchers to take a bottom-up approach and use human stem cells, particularly human pluripotent stem cells (hPSCs), to construct synthetic models of human development (or human embryoid systems)[20,21].

1.3.1 Derivation and maintenance

hPSCs include human embryonic stem cells (hESCs) isolated from the ICM of a human blastocyst (Figure 1.4) and human induced pluripotent stem cells (hiPSCs)[22–24] which are derived from adult tissue cells via a reprogramming process involving forced expression of key transcription factors associated with pluripotency. hPSCs are maintained in culture and are able to self-renew and expand for continued use for the development of *in vitro* models (Figure 1.4).
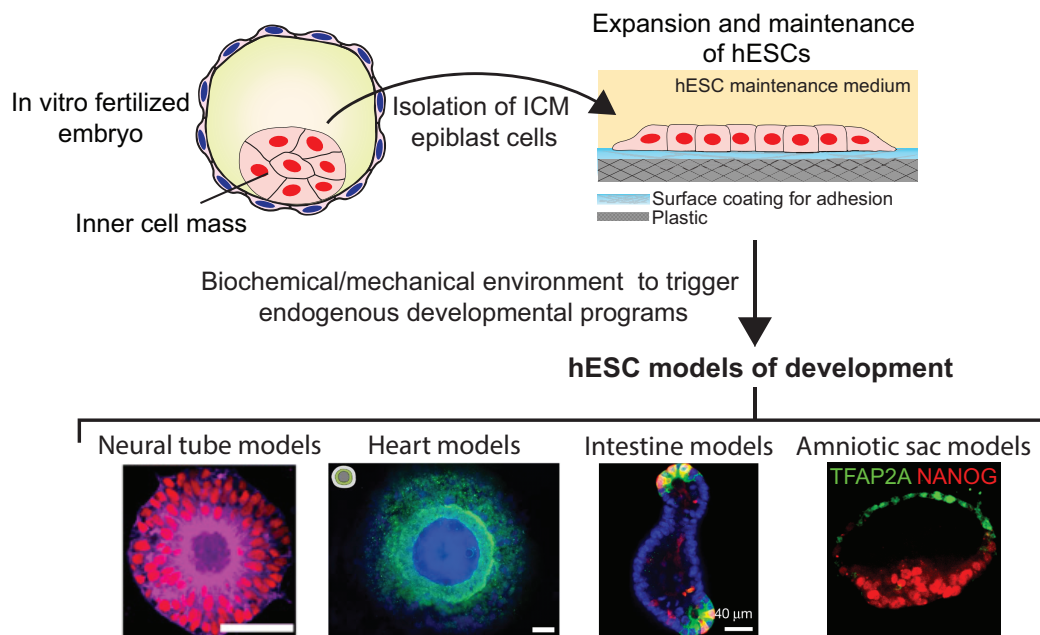
Figure 1.4 Derivation and maintenance of hPSCs for the construction of *in vitro* models of embryogenesis and organogenesis

1.3.2 Bioengineering approaches for the construction of stem cell-based models

Stem cell behavior is governed by both endogenous, genome-encoded mechanisms, and exogenous signals provided by the surrounding extracellular matrix (ECM) and cell-cell interactions. In order to construct embryoid models, researchers must provide stem cells with an environment that recapitulates the necessary exogenous signals in the *in vivo* setting. Bioengineering tools can be used to create biomimetic niches in which parameters such as adhesive surface area, dimensionality, and mechanical and chemical properties of the environment can be dynamically modulated to guide the progressive development of embryoid models to recapitulate different aspects of embryonic development [25,26]. For example, microcontact printing is a commonly used tool to control adhesive surface area in order to obtain two-dimensional (2D) cell colonies of specific shapes and sizes. The technique consists of coating a micropatterned substrate or the stamp with adhesive ECM proteins before stamping it onto a planar 2D surface where cells will be seeded. Microcontact printing has allowed for the study of how biomechanical mechanisms involving cell density and shape affect stem cell fate [27] as well as for the study of signaling dynamics within stem cell colonies[28–31]. Basement membrane matrices and synthetic hydrogels are often used to create three-dimensional (3D) environments containing adhesive molecules. Stiffness of these 3D scaffolds can be tuned by changing the gel concentration or the crosslinker density in synthetic hydrogels. This allows for assay optimization to obtain desired behaviors of stem cells. Furthermore, it allows for studies of how the mechanical properties of the environment affect hPSC differentiation and self-organization[32]. These 3D scaffolds can be incorporated into a variety of bioengineering platforms that allow for the control of initial cluster sizes of stem cells, including microwells and microfluidic devices. The current state of microfabrication technologies is such that stem cell culture platforms can be made arbitrarily complex with the use of microfluidic devices capable

of creating dynamic chemical gradients and the fabrication of substrates with different topographies. Importantly, even simple bioengineering platforms that provide a homogeneous 2D or 3D environment are able to engender the formation of complex human embryoid structures from hPSCs[28,33,34].

1.3.3 Developmental potential

The pre-implantation human embryo or the human blastocyst contains two extraembryonic lineages (trophoblast and hypoblast) and one embryonic lineage (epiblast) (Figure 1.2). Human embryoids aiming to recapitulate the peri-implantation development of the human embryo would ideally include stem cell derivatives associated with all of the three cell lineages, either from co-culture or from using human stem cells with the potency to give rise to both extraembryonic and embryonic lineages. Significant recent effort has been devoted to the development of the human hypoblast stem cells from naïve hPSCs[35,36]. While progress has been made to derive human trophoblast stem cells from human blastocysts and placenta [37], many researchers have been using BMP4 treatments to coax primed hPSCs to differentiate into trophoblast-like cells [38]. Recent studies have also reported human expanded potential stem cells (hEPSCs) [39–41] that seem to be able to give rise to both embryonic and extraembryonic lineages. However, the potencies of primed hPSCs and hEPSCs to give rise to extraembryonic lineages are still under debate. A recent study by Guo *et al.* [12] has shed doubt on the validity of these methods. They show that both primed hPSCs and hEPSCs have lost the capacity to differentiate into trophoblast-like cells and that BMP4 treatment causes differentiation of both cell types into amniotic cells as shown by Shao *et al.*[32]. They further show that naïve hPSCs[42–44] are able to give rise to trophoblast-like cells but require an unpredicted dedifferentiation process to the earlier inner cell mass (ICM) state[12,45]. While it is clear that we still have gaps in our knowledge

7

regarding the potencies of different human stem cell types, researchers have been developing hPSC-based embryoid models that yield complex multicellular structures that recapitulate different aspects of the human embryonic development.

1.3.4 Embryoid models using human pluripotent stem cells

During implantation of the human blastocyst, the epiblast polarizes and undergoes lumenogenesis to form the pro-amniotic cavity (Figure 1.3). The lumenal epiblast structure subsequently undergoes symmetry breaking, leading to the formation of the bipolar amniotic sac, with squamous amniotic cells at the uterine-proximal pole and the epiblast comprising the embryonic disc at the hypoblast-proximal pole (Figure 1.3). Soon after the formation of the amniotic sac, gastrulation is initiated with the formation of the primitive streak at the prospective posterior end of the epiblast, resulting in the formation of the three germ layers and establishment of the body axes. Researchers have successfully leveraged the developmental potential and self-organizing property of hPSCs to develop embryoid models that help break down these developmental stages and enable mechanistic studies[28,33,34].

Simple and controllable human embryoid systems can be used to recapitulate a variety of developmental events. Combined with single-cell analysis and mathematical modeling, they can help us understand how processes at the subcellular level have a direct effect on tissue-level emergent behaviors. Warmflash *et al.* [28] used microcontact printing to create the first human embryoid model in which primed hPSCs are constrained to 2D circular adhesive islands (Figure 1.5a) . Uniform supplementation of BMP4 into culture medium resulted in ring-shaped gene expression patterns, with CDX2+ extraembryonic cells on the colony edge followed by mesoendoderm and ending with ectoderm in the colony center, mimicking certain aspects of germ layer patterning during the human gastrulation. Varying pattern size, an edge sensing

mechanism was discovered to be responsible for patterning in the 2D human gastrulation model. The use of knockout lines and computational modeling revealed that this edge sensing mechanism involved gradual cell polarization and diffusion of the BMP4 inhibitor NOGGIN[31]. Further work in the 2D human gastrulation model revealed that WNT and ACTIVIN-NODAL signaling act downstream of BMP4 to control primitive streak formation, timing of epithelial-mesenchymal transition (EMT), and patterning of mesoderm and endoderm (Figure 1.5a)[29]. A 2D partial differential equation (PDE) model was generated to show that their system dynamics could be described by waves resulting from bistability[30].
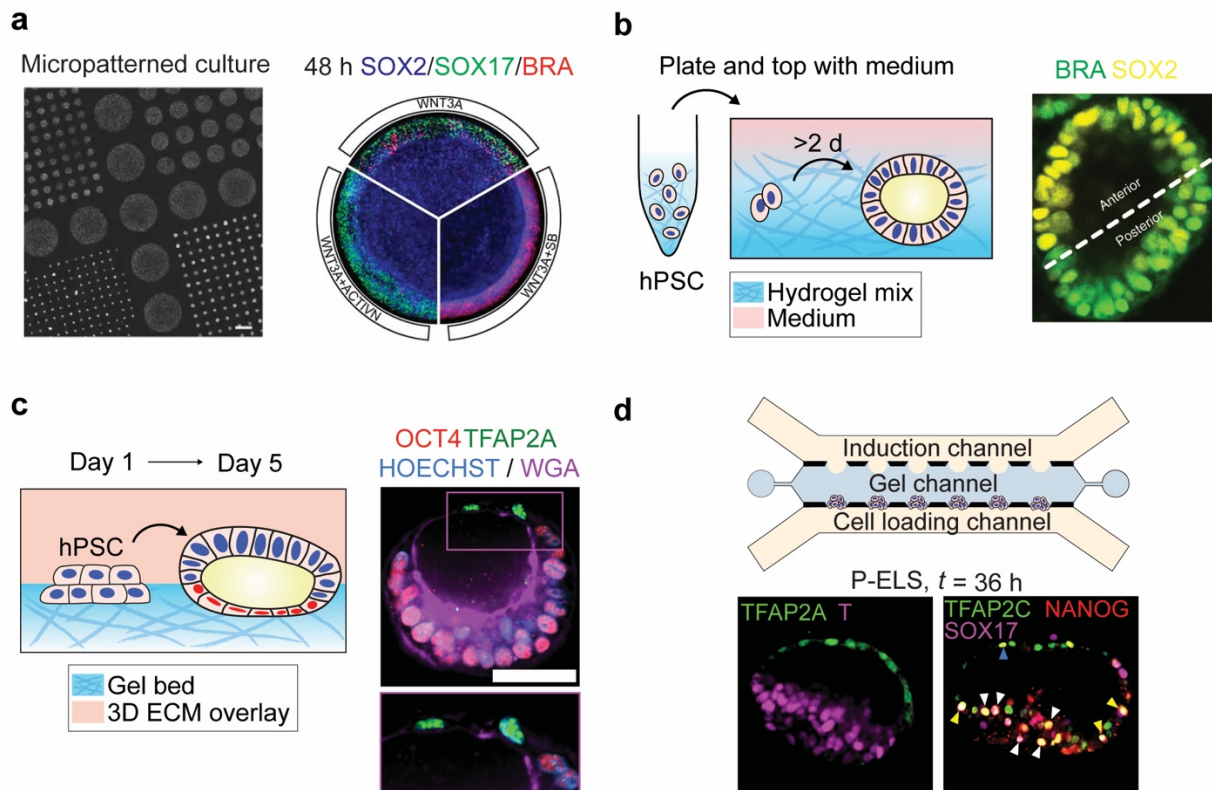
Figure 1.5 (a, left) hPSCs in micropatterned culture[28]. Scale bar, 500 μm. (a, right) hPSCs grown in 1000 μm micropatterns, stimulated with WNT3A, WNT3A+SB or WNT3A+activin and stained after 48 h for germ layer markers[30]. (b, left) Protocol used by[34] for creation of 3D environment. (b, right) Cyst with asymmetric expression of SOX2 (yellow) and BRA (green) resulting from BMP4 treatment. (c, right) Schematic of the biomimetic system used by [30] showing the formation of asymmetric cysts from hPSC. (c, left) Day 5 asymmetric cyst, stained for TFAP2A (green), OCT4 (red), and WGA (purple). HOECHST (blue) counterstains nuclei. Scale bar, 50 μm. (d, top) Schematic of microfluidic device designed by[46], showing three parallel channels partitioned by trapezoid-shaped supporting posts. (d, bottom) P-ELS at 36 h stained for TFAP2A (green) and T (purple) (left) and NANOG (red) and SOX17 (purple) (right). TFAP2C+SOX17+ hPGCLCs are marked by color-coded arrowheads to indicate

spatial localizations (blue, amniotic ectoderm-like compartment; yellow, amniotic ectoderm– epiblast junction; white, PrePS-EPI-like compartment).

While 2D embryoid models are more amenable to live imaging and quantitative measurements and perturbations, spatial constraints of hPSCs on 2D lead to a lack of 3D cellular architecture and morphogenetic events that are needed to model the 3D human embryo development. To address this issue, Shao *et al.* [33] engineered a 3D biomimetic platform with a soft gel bed made with the basement membrane matrix Geltrex[TM] and a 3D overlay made with a low concentration of Geltrex[TM] diluted in culture medium (Figure 1.5c). hPSCs plated as single cells cluster and undergo lumenogenesis to form pluripotent cysts enclosing a central lumen, mimicking the formation of the pro-amniotic cavity in the epiblast of the peri-implantation human embryo. Importantly, a subset of these cysts undergo symmetry breaking and yield the post-implantation amniotic sac embryoid (PASE). This asymmetric cyst consists of a central lumen, enclosed by amniotic cells at one pole and epiblast-like cells at the opposite pole (Figure 1.5c), resembling the post-implantation human amniotic sac structure (Figure 1.3). The PASE was further able to recapitulate key events of gastrulation including basal lamina breakdown, EMT, mesoderm induction, and cell dissemination. BMP-SMAD signaling was found to be important for amniotic differentiation of hPSCs [33]. However, the low yield of the asymmetric PASE structure and the lack of controllability of its symmetry breaking and embryonic-extraembryonic axis formation makes in depth mechanistic studies difficult. More recently, Simunovic *et al.*[34] developed a 3D hydrogel environment in which to study BMP4-induced symmetry breaking events in lumenal pluripotent cysts formed by primed hPSCs (Figure 1.5b). By varying BMP4 concentration, Simunovic *et al.* were able to identify a condition that resulted in spontaneous patterning with SOX2+ and BRA+ cells occupying two opposite poles of the cyst

(Figure 1.5b). Progressive development of this 3D human embryoid model further displayed the formation of a primitive streak-like structure, cell dissemination, and mesoderm/endoderm marker expression.

Mechanistic studies using embryoid models would benefit from platforms that allow for precise control of system parameters and for a clear understanding of how the cells and the structures they form are oriented with respect to their surroundings. Zheng *et al.*[46] recently developed a microfluidic platform that provides a highly controllable 3D biomimetic environment for the PASE development (Figure 1.5d). This microfluidic device contains three parallel channels, with an induction channel and a cell loading channel separated by a central gel channel. Addition of Geltrex™ to the gel channel leads to the formation of pockets in which hPSCs will settle and aggregate. The size of the initial cell clusters can be controlled by modulating initial cell seeding density. Both the induction and cell loading channels can be used for exogenous chemical stimulations. Notably, exposing clusters of primed hPSCs to BMP4 through the induction channel leads to the formation of a posteriorized PASE or posteriorized embryonic-like sac (P-ELS) (Figure 1.5d), which begins with patterning of the lumenal pluripotent hPSC cysts where the cells facing BMP4 differentiate into amnion-like cells (AMLCs) and the cells at the opposite pole show markers suggesting a pre-primitive streak epiblast (PrePS-EPI) phenotype. NANOG+TFAP2C+SOX17+ human primordial germ cell-like cells (hPGCLCs) are evident and scattered throughout the posteriorized PASE structure. Progressive development of this structure showed gastrulation-like events, with EMT, mesoderm induction, and cell dissemination from the PrePS-EPI compartment. The microfluidic PASE model has shown superior controllability and reproducibility for recapitulating consecutive hallmarks of early human post-implantation development.

11

1.3.5 Mouse embryoid models incorporating extraembryonic stem cells

The 3D hPSC-based embryoid models lack extraembryonic lineages associated with the trophoblast and the hypoblast, which may be critical for the proper organization and continuous development of these human embryoid models. Given the availability of mouse trophoblast stem cells (mTSCs) and extra-embryonic endoderm (XEN) cells[47], which represent the stem cell populations of the extraembryonic ectoderm and the primitive endoderm in the mouse embryo, researchers have created mouse embryoid models that incorporate these extraembryonic cell lineages. Rivron et al.[8] co-cultured mouse ESCs (mESCs) with mTSCs in non-adherent hydrogel microwells to induce the cells to self-organize into structures mimicking the mouse blastocyst (blastoid; Figure 1.6a). Harrison et al.[6] co-cultured mESCs and mTSCs in a 3D Matrigel culture to generate a mouse embryoid model with mESC- and mTSC-derived compartments (coined ETS embryoid model) to recapitulate certain events of the post-implantation mouse development, including mesoderm and PGC induction. Adding mouse XEN cells to the ETS embryoid using an inverted pyramid microwell plate resulted in the ETX embryoid, showing adjoining mTSC- and mESC-derived compartments surrounded by XEN cells [9]. Importantly, the ETX embryoid showed an improved efficiency and was able to recapitulate axial mesoderm and definitive endoderm specification in the mouse embryo.
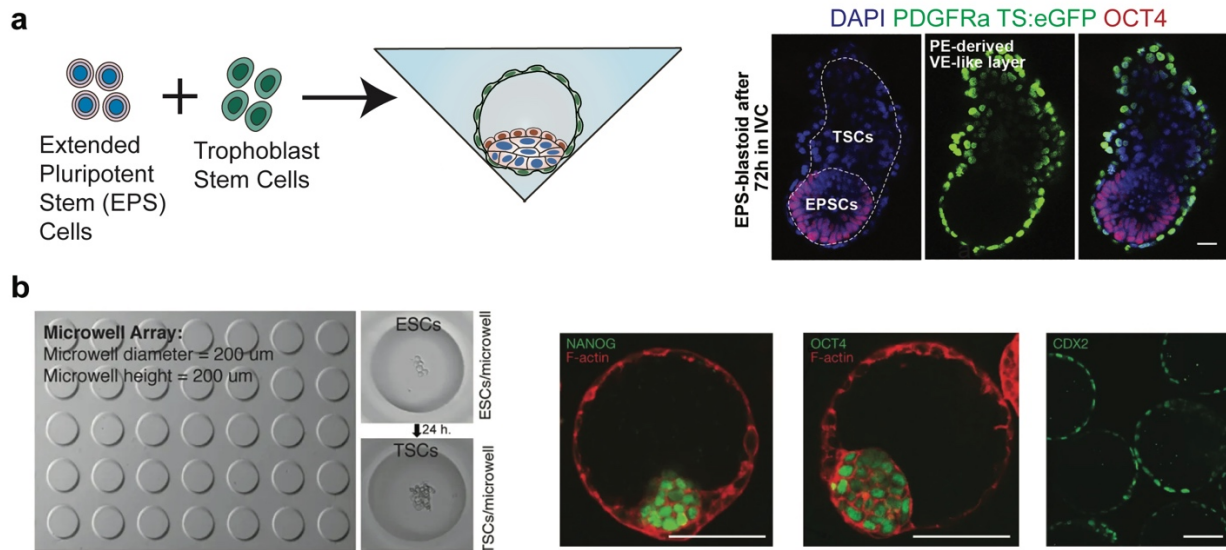
Figure 1.6 (a, left) Platform used by[48] for generating EPS-blastoids. (a, right) EPS-blastoid at 72h showing an implantation-like structure post-implantation-like structure with TS cell and EPS cell compartments. (b, left) Non-adherent hydrogel microwell arrays used by[8] for co-culture of mPSCs and TS cells. (b, right) Immunofluorescent staining of blastoids showing expression of NANOG (green, left), OCT4 (green, middle), and CDX2 (green, right). Scale bars, 50 μm.

Researchers have begun to use mouse EPSCs (mEPSCs) to generate mouse embryoid models. The first attempt by Li *et al.*[10] showed that mEPSCs cultured in inverted pyramid arrays could spontaneously give rise to a blastoid structure capable of implanting *in utero*. In a more recent work, Sozen *et al.*[48] co-cultured mEPSCs with mTSCs in inverted pyramid arrays (Figure 1.6b) to obtain a blastoid structure, in which a mTSC cyst enclosing the blastocoel-like cavity had an internal mEPSC compartment. The cavity was then flanked by mEPSC-derived primitive endoderm-like cells. Importantly, this blastoid was able to develop further and form an elongated, cylindrical structure with epiblast and extraembryonic ectoderm-like compartments covered by a visceral endoderm-like cell layer (Figure 1.6b).

The range of successive developmental events that these mouse embryoid models are able to recapitulate highlights the importance of stem cell potency and crosstalk between embryonic and extraembryonic stem cells to promote the proper organization and continuous

development of embryoid models. However, the development of extraembryonic mouse lineages is still an ongoing effort. The potential of mEPSCs to yield extraembryonic lineages has been challenged by a recent study[49]. Further, the aforementioned mouse embryoid models still suffer from low yields, which hinders their use for in-depth mechanistic studies. Regardless of current pitfalls, the success seen with mouse embryoid systems has set a precedent for the future development of human embryoid models. It is clear that the development of human embryoid models that more closely resemble the *in vivo* system will require the integration of extraembryonic lineages. However, there is still work to be done for the generation of human extraembryonic stem cell lineages. It also remains to be determined whether hEPSCs could be used to create human blastoid structures, like their mouse counterpart, the mEPSCs.

1.4 In silico models of human development

1.4.1 Agent-based models

As has been discussed, human embryonic stem cells have an amazing ability to self-organize and give rise to extremely complex structures both in form and function. But how do morphogenesis and patterning occur in such a robust manner? Given the low accessibility to human embryos and low efficiency of many stem cell-based models, how do we uncover the mechanisms at work behind cell migration, growth, and fate specification? The answer lies in the use of mathematical and computational models.

The start of morphogenetic observations can be traced back to D'Arcy Thompson's work *On Growth and Form*, in which he established similarities between the shapes seen in biological systems and mechanical systems[50]. Years later, Alan Turing presented a theoretical explanation, the reaction-diffusion (RD) mechanism, in which diffusion or 'Turing' instability is able to yield periodic patterning in a two-molecule reaction system[51]. Many of the pattern formation models

that have been developed since Turing's work rely on the same principles of short-range activation and long-range or lateral inhibition[52,53]. In this mechanism, a fast-diffusing inhibitor coupled with a slow-diffusing activator leads to nonlinear reaction dynamics that result in repetitive patterns[54]. RD models typically utilize systems of PDEs that are solved using numerical approaches. They are often used to represent intra- and extracellular molecular binding and diffusion[55–59]. Simple PDEs can be used to study the effect of morphogen gradients. However, modeling morphogenesis of complex structures such as the blastocyst or a developing limb bud becomes difficult because of PDE's difficulty in capturing the heterogeneous environments in which cells exist[60]. In the 1970s, Lewis Wolpert proposed the mechanism of positional information (PI) as a driver of pattern formation in morphogenesis[61]. In his proposed mechanism, a cell is able to determine its position relative to the other parts of the organism by sensing the local concentration of a morphogen. And based on this understanding of its relative position in the structure, it will undergo fate specification[62,63]. The mechanisms of RD and PI still form the basis for many of the current approaches to the study morphogenesis[60].

Agent-based modeling (ABM) is a cell-based modeling approach that can be seen as a consolidation of the two aforementioned theories[60]. Originally used to study the dynamics of replication[64], agent-based models including cellular automata (CA), cellular Potts, and hybrid models have been used to study a variety of biological systems including biochemical reaction networks, stem cell proliferation and differentiation, tumor angiogenesis, and metastasis[65–68]. ABM is a discrete modeling approach with three main components: agents, the environment, and rules. The agents are autonomous entities that can interact with each other and the environment. The environment is a two- or three-dimensional space in which agents exist. And the rules dictate what action an agent will take based on its local input. ABMs are flexible and permit for

scaling so that each agent can be as large as a group of organisms or as small as a subcellular membrane component. Because of this range, ABMs have been used to study a variety of processes including tissue electrical conduction, cell trafficking, tissue mechanics, immunomodulation, arterial remodeling, and inflammation[69–76].

As opposed to the more traditional theoretical models of morphogenesis and patterning, ABMs can capture the discrete nature of cells by modeling each cell as an individual agent. This cell-based approach is particularly compatible with morphogenesis, a decentralized process in which events at the single-cell level lead to complex global form and patterning. Multicellular systems are complex, with multiple processes occurring in series and in parallel, across different temporal and spatial domains to guide growth, migration, and differentiation. The scale at which a system will be modeled and the degree of detail or abstraction in an ABM depends on the object of study and what is known about it. In cases in which the system is poorly understood, a successful approach begins by simplifying the system to its bare essentials and gradually adding more components as it is interrogated and understood[77]. Further, not all model parameters required to build the model will be measurable. For this reason, it is important to explore the use of parameter estimation techniques to better parameterize the models.

1.4.2 Genetic algorithms for model construction

The problem of large parameter space exploration is one encountered in all fields interested in the study of complex systems. Oftentimes, the exploration of all possible combinations in parameter space is intractable. So how do we structure parameter space exploration so that optimal solutions can be found within acceptable timeframes? Genetic algorithms (GAs) are a parameter space exploration method based on the mechanisms of natural selection and genetics developed by Holland in the 1970s[78]. Key questions he was interested in

exploring were (1) how does evolution work to create organism of increasing fitness and (2) how can we build problem-solving programs based on what should happen rather than how it should happen[79]. In evolution, the adaptation of an organism involves the progressive change of structures based on environmental pressures. As a process, adaptation has three main components: the environment of the adapting system, the adaptive plan enacting change in the system, and the measure of performance or fitness of a system[79]. As explained by Holland in his effort to formalize GAs as a tool for parameter space exploration, a successful adaptive plan must be able to retain advances as well as progressively increase the proportion of fit structures[79]. This adaptive plan can be seen as a genetic plan through which a population of organisms or systems are allowed to reproduce and undergo genetic changes.

Analogous to evolution, GAs work on the basis of populations of candidate solutions[80]. Each candidate solution constitutes an 'organism' whose function or behavior is dictated by a genetic code. Parameter space will be explored as the genetic code of the organisms in the population evolves via the following steps[80]:

1. Initialization: the creation of an initial population drawn randomly from parameter space.

2. Evaluation: the fitness of each organism is assessed.

3. Selection: organisms are chosen to reproduce with a probability proportional to their fitness.

4. Crossovers: new organisms are generated by combining sections of two or more existing organisms.

5. Mutations: the genetic traits of an organism are randomly modified with some probability.

6. Replacement: the current generation is replaced by the new generation obtained after selection and genetic modifications.

7. Repetition of steps 2-6 until a satisfactory solution is found or a terminating condition is met.

An important consideration during initialization is population size. While small populations might lead to premature convergence on substandard solutions, large populations might require unreasonable amounts of computational time. For the evaluation, the fitness of an organism could be determined by an objective function involving a mathematical model or a subjective function determined by the user[80]. During selection, organisms of higher fitness have a higher probability of being chosen for the next generation. The two genetic operators, mutations and crossovers, serve distinct purposes. In a crossover, a new structure or organism is formed by exchanging a set of attributes (i.e. a section of genetic code). This process allows for the thorough exploration of local optima. During a mutation, each position or trait encoded in the genome has some probability of being randomly changed to a different value. The existence of this mechanism ensures that no trait can permanently disappear from the genetic pool[79]. Because the framework of GAs can be applied to any structure or system, since its inception the use of this tool has spread to a myriad of fields including optimization, automatic programming, machine learning, economics, immune systems, ecology, and social systems among others[81]. In this work, GAs are utilized in the context of artificial intelligence. The systems or 'organisms' being evolved are agent-based models. And the fitness measures are based on how well a model is able to yield a desired structure with a desired pattern.

1.5 Research topics and dissertation outline

In this work, I sought to study the emergence of form and function in the context of early human development, specifically focusing on our *in vitro* model of the epiblast cyst. Presented in Chapter 2, the first part of my work focuses on characterizing our *in vitro* model by identifying individual cells and recording cell-level events. Presented in Chapter 3 and Chapter 4, the second part of my work focuses on the construction of a computational framework for generating models of growth and patterning. With the combined use of these tools, we hope that future work will increase our understand of how cell-level decisions can lead to the emergence of complex patterns.

In **Chapter 2**, I present a pipeline for the characterization of morphogenesis from live cell imaging data. This machine learning-based analysis tool incorporates image pre-processing, cell tracking, and identification of cell state changes such as division and death. This tool enables us to parse the relationship between the properties of the local environment and cell-level decisions that lead to emergent behaviors like hPSC cyst formation and growth.

In **Chapter 3**, I present a framework for the construction of *in silico* models of development. Using a combination of agent-based modeling (ABM) and genetic algorithms (GA), we are able to generate computational models capable of giving rise to growth and patterning.

In **Chapter 4**, I demonstrate the use of the ABM-GA framework for the study of morphogenesis and patterning of the post-implantation amniotic sac embryoid (PASE) model. The models generated using evolution are tested in a variety of *in silico* experiments and are shown to be able to capture some of the dynamics seen *in vitro*.

Finally, in **Chapter 5**, I summarize this dissertation work and provide perspectives on future efforts in combining the use of *in silico* modeling, machine learning, and stem cell-based *in vitro* models of development for the study of embryogenesis and organogenesis.

Chapter 2 Machine Learning-Assisted Imaging Analysis of a Human Epiblast Model

The work in this chapter has been published in Integrative Biology[82].

2.1 Introduction

Human embryo development is a complex process in which cells go through major reorganization and progressive fate specification. Pre-implantation human development leads to the formation of the blastocyst, a hollow sphere of trophectoderm cells with an inner cell mass (ICM) composed of both epiblast cells (i.e. embryonic stem cells), which will later form the embryo proper, and hypoblast cells, which will later go on to form the yolk sac. Once the blastocyst begins implantation into the uterine wall, there are a number of developmental events all working in parallel and affecting each other in ways we still don't understand. These processes include the invasion of trophectoderm cells into the uterine wall and their differentiation into cytotrophoblast and syncytiotrophoblasts as well as the development of the epiblast into a lumenal rosette structure enclosing a central cavity. Soon after, the epiblast cells next to invading trophectoderm cells differentiate into the amnion, with the remaining epiblast cells next to the hypoblast remaining pluripotent, leading to the formation of a bipolar epiblast-amnion tissue. While crucial to a successful pregnancy, these developmental events are difficult to study due to both technical limitations and ethical considerations[19,83]. For years, researchers have tried understanding human development with the use of animal models including mouse and monkey models[1,2,9,48,84]. Recently, there have been increasing efforts towards the

development of *in vitro* models of human development with the use of human pluripotent stem cells (hPSCs) including human embryonic stem cells (hESCs)[21,85–87].

Studies have shown that hPSCs have an intrinsic property to self-organize and differentiate to form complex *in vivo*-like structures. Leveraging this capability, researchers have successfully created a variety of hPSC-based embryo models that recapitulate key steps in early human development[28,33,34,46]. A developmental process of particular interest to our group has been the formation of lumenal hPSC cysts and the differentiation of hPSCs into amnion cells. Shao *et al.*[33]were the first to show that hPSCs could differentiate into amnion cells. They engineered a 3D biomimetic platform with a soft gel bed made with the basement membrane matrix Geltrex™ and a 3D matrix overlay made with a low concentration of Geltrex™ diluted in culture medium. In this system, hPSC clusters would undergo lumenogenesis and form lumenal structures containing a central cavity. Over time, three types of cysts resulted from lumenal hPSC clusters: cyst composed of amnion cells, cysts composed of undifferentiated hPSCs, and asymmetric cysts containing amniotic cells at one pole and undifferentiated hPSCs at the opposite pole (Figure 2.1). The percentage of each type of cyst was shown to depend heavily on the initial cell plating density. While BMP-SMAD signaling was found to be important for amnion differentiation, the mechanism(s) that led to the initiation of amnion differentiation in the 3D structure has remained elusive.

hPSC cluster

Cell plating density
Biochemical signals
Biomechanical signals

Amniotic cyst | Pluripotent cyst | Asymmetric cyst
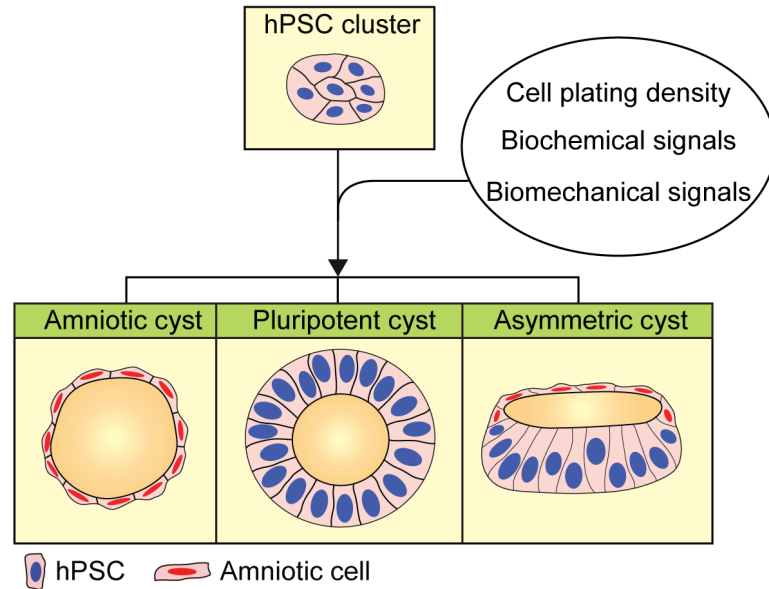
hPSC   Amniotic cell

Figure 2.1 Development of hPSC clusters into three distinct types of lumenal structures: amniotic cyst, pluripotent cyst, and asymmetric cyst. Amniotic cysts are composed of amniotic cells, whereas pluripotent cysts are composed of undifferentiated hPSCs. Asymmetric cysts contain amniotic cells at one pole and undifferentiated hPSCs at the opposite pole.

The ability of hPSCs to self-organize and differentiate into *in vivo*-like structures in *in vitro* settings posits the existence of endogenous developmental programs. Consequently, a crucial characteristic of *in vivo*-relevant, stem cell-based embryo models is their ability to leverage these programs in order to capture the progressive nature of human development. Triggering these developmental programs, however, is not a trivial endeavor; it requires cell culture environments engineered with correct dimensionality as well as correct mechanical and biochemical properties. Having taken the necessary first step of creating a hPSC model that recapitulates a developmental period of interest, the next step becomes the elucidation of the mechanisms at work in the system. *In vivo*, progressive development entails branching of distinct lineages and progressive differentiation into cell types with increasingly restricted potential[88]. Studying these processes in hPSC models in a tractable manner requires the use of computational tools that minimize manual curation and bias. Machine learning tools have come to the forefront

and are increasingly used to parse the mechanisms at work in these systems. To date, however, many of these efforts have been directed towards the application of single-cell RNA-sequencing (scRNA-seq) data analysis tools[89–93] or examining global features of the observed structures at discrete time points[94,95]. While these approaches are useful, their discrete nature limits their use for understanding how factors in the local cell microenvironment trigger and guide cell state changes that lead to the emergence of relevant structures. Understanding this requires the ability to continuously monitor individual cells in the system and record division events for later lineage tracing.

There have been several efforts towards the creation of classifiers for the identification of dividing cells. While the methods are varied, the models can be divided into two categories: (1) models that use spatial features[96–98] and (2) models that use both spatial and temporal features[99–105]. Many of the models that rely only on spatial features for classification utilize morphological feature extraction that leverages the clear differences in visual characteristics between dividing and non-dividing cells[96–98]. Models that lack temporal information, however, face the additional challenge of having to consider how the timing at which the event is captured will affect the features of interest. This is not an issue for spatiotemporal models in which many stages of the division process can be captured and used for the classification. However, rather than focusing on the nucleus, which shows the most obvious visual changes during division, many of the existing models rely on phase contrast microscopy images[99–102,104,105] that complicate classification because of the confounding factor of varying cell shape. In this work, we present a computational tool for the comprehensive analysis of live cell imaging data of hPSC cyst formation using a unique nuclear GFP H9 hESC reporter line. Using Python, we created a pipeline that is able to process the images and identify all individual cells in a developing hPSC

cyst. The pipeline captures information on both the cell properties and cell neighborhood at each time point. Further, we trained a machine learning model for event recognition that is able to identify changes in cell state such as division and death by looking at spatiotemporal properties of the nuclei. With this tool, we hope to parse the relationship between the properties of the local environment and cell-level decisions that lead to emergent behaviors like hPSC cyst formation and growth.

2.2 Materials and Methods

2.2.1 Cell culture substrate preparation

An array of 100 μm-diameter circular adhesive islands was created using a two-step micropatterning method as described previously[106]. Briefly, a poly-dimethylsiloxane (PDMS) elastomeric stamp with an array of circular posts was generated using replica molding from a silicon mold fabricated by standard photolithography and deep reactive ion etching (DRIE)[107,108]. The center-to-center spacing between adjacent posts on the PDMS stamp was 150 μm, and the post height and diameter were 30 μm and 100 μm, respectively. The PDMS stamp was coated in 1% Geltrex (Thermo Fisher Scientific; derived from Engelbreth- HolmSwarm tumors similarly as Matrigel®) solution for 24 h at 4 °C and subsequently rinsed with distilled water and blown dry with nitrogen. Before stamping, the cell culture substrate was prepared by coating a glass coverslip with PDMS and treating it with ultraviolet (UV) ozone (UV-ozone cleaner; Jelight, Irvine, CA) for 7 min to oxidize the PDMS surface. The PDMS stamp was then placed in conformal contact with the PDMS-coated coverslip for 5 s to transfer Geltrex from the stamp to the coverslip. To restrict cell attachment to the circular adhesive islands, the coverslip was treated with Pluronic F127 NF dissolved in PBS (0.2%, *w* / *v*; BASF, Ludwigshafen, Germany) for 1 h at room temperature and rinsed with distilled water. The coverslip was then immersed in

mTeSR (STEMCELL Technologies) for a minimum of 2 h to further block the non-functionalized surface of the coverslip. Finally, the coverslip was submerged in mTeSR medium containing 1% Geltrex for 1 h. The coverslip was washed with PBS before cell seeding.

2.2.2 Generation of mTnG cells

For live cell imaging of hESC cyst formation, a membrane tdTomato, nucleus-EGFP (mTnG) H9 hESC line was generated. H2B-EGFP was PCR amplified from a gift plasmid Tcf/Lef:H2B-GFP (Addgene plasmid #32610). The PCR product was then ligated into the ePiggyBac vector with a constitutively active puromycin selection cassette[109]. membrane-tdTomato was PCR amplified from a gift plasmid pQC membrane TdTomato IX (Addgene plasmid #37351). The PCR product was then ligated into the ePiggyBac vector with a constitutively active neomycin selection cassette[109]. These two plasmids (1.5 µg each) were co-transfected with 1 µg pCAG-PBase (ePiggyBac transposase helper plasmid obtained from Dr Ali H. Brivanlou[109]) using GeneJammer (Agilent Technologies) into H9 hESCs that were plated at 50,000 cells cm^{-2} 24 h prior to transfection. Puromycin selection (2 µg mL^{-1}) and G418 selection (250 µg mL^{-1}) started at 4 days after transfection. The cells were selected for 7 days. After selection, the cells were dissociated to single cells and replated at low density (400 cells cm^{-2}) for clone picking. 12 clones were hand-picked and evaluated for brightness and pluripotency. 3 clones were expanded at the end (mTnG #1, 2, 3). mTnG #1 hESC line has the brightest fluorescent signal and is used in the current study.

2.2.3 Cyst formation assay

Cultured hESC colonies were dissociated into single cells with Accutase (Sigma-Aldrich) at 37 °C for 10 min before the cells were centrifuged and re-suspended in mTeSR1 medium

containing 10 μM ROCK inhibitor, Y27632 (Tocris), to avoid dissociation-induced apoptosis[110].

Cells were then plated onto coverslips pre-coated with circular adhesive islands at a density of

300,000 cells cm^{-2}. To establish 3D ECM overlay, culture medium was changed to fresh

mTeSR1 medium containing 10 μM Y27632 and 4% (*v* / *v*) Geltrex 2 h after initial cell seeding.

Y27632 was removed 24 h after initial cell seeding, at which time the coverslip was transferred

to fluorescence microcopy for live cell imaging.

2.2.4 Live cell video acquisition

mTnG hESCs on the coverslip were imaged using the Zeiss Axio Observer Z1 inverted

epifluorescence microscope enclosed in the XL S1 incubator (Carl Zeiss MicroImaging) to

maintain cell culture at 37 °C and 5% CO_2. Fluorescence images were recorded with a 20×

objective for a period of 24 h, with an exposure time of 3 s and a time frame of 10 min to

minimize phototoxic effects on cells. A GFP filter set was used for fluorescent imaging of the

nuclei of mTnG hESCs.

2.2.5 Image pre-processing

A customized Python program was used to process raw images collected from live cell

imaging using fluorescence microscopy. First, contrast was enhanced using adaptive image

enhancement developed by Peng *et al*[111]. Specifically, each pixel in the image is normalized

using the mean and variance of a local region surrounding the pixel. This local region is

determined adaptively. For a given pixel, the program starts from a given initial size and expands

until the standard deviation of the region is equal to or more than a given threshold. For

computational tractability, maximum radius was set at 5 pixels. The threshold is in the range of

0.2 to 0.8 and is meant to ensure that the local region has enough relevant structures to classify a

pixel as being part of the background or part of an object. After obtaining the region size, the pixel is normalized by subtracting the local mean and dividing by the local standard deviation. This will account for varying background intensity and varying contrast, respectively. A background mask is then obtained by binarizing the resulting image with a binary threshold (cv2.THRESH_BINARY). This background mask is further refined with a dilation (cv2.dilate, kernel size = (3,3), iterations = 2) followed by an erosion (cv2.erode, kernel size = (5,5), iterations = 2). Multiplying this resulting background mask by the original image eliminates background noise. Second, contours of cells were identified with the use of adaptive gaussian threshold (cv2.adaptiveThreshold with blockSize = 23, C = 1) (Figure 2.3b). For each pre-processed image, contours were extracted with the use of cv2.findContours with cv2.RETR_TREE and cv2.CHAIN_APPROX_NONE. The third step is to carry out segmentation to find individual cell contours (Figure 2.3d). The pipeline measures the area and circularity of each contour. Contours identified as individual cells are stored. The contours identified as cell clusters undergo concavity point-pair segmentation, a method developed by Farhan *et al.* based on finding concavity point-pairs using a variable-size rectangular window[112]. In brief, using an established interval, a list of contour coordinates is first extracted from the binary image of the cell cluster. For each coordinate in the list, lines are drawn to the next two points in the list. Once a line passes through the image background (*i.e.*, a pixel with value 0), the algorithm finds the contour coordinate at which the line no longer passes through the background and establishes this coordinate as a concavity point. After filtering the resulting point list to account for contour irregularities, the program finds the directionality vector of each concave area. Using this vector, each concavity point establishes a rectangular window in which to search for other concavity points. Once all concavity points have paired up, a line is drawn

between them and the cluster is segmented. Farah *et al.* validated the method with the use of three data sets, two of which contained bright field microscopy images of yeast cells, and one which contained fluorescent microscopy images of yeast cells[112]. They showed that the concavity point-pairing segmentation method was highly effective, with precision averaging at 0.98[112].

2.2.6 Image selection for CNN-LSTM

The machine learning classifier used was a deep learning model consisting of a convolutional neural network (CNN) connected to a long short-term memory (LSTM) network. The data set for supervised training contained sequences of three time points showing three different classes of cells: dividing, dying, and non-dividing. These sequences of dividing, dying and non-dividing nuclei were manually cropped from live cell videos (Figure 3a). The sample set contained 450 samples, with an equal amount of every class.

2.2.7 Parameters for CNN-LSTM

The CNN-LSTM model was constructed using keras.Sequential, which yields a linear stack of layers. The CNN layers consisted of a repeating pattern of convolution, max pooling, and batch normalization followed by one dropout and one global max pooling layer. The CNN model output for each sequence of images was passed on to an LSTM layer via a TimeDistributed layer. This layer extracts features from each image in the sequence and passes it to the LSTM. The final layers in the model create a fully connected network with the use of dense layers. Rectified linear units (ReLU) were used as the activation function in all of the convolutional layers and dense layers, except for the last one. The last dense layer used softmax activation in order to carry out multiclass classification. The model was compiled using Adam as the optimizer, categorical crossentropy for the loss calculation, and accuracy as the metric

evaluated by the model. The number of epochs was 100. During training, callback with ModelCheckpoint was used to store the best model based on validation accuracy. A 70-20-10 split was used to create the training, testing, and validation data sets. Because the data set is small, data augmentation was carried out using a data generator class. Using this generator, the data set was randomly shuffled and images were transformed using rotation (range of 5), shifts in height (range of 0.1) and shifts in width (range of 0.1). Transformations in this data generator class were carried out using the Keras ImageDataGenerator class. One-hot encoding was applied to the labels before training.

2.2.8 Video analysis pipeline

Live cell videos were analyzed with a Python pipeline. Images were pre-processed and all individual nuclei were identified. A cell tracker python class was used to give each cell a unique identification (ID) number and track cells from one time point to another using Euclidean distance. The nuclei were cropped from the image and stored in a Python dictionary. For event recognition, the cropped nuclei of the current time point and the cropped nuclei from the previous two time points were passed as input to the CNN-LSTM classifier. Whenever a new cell would appear in the environment, the parent cell would be identified using a parent score (ps) parameter. Newly divided cells tend to be small, bright, and similar in size. For this reason, the ps takes into account both the classification of the cells in the previous time point and the similarity in area and brightness of the nuclei between the new cell and the possible sister cell. The cell with the highest ps in the local neighborhood of the newly appeared cell would be assigned as the parent. At this point, the daughter cell with the parent ID receives a new ID, and the parent ID of the daughter cells is stored. The number of neighboring cells and the average distance to neighbors are stored for each cell at every time point. The output of the video cell

analysis is a Python DataFrame with the cell IDs, cell positions, number of neighbors, average

distance to neighbors and parent IDs for each time point.

2.3 Results

2.3.1 Cyst formation analysis pipeline

Analysis of the morphogenesis of a multicellular structure at discrete time points can give

insight into the system-level dynamics governing its growth and development. However,

important cell-level dynamics and the degree of stochasticity and heterogeneity in a multicellular

system remain difficult to elucidate. To better study the dynamic process of hPSC cyst

formation, we developed an experimental platform to provide a biomimetic niche for the

formation of hESC cysts in a controllable and robust manner. Specifically, an array of 100 μm-

diameter circular adhesive islands was created on a coverslip, before hESCs expressing

membrane tdTomato and nucleus-EGFP (mTnG) were seeded onto the coverslip. Two hours

after cell seeding, culture medium was changed to fresh mTeSR1 medium containing 10 μM

Y27632 and 4% (*v* / *v*) Geltrex, to establish a 3D ECM overlay. The coverslip was transferred to

a Zeiss Axio Observer Z1 inverted epifluorescence microscope 24 h after cell seeding. To track

the dynamics of hESC cyst formation, live cell imaging was conducted for 24 h with a depth of

focus that captured all the cells in the system. Given their self-organizing property, hESCs

confined on adhesive islands on the coverslip formed small clusters and underwent epithelization

and lumenogenesis. Throughout cyst formation, cells showed limited movement ability in the z-

direction. Live cell imaging data were then processed with the use of a Python pipeline capable

of image processing, cell tracking, and event recognition. With these extracted data, a

comprehensive characterization of cell states and actions during hESC cyst formation could be

conducted, using a workflow that includes image processing with machine learning, and the

characterization of cell state and cyst growth (Figure 2.2).
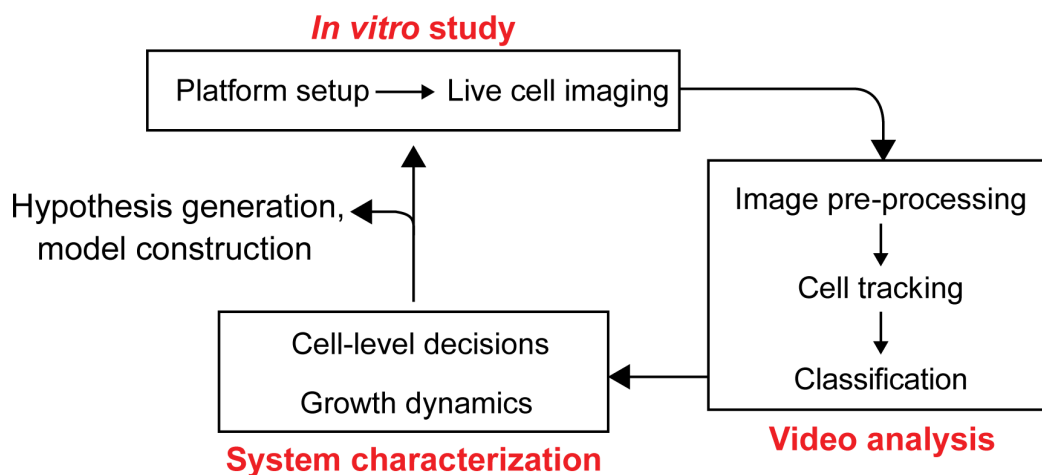
Figure 2.2 Workflow for live cell imaging data analysis of cyst formation, including *in vitro* experimentation, video analysis, and system characterization.

2.3.2 Event recognition and image processing

Parsing the relationship between cell actions and their local microenvironment is a

necessary step for elucidating the mechanisms that drive hESC cyst formation and development.

As a first step to carrying out this analysis, we sought to develop a machine learning model

capable of detecting two important changes in cell state: division and death. While CNNs are

often used for image classification, we sought to add robustness to the model by also leveraging

temporal information with the use of an LSTM network. CNN-LSTM has been utilized for

imaging analysis to detect mitotic cells recorded using time-lapse phase-contrast microscopy[102].

Following this logic and utilizing mTnG hESCs that show significant changes in nuclear shape

and area for both dividing and dying cells, we trained a CNN-LSTM classifier. The classifier

identifies three cell states: dividing, dying, and non-dividing (Figure 2.3a). While the use of a

single image could lead to correct classification, the CNN-LSTM model is able to leverage

information on the temporal changes in nuclear shape using live cell imaging data (Figure 2.3b).

By using a set of 450 manually labeled images with an equal number of each class, with a 70-20-10 test-train-validation split and data augmentation, a 96.3% overall accuracy in event recognition was achieved using the CNN-LSTM classifier (Figure 2.3c).
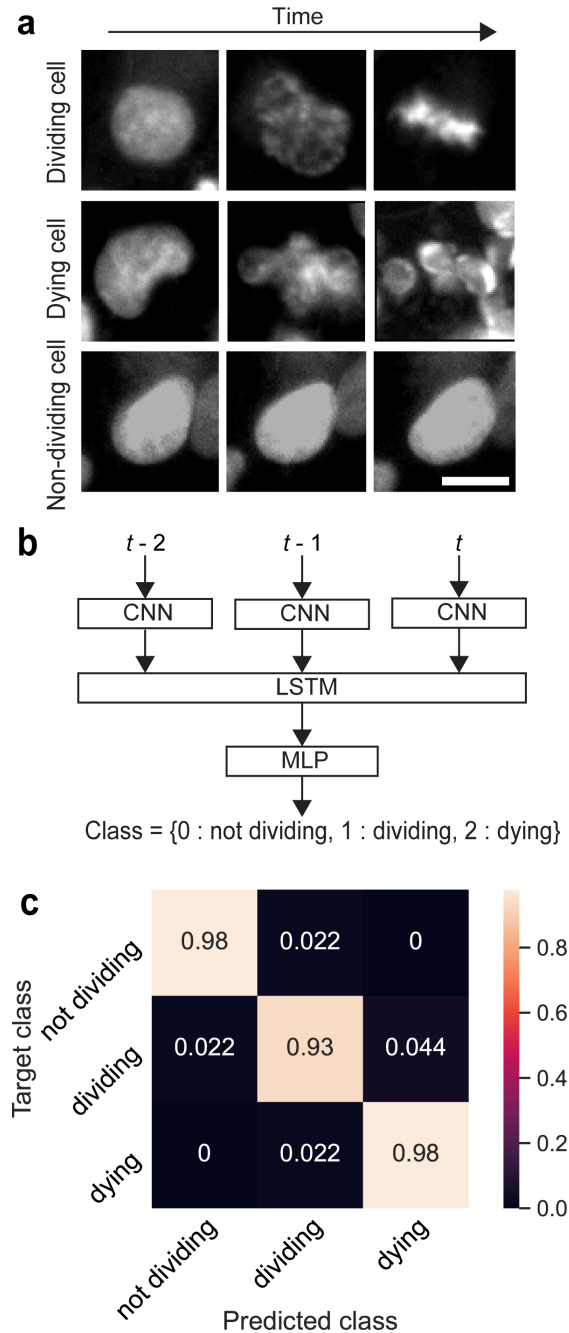
Figure 2.3 (a) Input image sequences to CNN-LSTM classifier. Images show the GFP channel of imaged mTnG H9 hESCs. The top, middle and bottom rows correspond to a dividing, dying, and non-dividing cell, respectively. Scale bar, 10μm. (b) CNN-LSTM framework followed by a multilayer perceptron (MLP) for multiclass classification. (c) Confusion matrix for CNN-LSTM classifier.

Having successfully trained a machine learning model for event recognition, we next sought to create an image processing pipeline capable of identifying individual cell nuclei. While there are a number of computational tools available for identification and tracking of cells in a multicellular system[113–115], identifying and tracking cells within a forming cyst presents unique challenges that require the use of more catered approaches. For example, in our test of the commonly used watershed method for segmentation, it is difficult to carry out correct segmentation of cell clusters, likely because of the compact arrangement of cells in hESC cysts. Additionally, global thresholding methods were unsatisfactory for finding hESC clusters, likely because of variations in nuclear GFP intensity. To address these challenges, we devised an imaging processing pipeline uniquely suited to carry out thresholding and segmentation in tightly packed hESC cysts. The image processing steps consist of thresholding, denoising, segmentation, and identification of individual cells in hESC cysts (Figure 2.4). For the segmentation step, we utilized adaptive local enhancement[111] to enhance the contrast between nuclei and background. For binarization, we utilized adaptive Gaussian thresholding to ensure that cells that are slightly out of focus can still be identified. After binarization and denoising, hESC clusters were segmented with the use of concavity point pairing analysis[112]. Cells are given a unique identification (ID) number and tracked from one time point to another with the use of Euclidean distance.
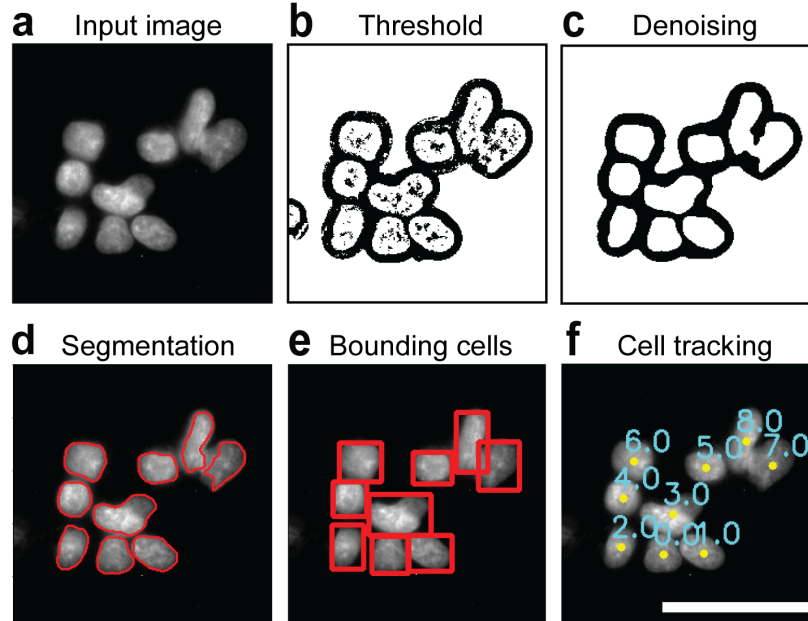
Figure 2.4 Image processing pipeline. Cell cluster is first isolated from original images obtained from live cell imaging of mTnG H9 hESCs (GFP channel is shown here) (a), before going through adaptive local enhancement and adaptive Gaussian thresholding (b). The image is then denoised with the use of open, erode, and filtering by connectivity (c). Cell clusters are segmented with concavity point analysis and individual cell contours are established (d). Bounding rectangle is then inputted into cell tracker (**e**), and cells are given a unique ID (f). Scale bar, 50 μm.

2.3.3 Live cell data processing and system characterization

As mentioned earlier, the experimental platform for the formation of hESC cysts consisted of an array of micropatterned Geltrex islands with a diameter of 100 μm generated with a two-step micropatterning process[106] (Figure 2.5a). After a period of 24 h in which mTnG hESCs were allowed to attach and cluster onto the adhesive islands, the experimental platform was transferred to a fluorescent microscope for live cell imaging for a period of 24 h (Figure 2.5b&c). To avoid cytotoxic effects, images were recorded at intervals of 10 min.

a

3D Geltrex overlay
Microcontact printed Geltrex

b

Plate hPSC
+Y27632   Fresh culture medium

-24 -22   0   24
+3D ECM

c $t = 0$ h $t = 6$ h $t = 12$ h $t = 18$ h $t = 24$ h

Membrane tdTomato / nucleus EGFP
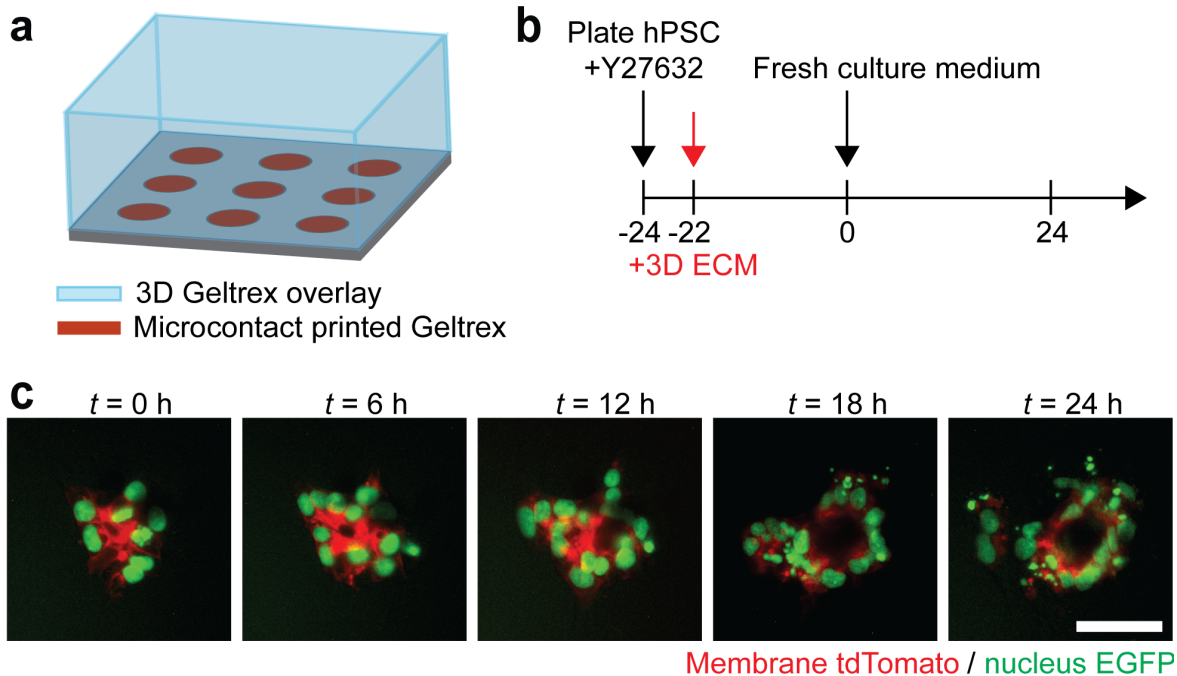
Figure 2.5 hPSC cyst formation experiment. (a) Cartoon of experimental platform consisting of micropatterned adhesive islands and a 3D Geltrex overlay. (b) Experimental protocol timeline. (c) Live imaging of a hPSC cluster at different time points. In this assay, the mTnG H9 hESC line was used. Scale bar, 50 μm.

Having obtained live cell data from various hESC cysts, we processed the images using the Python pipeline. Figure 2.6 shows the growth profiles of four different hESC cysts. The growth profiles vary greatly between the cysts. While there are periods of a sustained increase in cell number like the one seen in Figure 2.6a between 500 min and 750 min, we can also find periods of a sustained decrease in the number of cells as seen in Figure 2.6c between 0 min and 250 min. Regardless of the growth profile, however, the number of cells seem to plateau for all the cysts. While the final cell numbers might be similar among the four hESC cysts, there are a number of different growth trajectories that could not have been inferred from looking at the final cyst configuration. In the context of modeling human development, this information facilitates the study of how these changes in growth dynamics correspond to relevant cell specification events. Further, as cells progressively differentiate and more populations appear in

a system, we can start to study the growth dynamics of specific populations and how they relate

to correct form and function in the structure.
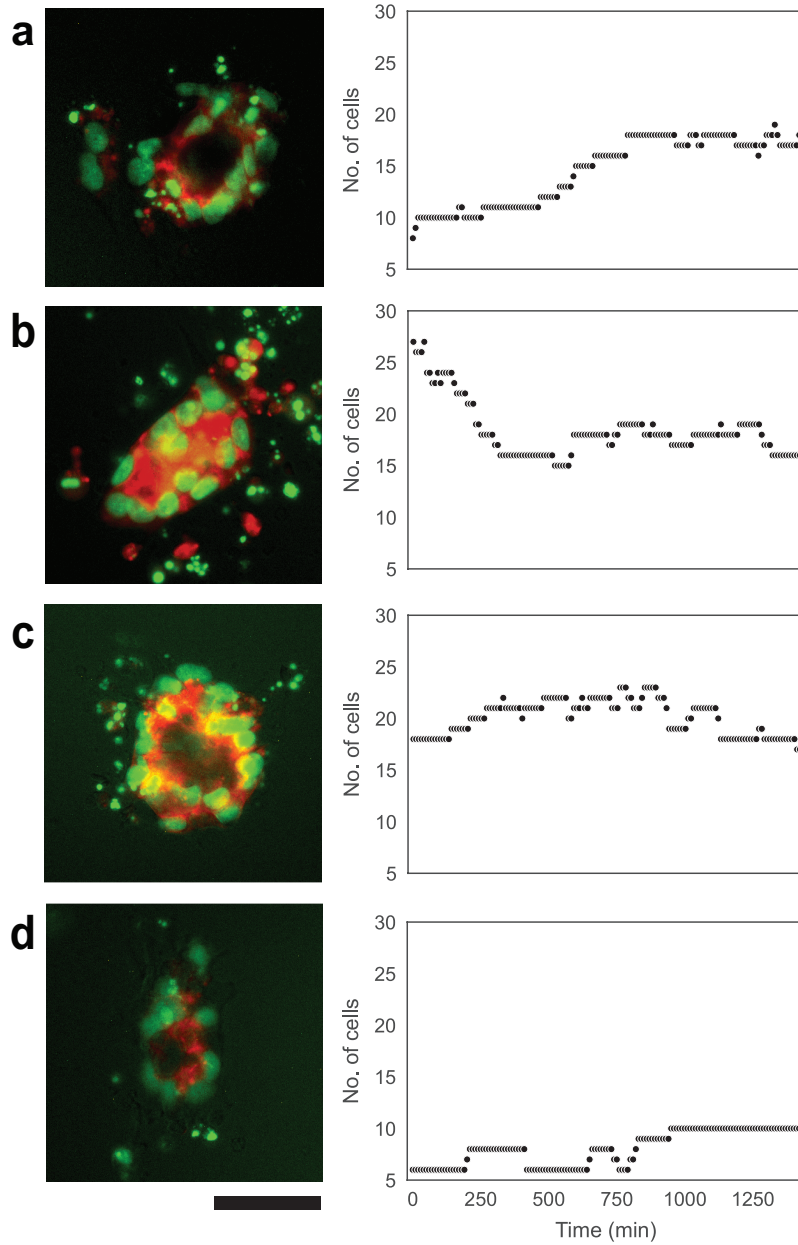
Figure 2.6 Growth dynamics of four different hPSC cysts. Left and right columns show the cysts after 24 hours of growth and the number of cells in the cyst through the period of 24 hours, respectively. In this assay, the mTnG H9 hESC line was used. Images show merged nuclear EGFP (green) and membrane tdTomato (red). Scale bar, 50 μm.

Lineage tracing is a powerful tool for parsing the mechanisms guiding morphogenetic

events in a multicellular system. It has many uses including providing insight into the timing of

differentiation of cell types of interest and helping identify lineage-specific precursor cells.

Combined with the ability to record properties of the local cell microenvironment, it can help

parse when and why different cell types arise. Figure 2.7a shows a network representation of the

cyst shown in Figure 2.6a at different time points. Edges between cells of a given cyst are

connected to each other based on an established threshold distance. The solid line going from

one time point to another indicates a chosen cell lineage, with an additional dotted line indicating

cell divisions. Having tracked a cell and established its lineage, we can characterize the local cell

density experienced by the cells in the lineage throughout time. As can be seen in Figure 2.7b,

the number of cells in the neighborhood and the average distance from neighbors of the selected

cell continue to vary even after cell number in the cyst has plateaued (Figure 2.6a). We can also

see from Figure 2.7d that the cyst radius continues to increase after cell number has plateaued,

suggesting that structure growth does not necessarily correlate with increased cell number. From

the MSD plot in Figure 2.7d we can see that the average MSD in the system remains low

throughout the 24 hours. This is likely a result of the confinement provided by the adhesive

islands in which the cells exist. As different cell types start to arise, this pipeline output can be

used to assess differences in movement dynamics between different cell populations and their

spatial segregations. For example, mesoderm cells, which are studied in our post-implantation

amniotic sac embryoid (PASE) model[33], are known to be more migratory as compared to other

populations like epithelial ectoderm and endoderm cells. While the model presented here is

limited to the first 24 h of cyst formation, future efforts can be devoted to extending this

timeframe to include important events such as symmetry breaking caused by the appearance of

amnion-like cells (AMLCs) in the PASE. In our PASE model, we found that the initial cell seeding density has a significant effect on morphogenesis and differentiation of hESCs[33,106]. With the information that can be obtained from this pipeline, we can begin to understand the role of initial conditions like cell seeding density, and we can begin the work of relating changes in the local environment with cell-level decisions that lead to the cyst-level growth and patterning.
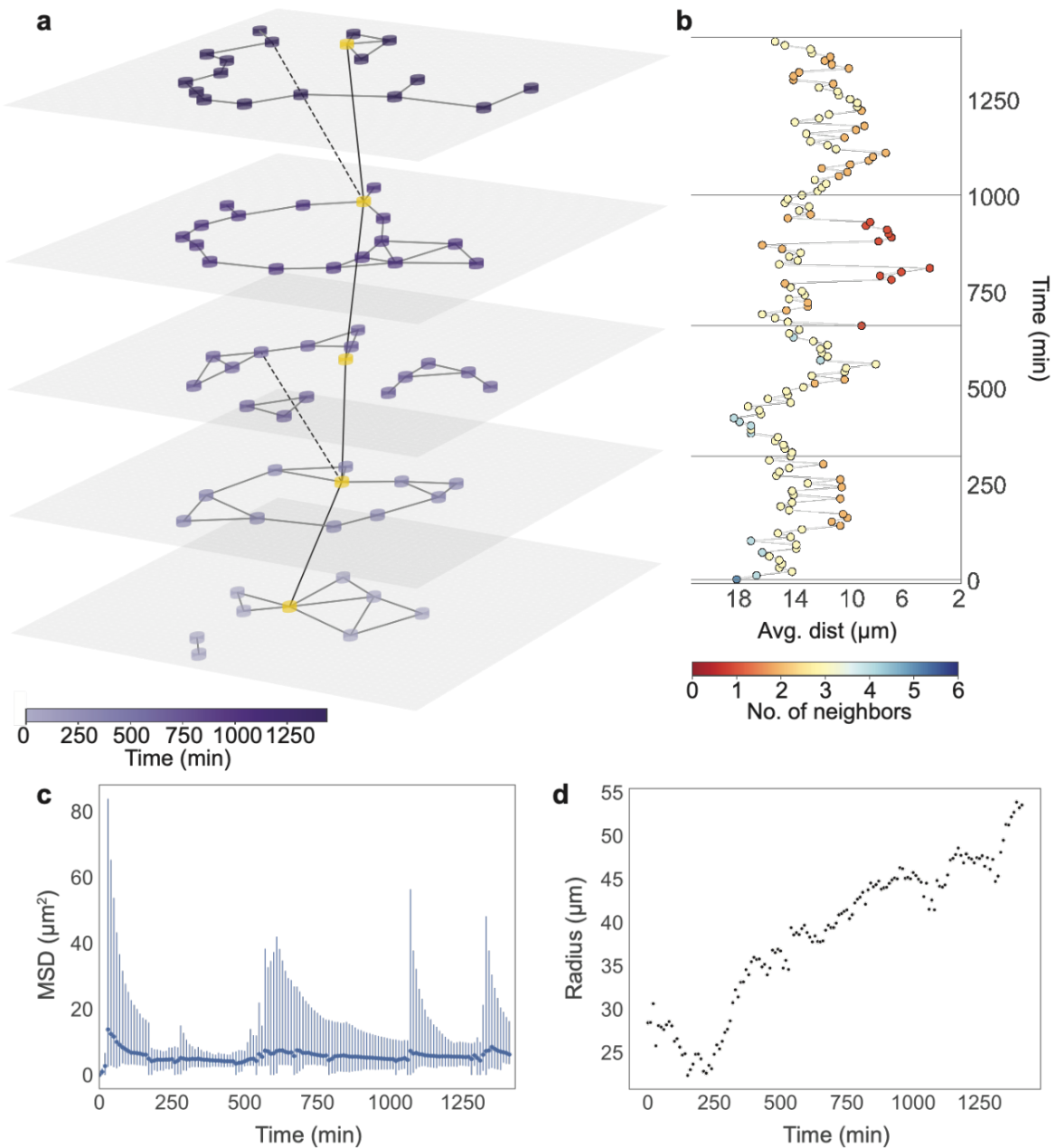
Figure 2.7 Characterization of hPSC cyst formation dynamics. (a) Network representation of a cyst at different time points. The color of the nodes in each plane indicates the time point. Each node represents a single cell in the cell cluster. Connections between cells are established based on a threshold distance. Cell lineage of the gold-colored cell is shown with the black line connecting the cell at different time points. Dotted lines indicate instances of cell division. (b) Average distance to neighbors and number of neighbors corresponding to the gold-colored cell in a as a function of time. (c) Average mean squared displacement (MSD) of all the cells in the cyst as a function of time. Dark blue points and bars represent the average MSD and the range of MSDs, respectively. (d) Radius of the cyst as a function of time.

2.4 Conclusions

The successful generation of human embryo-like structures is a crucial step in advancing fundamental understanding of human development, without using intact, natural human embryos. However, limitations on the insights gained through analysis of human embryo-like structures at discrete time points drove us to create a live cell video processing pipeline catered for the unique challenges of our system. With the use of both spatial and temporal information, we were able to create a machine learning model for event recognition. Furthermore, this model was integrated into an image processing pipeline that leveraged specialized image processing tools for the identification and tracking of individual cells in our system. With this integrative pipeline we were able to characterize the cell states and actions during the dynamic growth and morphogenesis of lumenal hESC cysts. Combining this tool with reporter lines for cell types of interest, we hope to advance in our goal to elucidate the mechanisms driving lumenogenesis, cyst growth, and cell fate specification in our *in vitro* hESC models of human development.

Chapter 3 Agent-Based Modeling-Genetic Algorithm Framework for the Study of Morphogenesis and Patterning

3.1 Introduction

The development of the human body is arguably the most complex process that occurs in our lifespan. In this process, the developing embryo must undergo proliferation, reorganization, lineage diversification, and dozens of cell fate specification events. These processes are guided by endogenous signaling that is affected by and allows for the correct interpretation of exogenous signaling (i.e. the cells receive 'instructions' from the environment that, combined with endogenous signaling, trigger the correct state changes). As the individual cells take actions based on their local microenvironments, global form and function emerge. Through self-organization, an initially unstructured cellular system attains complex structure and pattern via seemingly endogenous programs[116]. This self-organization can be subdivided into self-patterning and morphogenesis[116]. Self-patterning is initiated by a symmetry-breaking event where a subpopulation of the cells undergoes cell fate change. Maturation and stabilization of the self-patterned structure can be guided by a variety of mechanisms including reaction-diffusion mechanisms, gene regulatory networks, and cell-cell interactions. Morphogenesis is the process by which the structure achieves a complex, functional form via cell divisions and movement. The self-organizing abilities of our developing cells has been heavily leveraged in the field of stem cell-based models. In the last decades, researchers have built stem cell-based models that

41

are able to capture a wide array of key developmental processes in *in vitro* structures coined embryoids and organoids. In these *in vitro* models, the mechanical and chemical characteristics of the environment are used to trigger the seemingly endogenous developmental programs of stem cells that lead to self-organization. To date, researchers have been able to create models of the amniotic sac, neural tube, optic cup, and intestine among many others[33,116,117]. In these systems, structures undergo processes including polarization, lumenogenesis and symmetry breaking (Figure 3.1). While the specific mechanochemical factors that lead to these structures are unique, some key unifying features include the ability of structures to undergo symmetry breaking in homogeneous environments and a high sensitivity to initial cell cluster size[33,116,117]. Parsing the mechanisms that lead to the emergence of form and function in stem cell models is nontrivial and further complicated by cell-level heterogeneity and low efficiency. Computational modeling presents itself as a promising method for exploring possible mechanisms. Because of its ability to model cells as discrete agents, agent-based modeling (ABM) is particularly suited for the study of multicellular systems.
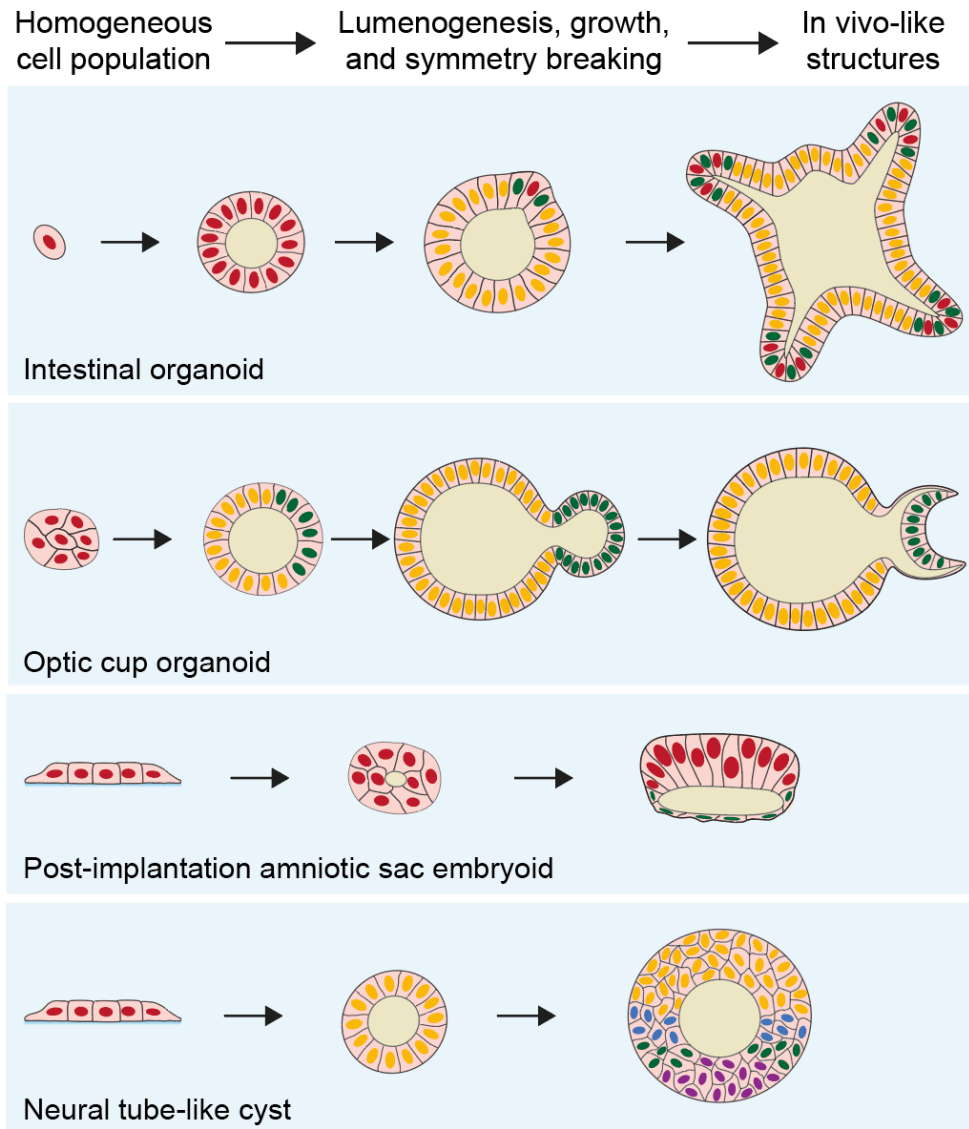
Figure 3.1 Schematic showing symmetry breaking and growth as seen in in vitro stem cell-based models.

Building a model when a system is new and there are many unknowns is understandably difficult. Even with some *a priori* knowledge, some parameter values required for the simulation might be difficult or even impossible to measure[76]. Parameter estimation techniques then become necessary. Approximate Bayesian computation (ABC) is one such method for parameter estimation[118,119]. This method compares simulations from a model with experimental data and

accepts simulations if the statistics for model data lie within an acceptable distance threshold. A faster adaptation of this method called approximate approximate Bayesian computation (AABC) was used by Lambert *et al*. for parameter estimation in an ABM of kidney branching[120,121]. While useful for estimating parameters, these methods do not address the need for uncovering morphogenesis rules. A possible approach to rule establishment is machine learning methods like deep learning (DL), which are capable of learning models from data without the need for prior domain knowledge[122]. An example of this is the work done by Wang *et al.* where they used deep reinforcement learning to study cell movement during embryonic development of *C.elegans*[123]. DL methods, however, have limitations when it comes to building ABM logic; they require a differentiable loss function[124,125] and lack transparency, which makes interpreting and trusting these models difficult[126,127]. A promising approach to model building is the use of genetic algorithms (GA). GA are a parameter space exploration method based on the mechanisms seen in evolution[78–81]. While GA are ubiquitous in fields such as computational evolution and robotics, they have yet to be used to their full potential in the field of multicellular systems. In this chapter, we present a novel combination of ABM with a GA. This ABM-GA framework evolves populations of ABMs. With the use of selection, mutations, crossovers, and reproduction, our GA yields rule sets that plausibly capture key aspects of self-organization and self-patterning.

3.2 Materials and Methods

3.2.1 Agent-based model in NetLogo

The ABM was constructed in NetLogo, a widely-used and open-source ABM platform[128]. In the model, cells exist in a hexagonal lattice with a field of view limited to their six immediate neighbors (Fig 3.2a). To facilitate finer diffusion profiles, the cells are larger than a single patch,

with size ratio of 1:19 (cells:patches). Each position of a cell's neighborhood might be occupied by either a single cell, an array of matrix, or an array of lumenal space. Cells can die, move, form lumenal space, divide, secrete factors (activating and/or inhibiting signals), and differentiate (Figure 3.2b). The cell's neighborhood composition will determine which rule it follows, with a total of 20 morphogenesis rules for the different possible compositions (Figure 3.2c). Differentiation may take place via two mechanisms: (1) contact-mediated differentiation wherein differentiated cells are assumed to have inductive effects on their pluripotent neighbors and (2) presence of activator exceeding some established threshold.
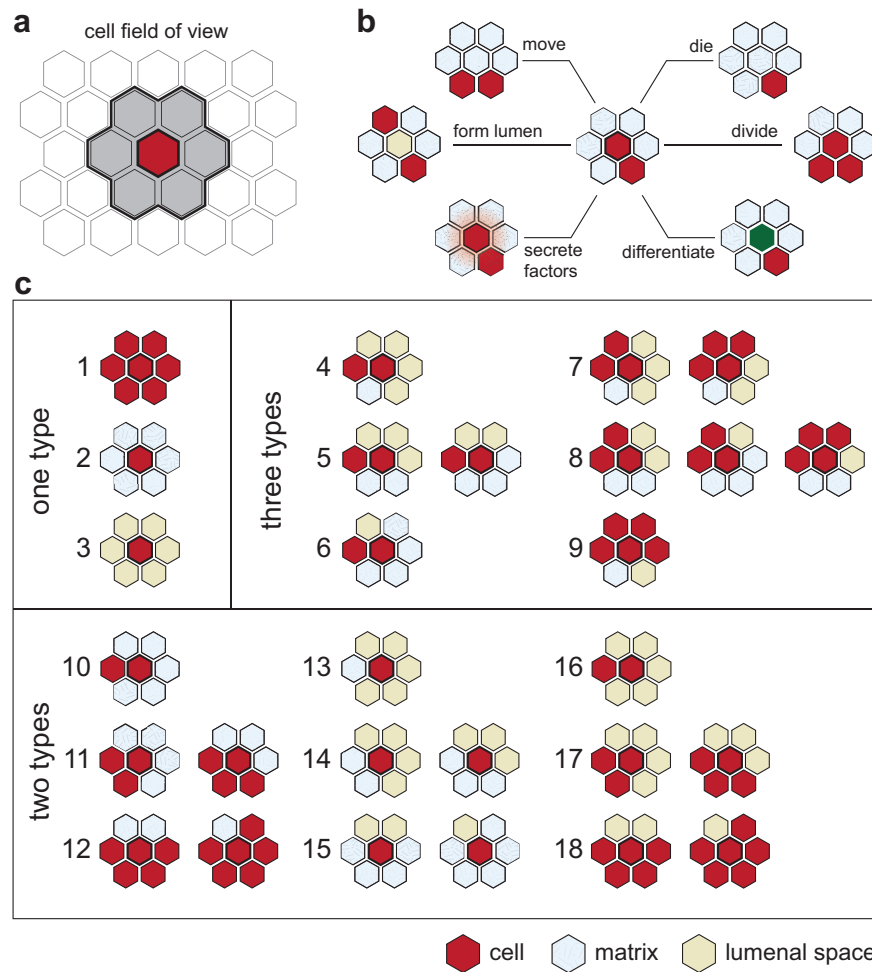
Figure 3.2 Agent-based model. (a) Field of view of a cell. (b) Possible actions carried out by a cell in response to the local microenvironment. (c) ABM rules for each type of neighborhood composition.

3.2.2 ABM-GA framework

The goal of the GA is to discover rule sets that plausibly capture features of self-organization and symmetry breaking as seen in stem cell-based models. In our GA, which was constructed using Python, we evolve populations of ABMs (i.e. each ABM is an 'organism' in a population undergoing evolution through the GA). Each model has a genome that specifies 31 model parameters. These parameters can be divided into three categories: initial conditions (2 rules), morphogenesis (20 rules) and differentiation (9 rules) (Figure 3.3). There exists a set of possible morphogenesis rules that a cell might follow given the composition of its local neighborhood. The ABM differentiation process is modular in order to accommodate different possible mechanisms. Evolving parameters include (1) cell type (differentiated or undifferentiated) responsible for activating signal secretion, (2) amount of activator being secreted, (3) cell type responsible for inhibitory signal secretion, (4) amount of inhibitor being secreted, (5) direction (basal or apical) of activating signal secretion, (6) direction of inhibitory signal secretion, (7) differentiation threshold, and (8) period of exposure needed for differentiation.
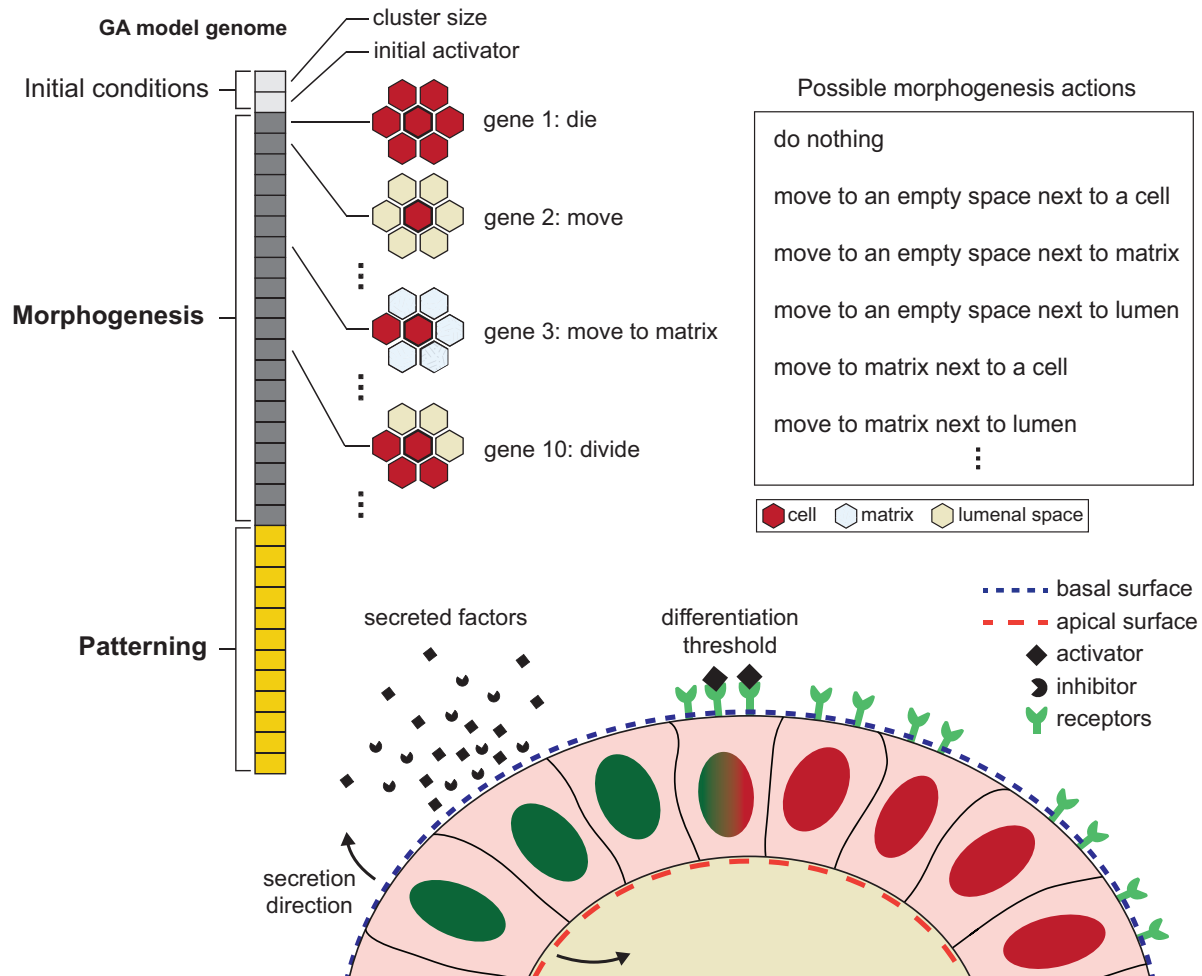
Figure 3.3 Genetic algorithm setup. Each model or 'organism' in the population has a genome containing the model parameters of interest. These can be divided into (1) initial condition parameters, (2) morphogenesis rules, and (3) patterning parameters.

In order to evolve the population of models towards the desired functionality, we established four distinct fitness objectives or parameters. Two of the parameters pertain to patterning and two pertain to morphogenesis. For morphogenesis, each organism has a structure score and a lumen score. The structure score accounts for structure integrity and cell connectivity. And the lumen score accounts for number of lumen. For patterning, there is a sections score and a proportion score. The sections score is the inverse of the error between the target number of differentiated and pluripotent sections and the actual number of differentiated and pluripotent

sections in the cyst. Similarly, the proportion score is the inverse of the error between the target proportion of differentiated to pluripotent section area and the actual proportion. The target number of sections, target proportion of cell types, and target number of lumen can be changed according to the particular system under study. Given our multi-objective fitness evaluation, we used the non-dominated sorting GA, NSGA-II[129], to rank the models in order to undergo selection for the next generation.

The ABM-GA framework functions as follows. At initiation of evolution, a population of a desired size is randomly generated by assigning random values to each position in the genome. For all positions in the genome in which the value is an integer, a range within which to choose a random value from is given at initialization. For all positions in the genome that pertain to a morphogenesis rule, there exists a pool of actions from which to randomly draw from (Figure 3.4). In order to mitigate overfitting, a user-defined percentage of morphogenesis rules are assigned the value of 'do-nothing' in order to render them nonfunctional. A child population is then generated with the use of mutations, and the combined populations are run in NetLogo (Figure 3.4). The fitness of structures resulting from the NetLogo run are then evaluated based on how well it corresponds to the target structure and composition. The next parent generation is constructed by selecting the highest-ranking individuals identified by the nondominated sorting algorithm as well as by a configurable percentage of crossovers and new randomly generated organisms. The connection between the Python GA and NetLogo is established using the PyNetlogo library[130].
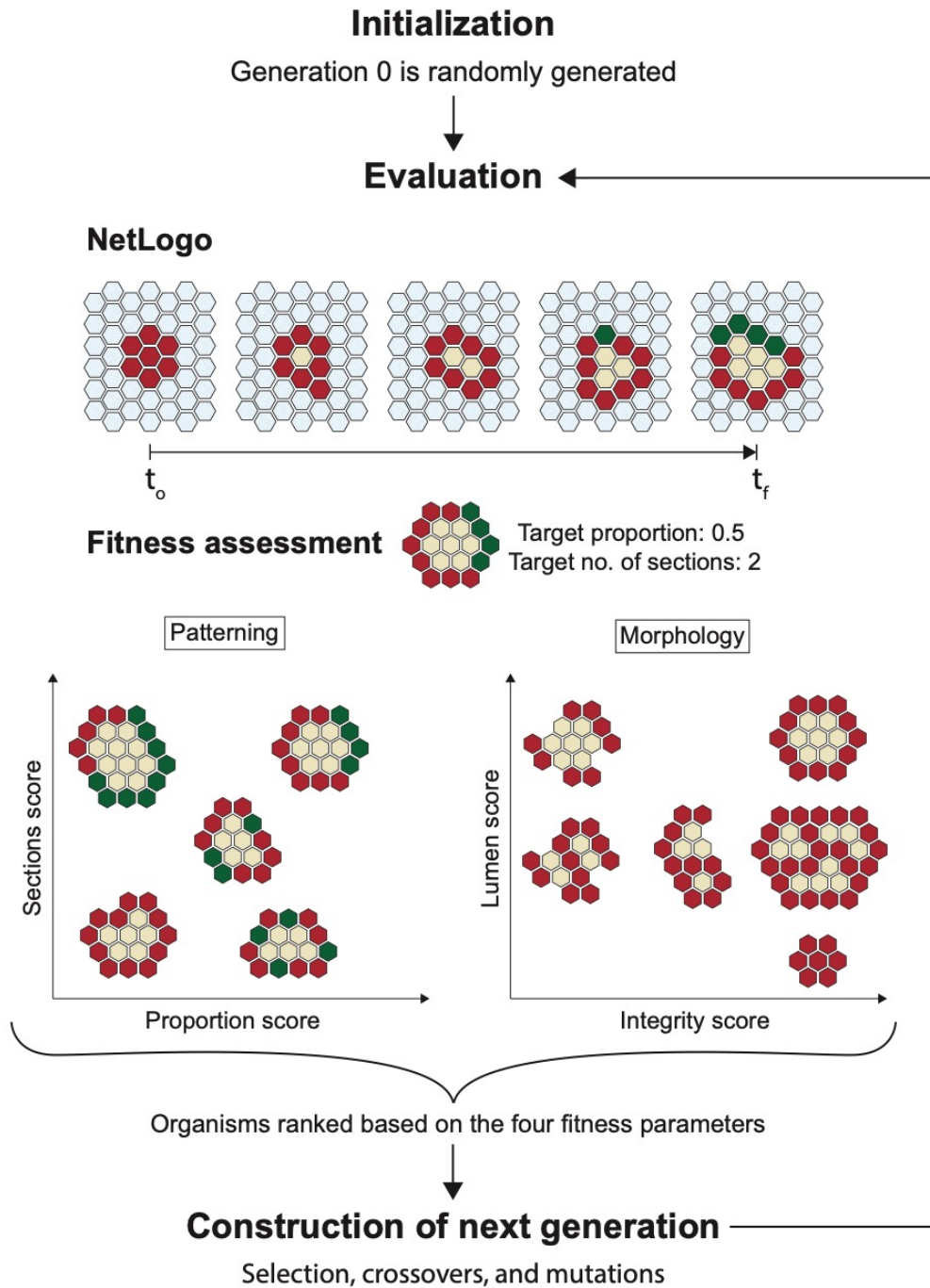
Figure 3.4 ABM-GA framework. The GA initializes with a population of randomly generated organisms. The population is evaluated by first running the models in NetLogo and then carrying out fitness assessment and genetic manipulation in Python. This process is repeated for each generation for an established amount of time.

3.3 Results

Having had established the framework, we sought to understand how an evolving

population behaves in time. The primary outputs we were concerned with were the sections score

and the diversity in the population. Because parameter space is large, we wanted an evolutionary

approach that would preserve diversity in the population by favoring exploration over

exploitation. We ran the framework for a section objective of four for a total of 100 generations

(Figure 3.5). The GA parameters were set as follows: population size of 100, percentage of

nonfunctional morphogenesis rules of 70%, mutation rate of 0.5, retention of top 2 organisms,

20% of the new generation from crossovers, and 10% of the new generation from new randomly

generated organisms. In Figure 3.5 we can see a random sampling of the evolving population for

generations 0 and 100. The average sections score of the population was relatively constant

throughout evolution (Figure 3.5b). As a proxy for diversity, we looked at the number of species.

Models were grouped into the same species if they shared at least 70% of their genome. For this

population of 100 models, the number of species remained above 50 throughout evolution,

indicating that diversity was indeed being preserved (Figure 3.5c). Finally, we looked at the

percentage of nonfunctional morphogenesis rules (Figure 3.5d). Interestingly, this value

remained relatively constant throughout evolution.
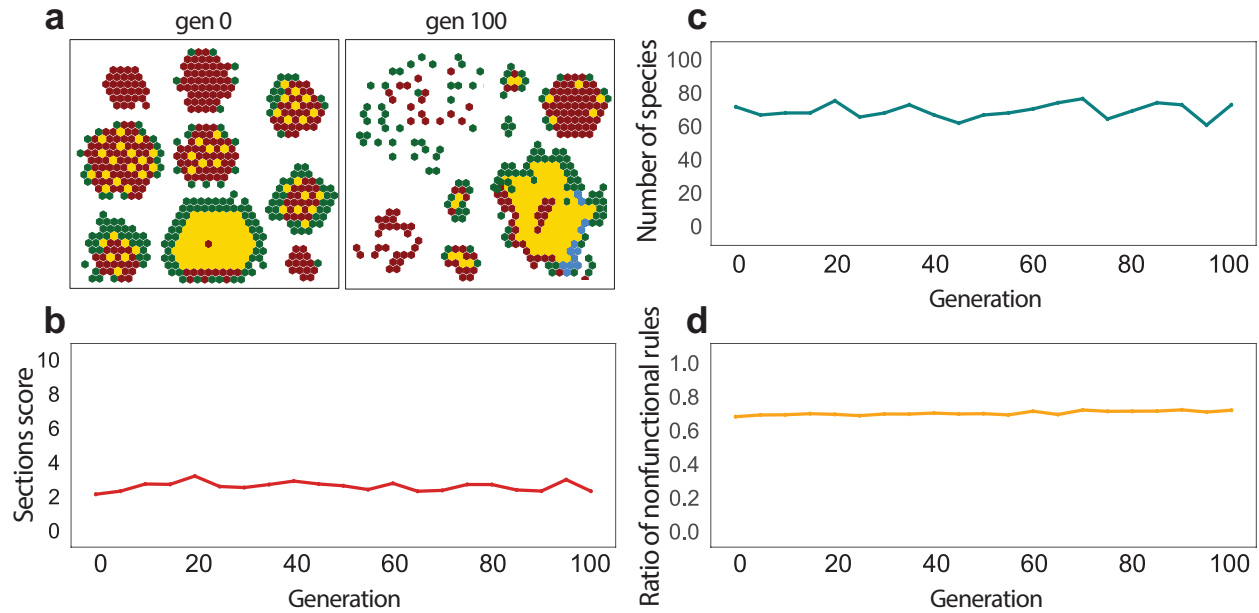
Figure 3.5 ABM-GA framework output for a section objective of four. (a) Sample organisms from generations 0 and 100. (b) Average sections score of the evolving population at each generation. (c) Number of species in the evolving population at each generation. (c) Average ratio of nonfunctional rules of the evolving population at each generation.

While diversity was preserved at the population level, we also sought to understand diversity within the set of successful models. In Figure 3.6a, we can see some sample successful organisms from generations 0-15, 50-60, and 90-100. The number of species for this subset of organisms seems to grow as evolution proceeds (Figure 3.6b), leading us to believe that our exploration-favored GA approach can be successful at generating a diverse set of possible solutions.
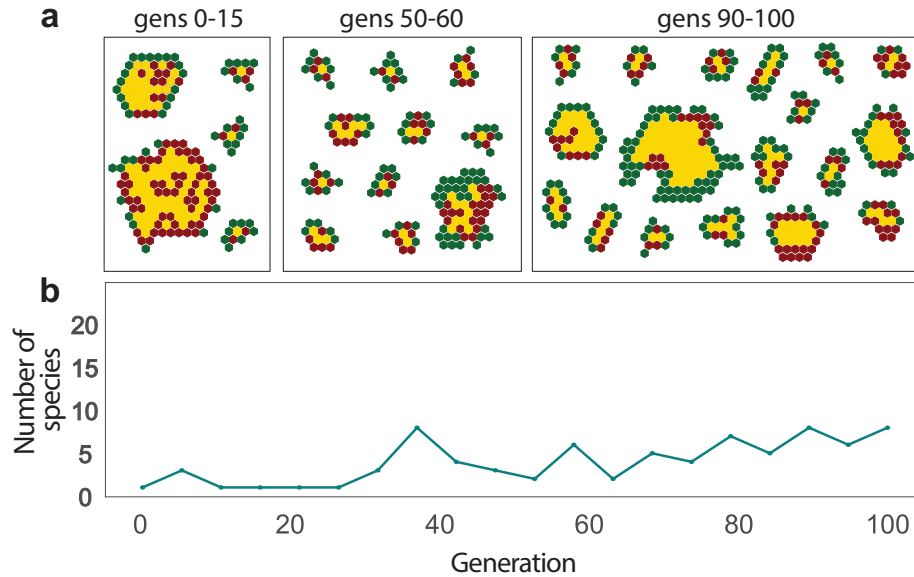
Figure 3.6 Subset of successful organisms resulting from the ABM-GA framework for a section objective of four. (a) Samples of successful organisms from generations 0-15, 50-60, and 90-100. (b) Number of species through time within the subset of successful organisms.

In order to assess if the ABM-GA is able to yield a variety of patterns, we ran the framework for two additional sections objectives: six sections and two sections. The GA parameters were set as follows: population size of 100, percentage of nonfunctional morphogenesis rules of 70%, mutation rate of 0.5, retention of top 2 organisms, 20% of the new generation from crossovers, and 10% of the new generation from new randomly generated organisms. Each population was left to run for 100 generations. The behavior was found to be consistent across the different runs; the average sections score, the number of species, and the percentage of nonfunctional rules remain fairly stable throughout evolution (Figure 3.7a,b,c). The sections objective was met for each of the populations (Figure 3.7d,e,f).
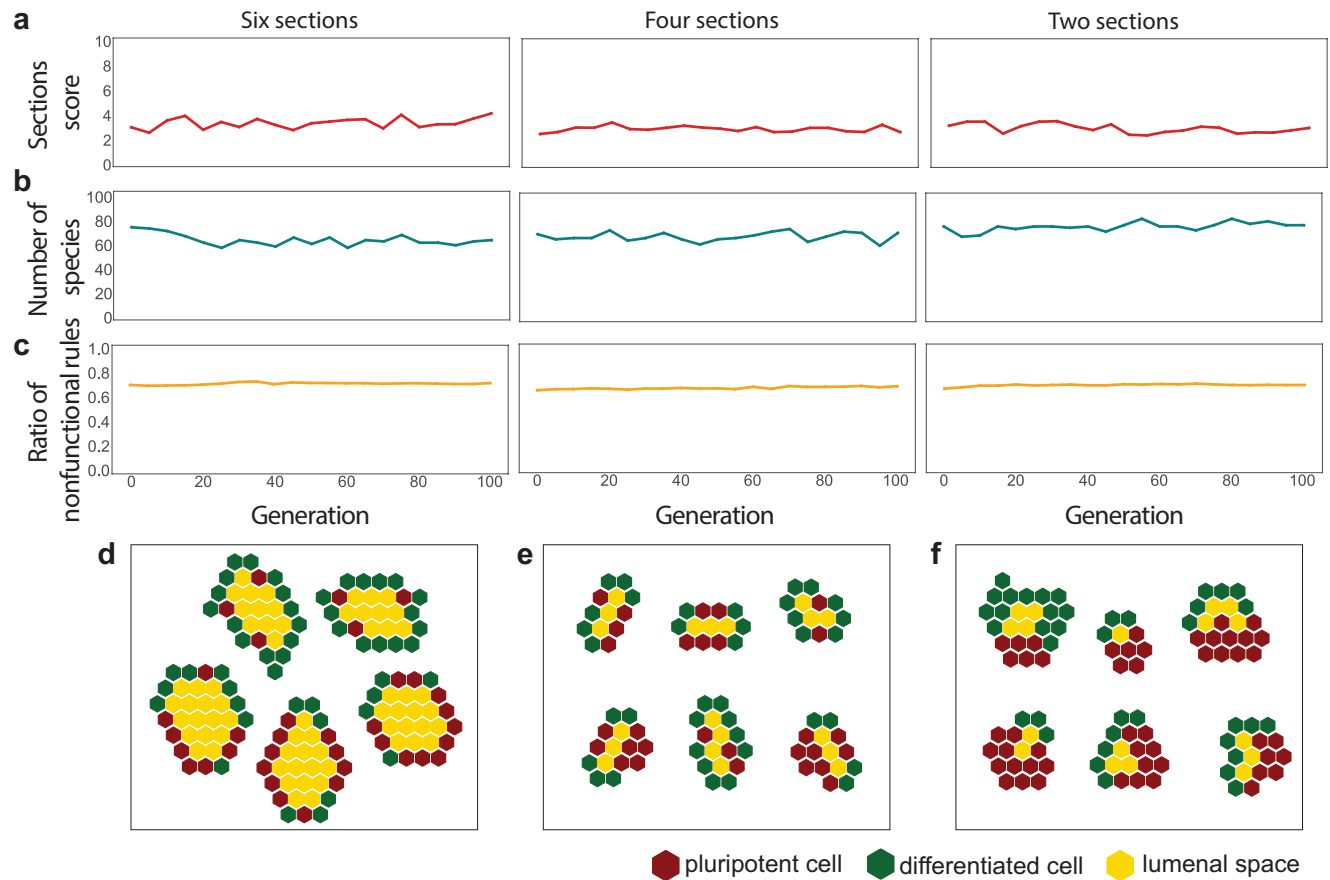
Figure 3.7 ABM-GA framework output for different sections objectives. (a) Average sections score of the evolving population at each generation for an objective of six, four, and two sections respectively. (b) Number of species in the evolving population at each generation for an objective of six, four, and two sections respectively. (c) Average ratio of nonfunctional rules of the evolving population at each generation for an objective of six, four and two sections respectively. (d-f) Sample patterned organisms of the evolving population at generation 100 for an objective of six, four, and two sections respectively.

3.4 Discussion

Understanding the mechanisms that lead to the emergence of form and function in biological systems is of great importance not only for our fundamental knowledge of development but also for the development of regenerative medicine. A variety of *in vitro* models of embryogenesis and organogenesis have been developed. Many of them, however, suffer from low controllability and therefore are not amenable for mechanistic studies that require a high

volume of samples and in which parameter space is large. *In silico* models present themselves as a great complement to *in vitro* experiments. But because these embryo and organ models are generally poorly understood, building such models becomes difficult. The ABM-GA framework presented in this work lays the foundation for a novel approach to model building as it pertains to the context of morphogenesis and patterning. The framework presented in this work is highly abstracted and modular, allowing for the modeler to begin with the simplest model possible and add details and complexity as needed. As shown, the framework is able to yield models that generate structures with specific desired patterns. Having this set of 'successful' models, the next step then becomes a more comprehensive comparison between the *in silico* model and the *in vitro* reality. Further filtering of the resulting models can be done by investigating whether the models can capture specific known behaviors of interest such as the effect of initial conditions on final structure composition. Having found those *in silico* models that demonstrate the most similarity to the *in vitro* reality, future work can focus on the use of such models for parameter space exploration, hypothesis testing, and prediction.

Chapter 4 Agent-Based Model of Epiblast Growth and Patterning

4.1 Introduction

One of the strict requirements for a successful pregnancy is embryo implantation into the maternal tissue. During this process, the epiblast cyst undergoes a major symmetry-breaking event that gives rise to the amniotic sac (Figure 4.1). In order for this to happen, cells at the maternal tissue-facing pole of the epiblast cyst must undergo differentiation into squamous amniotic ectoderm. This development of the amnion, or amniogenesis, is a critical developmental milestone[131,132]. Despite its importance, our understanding of the mechanisms underlying this cell fate decision and the stabilization of this asymmetric structure is very limited.
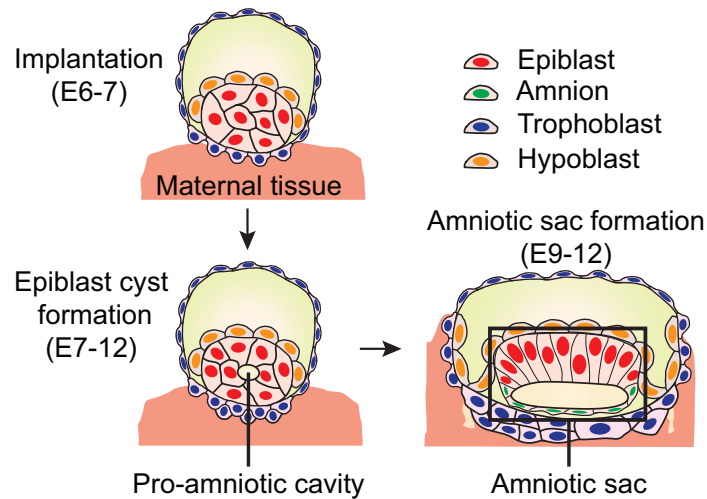
Figure 4.1 Schematic of epiblast cyst growth and development into the asymmetric amniotic sac during embryo implantation.

In the work done by our group, Shao *et al.* were the first to show that hPSCs could differentiate into amnion cells[32]. In this system, hPSC clusters would undergo lumenogenesis and form lumenal structures containing a central cavity. Given a biomimetic environment consisting of a soft gel bed made with the basement membrane matrix Geltrex[TM] and a 3D matrix overlay made with a low concentration of Geltrex[TM] diluted in culture medium, the cells in the cyst would undergo amniogenesis to give rise to squamous amniotic cysts. In later work done by Shao *et al.*, they identified a very small subset of stable asymmetric cysts in the system[33]. Time-lapse data revealed that morphogenesis involved lumenogenesis, growth, and thinning of a section of the cyst (Figure 4.2a). Characterization of the cells showed that these asymmetric cysts consisted of squamous amniotic cells at one pole and undifferentiated hPSCs at the opposite pole, much like what we see in the post-implantation human amniotic sac structure. This asymmetric cyst was named post-implantation amniotic sac embryoid (PASE)[33].
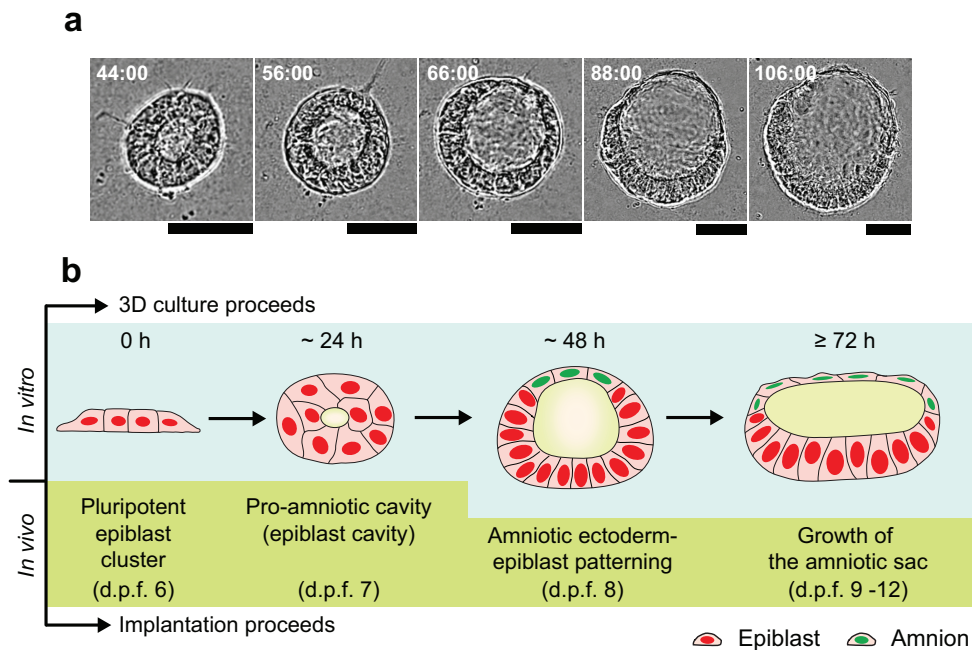
Figure 4.2 Developmental trajectory of PASE. (a) Representative time-lapse phase-contrast images showing dynamic morphogenesis during the development of a PASE. Time stamps indicate the total hours of culture. Scale bars, 50 μm. (b) Cartoon showing the time course of PASE development *in vitro*, compared with human amniotic sac development *in vivo*.

Interestingly, the efficiency for the generation of this asymmetric structure was shown to be heavily dependent on initial cell plating density. At the highest plating density tested (70,000 cells cm^{-2}), all cysts were composed of pluripotent columnar cells on day 4 (Figure 4.3a). At the lowest plating density tested (20,000 cells cm^{-2}), all of the structures became squamous amniotic ectoderm-like cysts by day 4 (Figure 2b). And at some intermediate plating density (50,000 cells cm^{-2}), the percentage of PASE structures peaked. Later work done in our lab by Esfahani *et al*. further solidified the importance of initial cell numbers. In this work, amniogenesis was shown to be cluster-size dependent; clusters that began with 4-7 cells would develop into amnion cysts while clusters that began with more than 8 cells would form pluripotent epiblast cysts[106]. Because amnion fate occurred only within a narrow window of cyst sizes, this work suggested that amniogenesis is linked to a community sensing mechanism.
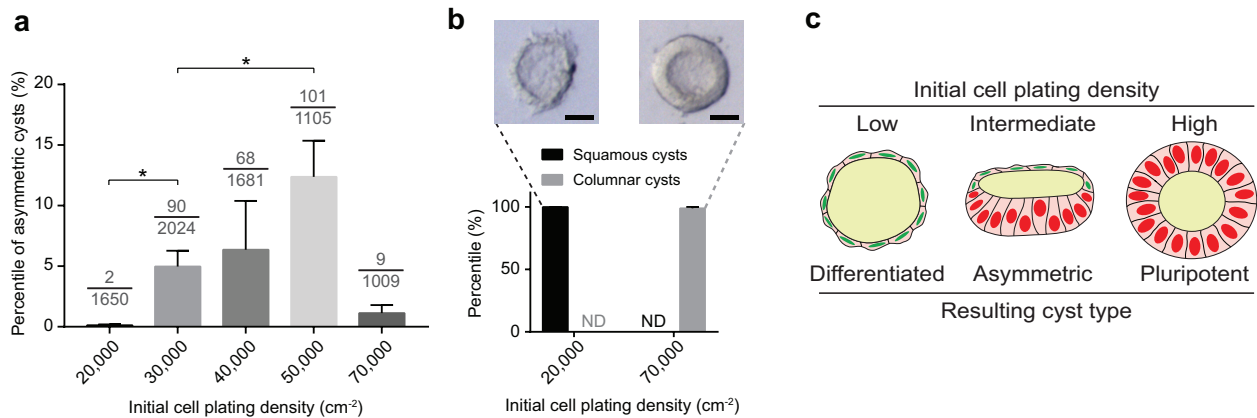
Figure 4.3 (a) Bar plot showing the percentage of asymmetric cysts formed on day 4 at different initial plating densities (indicated on the x-axis); all other conditions are identical. Data represent the mean ± s.e.m. The denominator of each fraction indicates the total number of cysts quantitated for that condition; the numerator of each fraction indicates the number of asymmetric cysts among the quantitated cysts. P-values were calculated using unpaired, two-sided Student's t-test. *P < 0.05. n = 3–5 biological replicates. n = 3 independent experiments. (b) Bar plot showing the percentage of fully squamous amniotic ectoderm-like cysts and fully columnar cysts under low-extreme and high-extreme cell plating densities. ND not detected. Phase-contrast images show representative squamous (left) and columnar (right) tissue phenotypes observed at these extreme conditions. n = 3 independent experiments. Scale bars, 30 μm. (c) Schematic summarizing the effect of initial cell plating density on the formation of the asymmetrically patterned epithelial cysts.

In the PASE model, endogenous bone morphogenetic protein (BMP) signaling is believed to trigger amnion differentiation, and Noggin (a known BMP signaling inhibitor) is believed to be involved in the stabilization of the asymmetry[46,133]. Knowing that patterning in this model could be controlled by the interaction of these two morphogens, our ABM which permits for the secretion and diffusion of an activator and inhibitor, has the potential to help parse out how BMP and Noggin work together to bring about genetic asymmetry. Further, with this framework, we hope to gain insight into possible mechanisms guiding growth and lumenogenesis.

4.2 Materials and Methods

In this work, we employ the ABM-GA framework presented in Chapter 3. The agents in this model are the cells. In an environment consisting of a hexagonal grid, each cell has six neighboring positions. These positions might be filled by other cells, matrix, or lumenal space. Cell actions can be divided into two categories: morphogenesis and patterning. Morphogenesis actions include dying, moving, and dividing. Patterning actions include differentiating, secreting activator, and secreting inhibitor (Figures 3.2 & 3.3). Cells will take actions based on the composition of their local neighborhood and their identity (pluripotent or differentiated).

As explained in Chapter 3, we utilize a genetic algorithm to construct ABMs. In brief, a population of models is evolved through mutations, crossovers, and selection guided by four fitness objectives: number of sections, proportion of one section to another, number of lumen, and structure integrity. For this study, the most important fitness objectives were structural integrity and that the number of sections in the cyst is 2.
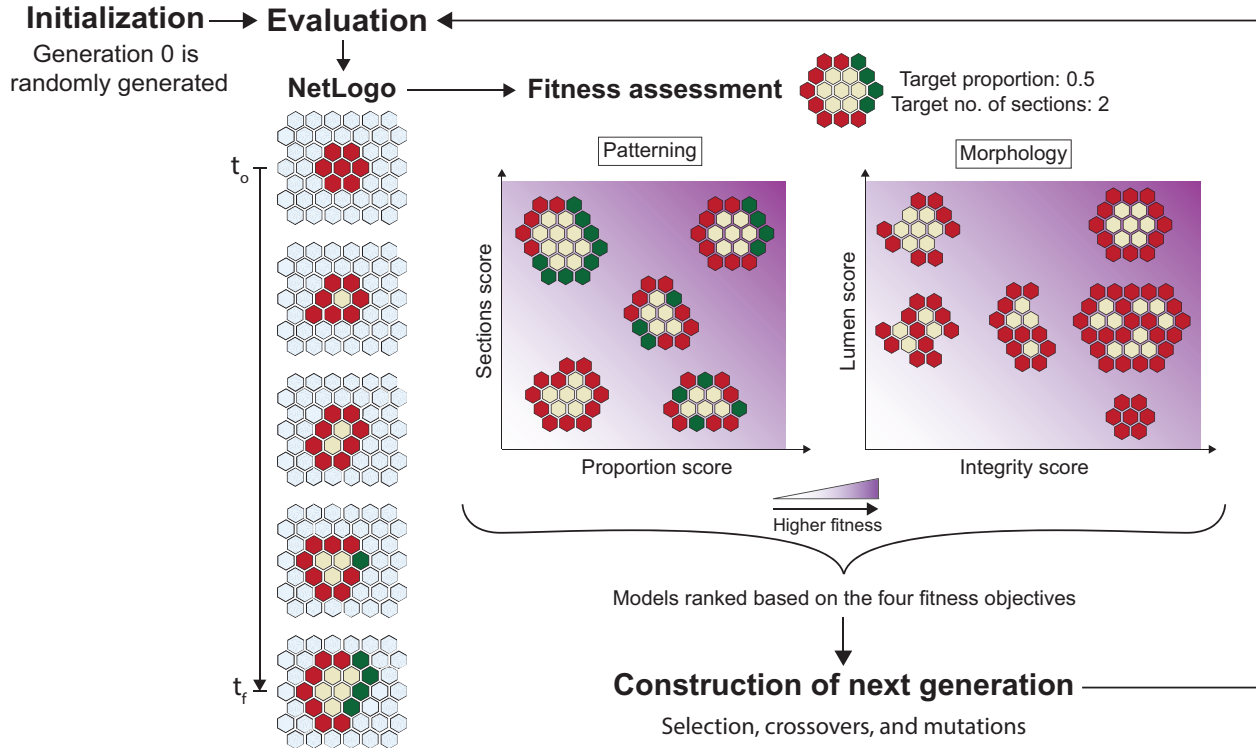
Figure 4.4 ABM-GA framework. The GA initializes with a population of randomly generated organisms. The population is evaluated by first running the models in NetLogo and then carrying out fitness assessment and genetic manipulation in Python. This process is repeated for each generation for an established amount of time.

4.3 Results

Understanding how morphogenesis and patterning take place in *in vitro* models of development can be difficult given the low efficiency of some of these systems[33,95]. *In silico* modeling then presents itself as an invaluable tool for hypothesis testing and parameter space exploration. Following this logic, we aimed to create agent-based models that could yield structures resembling our PASE and plausibly capture key rules or mechanisms involved in the process of growth and differentiation.

The ABM-GA was run with the objective of obtaining two-section structures. The GA parameters were set as follows: population size of 500, mutation rate of 0.5, retention of top 2 organisms, 20% of the new generation from crossovers, and 10% of the new generation from

new randomly generated organisms. Each population was left to run for 100 generations. Figure

4.5 shows results from selected models from the 100th generation. For each model, the simulation

was run in NetLogo 100 times. The graphs show the ratio of each possible cyst outcome.

'Patterned' refers to cysts that showed the desired asymmetric pattern consisting of two sections,

'unpatterned' includes both fully differentiated and fully pluripotent cysts, 'other pattern' refers

to incorrectly patterned cyst in which there was symmetry breaking that led to more than two

sections, and 'incorrect morphology' includes cysts that either lacked a lumen or lacked

structural integrity.
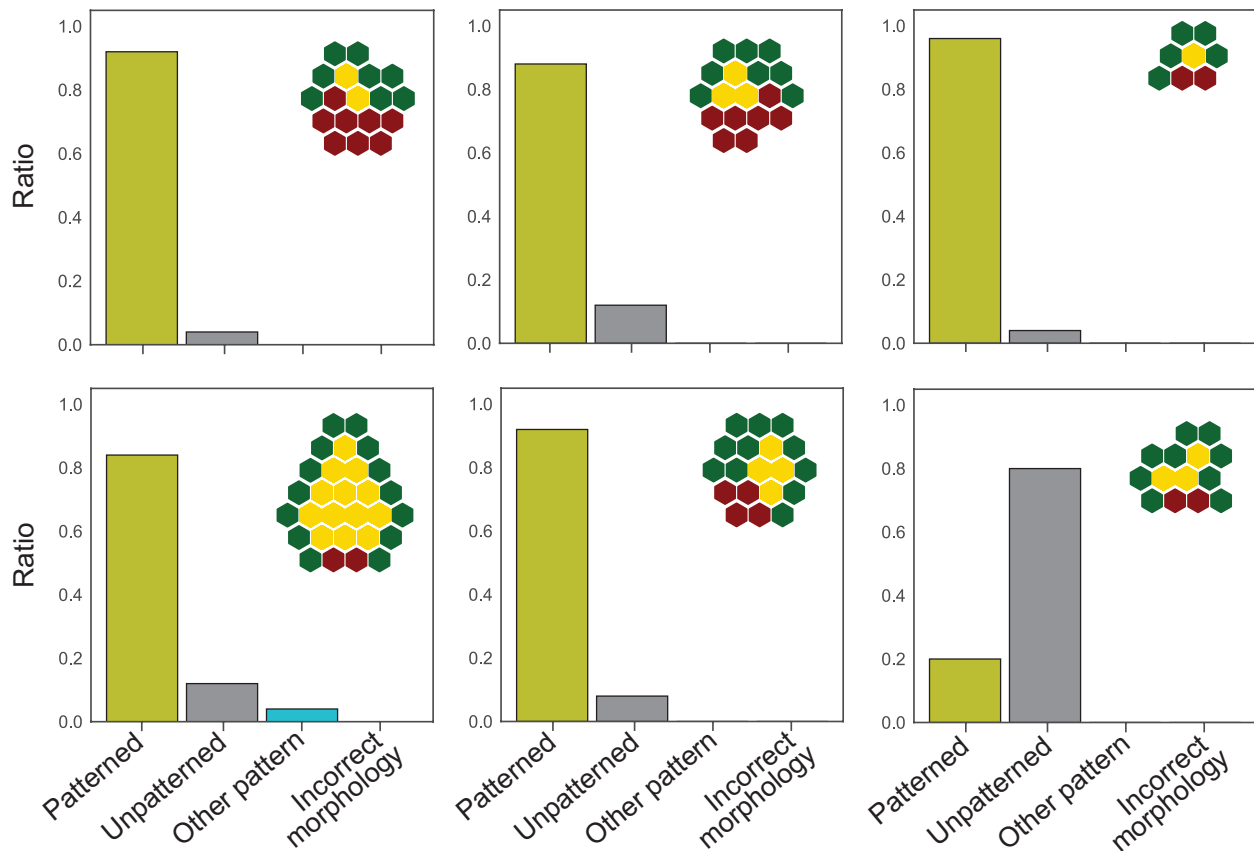
Figure 4.5 Efficiency of chosen successful models generated with the ABM-GA framework. Models were evaluated 100 times. Bars indicate the ratio of each possible outcome (patterned cyst, unpatterned cyst, incorrectly patterned cyst, and cyst with incorrect morphology).

The ABM-GA yielded a variety of models that are fairly successful at achieving patterning (Figure 4.5). The next question then becomes, can any of the models capture interesting dynamics seen in our *in vitro* systems? One of the most interesting aspects of our system is the observed effect of initial cell number. As stated previously, in all of our models pertaining to amniogenesis and amnion cyst-like structures, final cyst fate depends heavily on initial cell number[33,95,106]. To test if this effect can be seen in our models, we chose four successful models and ran the NetLogo simulation using a range of initial cell numbers, going from 30 percent less cells to 60 percent more cells than used in the original run (Figure 4.6). For each model, the graph shows the ratio of patterned cysts and the ratio of unpatterned cysts obtained after 100 simulation runs. Under each graph, we show a final structure for each tested initial cell number.
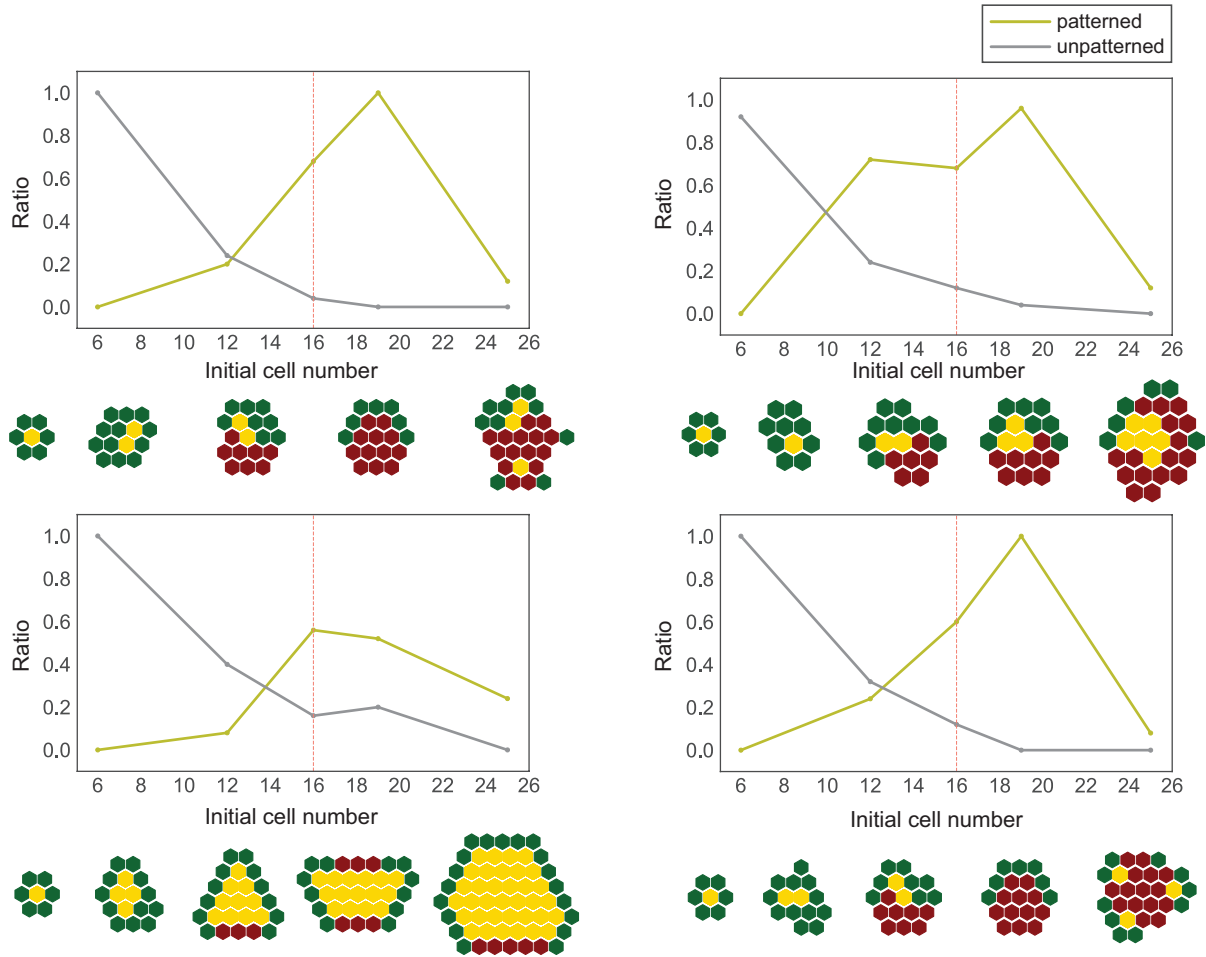
Figure 4.6 In silico experiment varying initial cell number. For each chosen successful model, a simulation was run with varying cluster sizes, starting from an approximately 30 percent smaller cluster to an approximately 60 percent larger cluster. The corral-colored line indicates the original cell cluster size. Under each graph, we show a final structure configuration for each tested initial cell number.

Interestingly, small initial cell numbers result in fully differentiated cysts for all of the tested models. As the cell number increases, we see a rise in correctly patterned asymmetric cysts. This increase, however, is not monotonic. At some point, the number of correctly patterned asymmetric cysts begins to decrease. The models, however, fail to capture the appearance of fully pluripotent cysts at high cell numbers. Instead, large cell numbers show either the persistence of the patterning or aberrant morphogenesis with an incorrect number of cavities.

One of the main advantages of *in silico* models is the ability to run low-cost experiments. For the purpose of demonstrating this, we ran a series of knock out (KO) experiments using our subset of successful models. First, we carried out knock outs of relevant patterning parameters in the genome including (1) the initial amount of homogeneous activator in the environment, (2) the amount of inhibitor released by the cells, and (3) the amount of activator released by the cells (Figure 4.7a). As expected, knocking out the initial activator (i.e. setting the amount to 0) meant that no model could yield patterning, as the initial signal is necessary for the initiation of chemical secretion by the cells (Figure 4.7b). Interestingly, while knocking out inhibitor secretion by the cells significantly decreased the number of successful structures, knocking out activator secretion by the cells did not (Figure 4.7b).
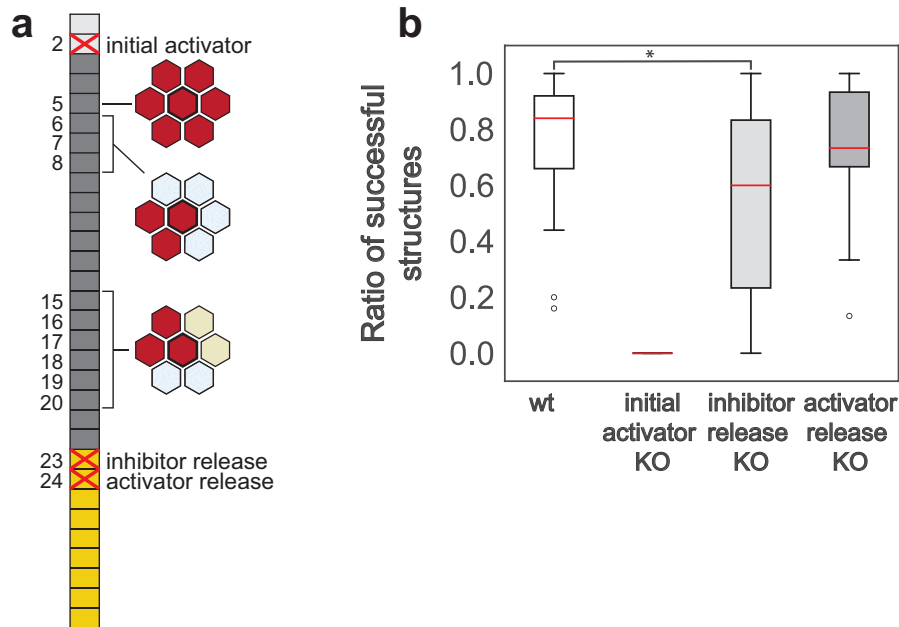
Figure 4.7 Patterning genes KO experiments on the subset of successful organisms. (a) Schematic of the genome showing the three parameters that were knocked out (2, 23, and 24). (b) Box plot showing the average ratio of successful structures for all the successful models for each experimental condition (indicated on the x-axis). P-values were calculated using two-sided Student's t-test. *P < 0.05.

A second set of KO experiments was done for investigating the importance of the morphogenesis rules (Figure 4.8). The KO of a morphogenesis rule was carried out by setting the value of that rule to 'do nothing' so that the cell did not respond to that specific neighborhood composition. Knocking out rule 5, which pertains to a neighborhood made up of cells, significantly decreased the percentage of successful structures (Figure 4.8b). Knocking out rule 6, which pertains to a neighborhood composed of cells and matrix, also significantly decreased the percentage of successful structures. KO of the remaining rules pertaining to a neighborhood of cells and matrix (Figure 4.8c) or a neighborhood of cells, lumen, and matrix (Figure 4.8d), did not lead to any significant decrease in the percentage of successful structures.
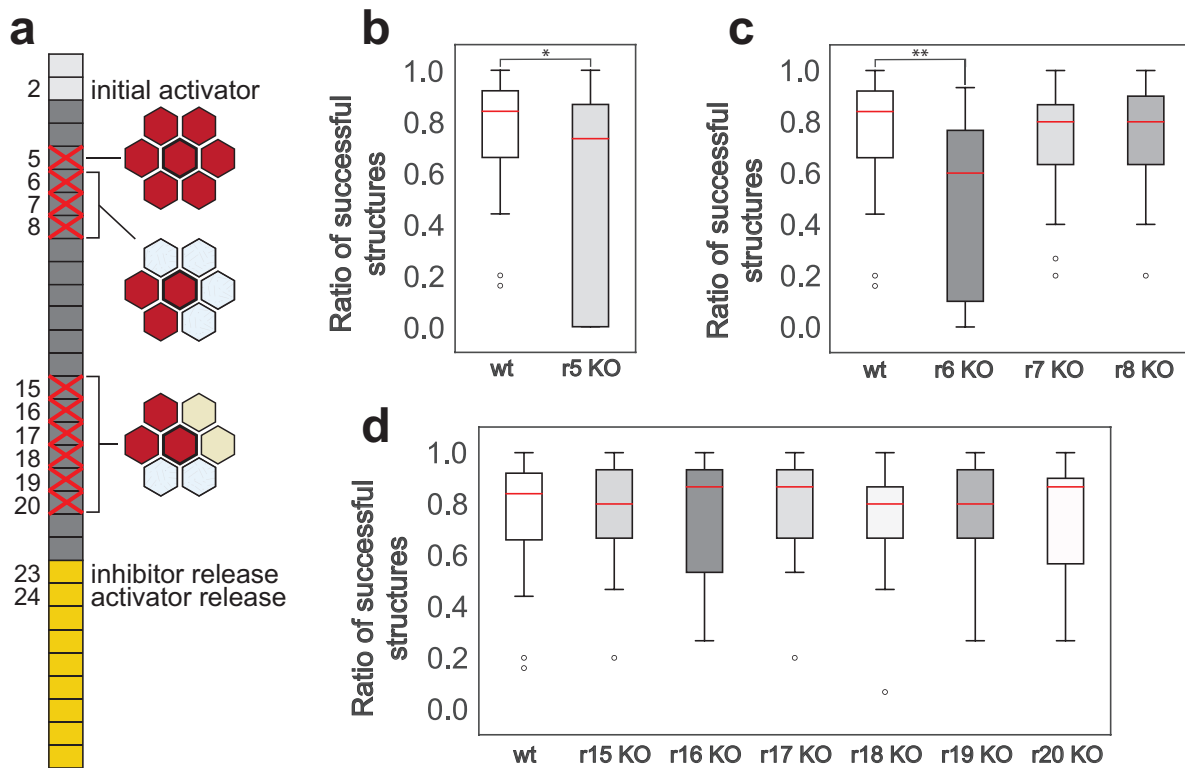
Figure 4.8 Morphogenesis rules KO experiments on the subset of successful models. (a) Schematic of the genome showing the ten parameters that were knocked out (5-8 and 15-20). (b-d) Box plots showing the average ratio of successful structures for all the successful models for each experimental condition (indicated on the x-axis). P-values were calculated using two-sided Student's t-test. *P < 0.05. **P<0.01.

4.4 Discussion

The formation of the amniotic sac is a pivotal event for the successful development of the human embryo. Yue *et al*. were able to model it with the use of hPSCs and uncover some of the signaling mechanisms involved in the process[33]. However, the low yield of the asymmetric PASE structure and the lack of controllability of its symmetry breaking makes in depth mechanistic studies with this model difficult. Little is known about what makes a cell in this seemingly homogeneous cluster in a seemingly homogeneous environment differentiate. And it is still unknown how this differentiation event stops in order to give rise to a stably asymmetric structure. In this work, we began the task of unraveling the mechanisms underlying morphogenesis and patterning of the PASE with the use of an ABM-GA framework. We showed that this method can yield ABMs that give rise to asymmetric cysts resembling the PASE structure. Further, we showed that a select group of models can partially capture the effect of initial cell number on final cyst fate with lower initial cell numbers leading to fully differentiated cysts. Additionally, we demonstrated how these models can be used to carry out *in silico* experiments such as KOs in order to understand the importance of different model parameters. Interestingly, KO of the cell's ability to secrete inhibitor significantly decreased the number of successfully patterned structures while KO of the cell's ability to secrete activator did not. This suggest that, in our *in silico* models, an initial amount of activator in the system combined with inhibitor secretion is enough to bring about correct patterning. Future work needs to be done to determine if this is also true for our *in vitro* system. With regards to morphogenesis rules, only the KOs pertaining to an all-cell neighborhood and a cell and matrix neighborhood led to a decrease in the percentage of patterned structures. This result is not surprising given that at time 0, cells exist in a cluster surrounded by matrix. Consequently, knocking out rules related to these

neighborhoods will hinder proper cyst morphogenesis. It is important to note that these KO experiments only show the effect of the KOs on the average success rate of a group of models, without looking at the effect on any specific model. Future work can be done to examine the effect of KOs on specific models and parse how exactly the disruption in morphogenesis leads to a disruption in the patterning.

This study presents a preliminary proof of concept that ABMs can be built with little to no *a priori* knowledge and still yield interesting results. More work is to be done, however, both in terms of exploring the importance of different parts of the model genome and increasing model complexity via more realistic reaction-diffusion simulations.

Chapter 5 Summary and Future Works

5.1 Summary

In this work, we sought to create novel tools to study the emergence of form and function, specifically during the development of the epiblast cyst. This endeavor required the ability to both characterize and model our *in vitro* stem cell-based model of epiblast cyst morphogenesis. In Chapter two, we presented a novel machine learning-assisted imaging pipeline that permits the careful characterization of cell-level events occurring in our system. While attempts were made to connect specific local environments (cell density) to specific cell actions, a direct causal relationship could not be established. Stochasticity and heterogeneity make the mechanistic studies of our systems difficult. In order to address the need for a platform capable of yielding mechanistic insights, chapter three presented a novel ABM-GA framework for the generation of plausible models of morphogenesis. The framework was shown to be able to generate diverse sets of ABMs that yielded structures with desired patterning. In Chapter four, we applied this framework for the generation of models of asymmetric cysts resembling the amniotic sac. With preliminary *in silico* experiments, we showed that the framework was able to output models that partially captured interesting behaviors of interest, specifically the effect of initial cell number on final cyst composition. We further showed that correct structure formation was heavily impacted by just a few model parameters including inhibitor secretion and morphogenesis rules involving the presence of cells and matrix. Used together, these tools have

the potential to shed light into the mechanisms guiding growth, movement, and cell fate specification in *in vitro* models of human development.

5.2 Future works

5.2.1 Characterization of cell-level events in the PASE model

In Chapter 2 we presented a machine learning-assisted image analysis tool capable of tracking individual cells and recognizing important cell-level events such as death and division. With this integrative pipeline we were able to characterize the cell states and actions during the dynamic growth and morphogenesis of lumenal hESC cysts. A logical future direction of this work is to use this tool to study the development of the PASE model. Preliminary data from our lab has identified ISL1 as a specific marker of nascent amnion. If an amnion reporter line were to be created, we could use live cell imaging to gain insight into the process of symmetry breaking. Some interesting questions to tackle are: (1) Which cell or cells are the first to differentiate? (2) Coming from an agent-based modeling perspective, can we predict which cells will differentiate based on the cell's local neighborhood composition? (3) What are the temporal dynamics of differentiation (i.e. in what order do cells differentiate and when does this differentiation stop to give rise to a stably asymmetric cyst)? (4) Why can different initial cell numbers lead to different cyst compositions? To start tackling this question we can start by investigating cell-level events in cell clusters of different sizes. Are cells generally following the same behaviors or is cluster size changing cell behavior? Understanding how initial conditions affect symmetry breaking could not only shed light into the mechanisms involved in symmetry breaking but also help in the development of strategies to increase the efficiency and facilitate future mechanistic studies.

5.2.2 Modifying GA to generate more robust models of morphogenesis and patterning

The current ABM-GA framework is able to yield models that comply with a variety of sections objectives. Desired behaviors such as different cyst outcomes depending on the initial cell cluster size are not considered during evolution. Future work with the ABM-GA framework could involve the incorporation of this and other behaviors as additional fitness objectives. For the purpose of considering the effect of initial cell number, the models would have to run at least three times (low, medium, and high starting cell number) during the evaluation. The population could then evolve in such a way that models that are able to yield fully differentiated cysts at low starting numbers, fully pluripotent cysts at high starting numbers, and asymmetric cysts at an intermediate starting number are favored during the selection process. Because simulations in NetLogo are one of the components of the framework that take the most time, the increase in model runs would make this approach more computationally expensive. To mitigate this, future work could be done to utilize more efficient ABM platforms such as MESA, a Python ABM library.

In Chapter 3 we introduced the construction of models with a controllable percentage of nonfunctional morphogenesis rules in order to mitigate possible overfitting. Results showed that, when starting with a percentage of nonfunctional rules of 70, complexity remains fairly constant throughout evolution. This result indicates that morphogenesis and patterning can be achieved with simple rule sets. In future work, the percentage of nonfunctional rules could be increased, even to 100 percent. It would be interesting to investigate how model complexity evolves and to determine the minimum set of rules required for successful structure formation. Further, models of differing complexities could be compared to determine the effect of complexity on model robustness.

5.2.3 In depth analysis of the two-section cyst ABM to investigate possible mechanisms at work in our in vitro amniotic sac model

In this work, we established the potential of the framework-generated ABMs for studying morphogenesis and patterning. More work can be done, however, to use this framework's output to its fullest potential. More KO experiments can be carried out in select models in order to determine which rules are most important for morphogenesis and patterning. Results can be compared between different types of models in order to determine if there are any rules that are indispensable for correct structure formation. Care should be taken so that *in silico* KO experiments can be mapped to *in vitro* KO experiments for validation of findings. Further validation of the ABM can involve experiments with initial activator, initial cell number, and addition of inhibitor. By comparing *in silico* and *in vitro* results, models can be either discarded or validated. Validated models can then be used to carry out further *in silico* experiments and gain novel insights into the mechanisms at work during morphogenesis and patterning.

In Chapter 2 we showed that we are capable of tracking individual cells and recognizing important cell-level events such as death and division. Having obtained the cell tracking output from various videos, we sought to relate these events to specific local environments (i.e. number of cell neighbors and distance to each one). Whether because of the effect of stochasticity or a lack of information, a causal relationship could not be established, even with the use of machine learning approaches. Having a set of potential ABMs, it would be interesting to see how closely the *in vitro* system behaves to the *in silico* models.

5.2.4 Use of ABM-GA framework to explore general mechanisms behind morphogenetic events in hPSC models

The field of stem cell-based models of development has advanced greatly in the last few decades. We have seen that hPSCs have an amazing ability to self-organize and recapitulate key morphogenetic events seen *in vivo*. In addition to our amniotic sac model, researchers have developed models of the neural tube, the optic cup, and the intestine among many others[33,116,117]. Events including polarization, lumenogenesis, and symmetry breaking are seen in most if not all the systems. Further, the systems show a high sensitivity to initial cell cluster size. While each system has been studied in isolation in order to understand the mechanisms behind growth and differentiation, there exists no unifying mechanistic picture that is able to capture the principles behind self-organization and explain the role of cell plating density on symmetry breaking. Using our ABM-GA framework we can begin to tackle this challenging task. The work would begin by generating models that capture morphogenesis and patterning as seen in different hPSC-based models. Having constructed the models, we can explore commonalities between the models and try to identify the rules that play the most important roles in correct growth and differentiation.

Bibliography

1. Ma, H. *et al.* In vitro culture of cynomolgus monkey embryos beyond early gastrulation. *Science (80-.).* **366**, (2019).
2. Niu, Y. *et al.* Dissecting primate early post-implantation development using long-term in vitro embryo culture. *Science (80-.).* **366**, eaaw5754 (2019).
3. Sasaki, K. *et al.* The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Dev. Cell* **39**, 169–185 (2016).
4. Nakamura, T. *et al.* A developmental coordinate of pluripotency among mice, monkeys and humans. *Nature* **537**, 57–62 (2016).
5. Sawai, T., Hatta, T. & Fujita, M. Japan Significantly Relaxes Its Human-Animal Chimeric Embryo Research Regulations. *Cell Stem Cell* **24**, 513–514 (2019).
6. Harrison, S. E., Sozen, B., Christodoulou, N., Kyprianou, C. & Zernicka-Goetz, M. Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. *Science (80-.).* **356**, (2017).
7. Beccari, L. *et al.* Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. *Nature* **562**, 272–276 (2018).
8. Rivron, N. C. *et al.* Blastocyst-like structures generated solely from stem cells. *Nature* **557**, 106–111 (2018).
9. Sozen, B. *et al.* Self-assembly of embryonic and two extra-embryonic stem cell types into gastrulating embryo-like structures. *Nat. Cell Biol.* **20**, 979–989 (2018).
10. Li, R. *et al.* Generation of Blastocyst-like Structures from Mouse Embryonic and Adult Cell Cultures. *Cell* **179**, 687-702.e18 (2019).
11. Rossant, J. & Tam, P. P. L. New Insights into Early Human Development: Lessons for Stem Cell Derivation and Differentiation. *Cell Stem Cell* **20**, 18–28 (2017).
12. Guo, G. *et al.* Trophectoderm Potency is Retained Exclusively in Human Naïve Cells. *bioRxiv* 2020.02.04.933812 (2020) doi:10.1101/2020.02.04.933812.
13. Blakeley, P. *et al.* Erratum to Defining the three cell lineages of the human blastocyst by single-cell RNA-seq (Development, (2015) 142, 3151-3165). *Dev.* **142**, 3613 (2015).
14. Molè, M. A., Weberling, A. & Zernicka-Goetz, M. Comparative analysis of human and mouse development: From zygote to pre-gastrulation. *Curr. Top. Dev. Biol.* **136**, 113–138 (2020).
15. Deglincerti, A. *et al.* Self-organization of human embryonic stem cells on micropatterns. *Nat. Protoc.* **11**, 2223–2232 (2016).
16. Shahbazi, M. N. *et al.* Self-organization of the human embryo in the absence of maternal tissues. *Nat. Cell Biol.* **18**, 700–708 (2016).
17. Xiang, L. *et al.* A developmental landscape of 3D-cultured human pre-gastrulation embryos. *Nature* (2019) doi:10.1038/s41586-019-1875-y.
18. Hyun, I., Wilkerson, A. & Johnston, J. Embryology policy: Revisit the 14-day rule. *Nature* **533**, 169–171 (2016).
19. Rivron, N. & Pera, M. Debate ethics of embryo models from stem cells. *Nature* **564**, 183–185 (2018).
20. Rossant, J. & Tam, P. P. L. Exploring early human embryo development. *Science (80-.).* **360**, 1075–1076 (2018).
21. Shahbazi, M. N., Siggia, E. D. & Zernicka-Goetz, M. Self-organization of stem cells into

embryos: A window on early mammalian development. *Science (80-.).* **364**, 948–951 (2019).

22. Takahashi, K. *et al.* Induction of Pluripotent Stem Cells from Adult Human Fibroblasts by Defined Factors. *Cell* **131**, 861–872 (2007).

23. Thomson, J. A. Embryonic stem cell lines derived from human blastocysts. *Science (80-.).* **282**, 1145–1147 (1998).

24. Yu, J. *et al.* Induced Pluripotent Stem Cell Lines Derived from Human Somatic Cells. *Science (80-.).* **318**, 1917 LP – 1920 (2007).

25. Shao, Y. & Fu, J. Integrated micro/nanoengineered functional biomaterials for cell mechanics and mechanobiology: A materials perspective. *Adv. Mater.* **26**, 1494–1533 (2014).

26. Sun, Y., Chen, C. S. & Fu, J. Forcing Stem Cells to Behave: A Biophysical Perspective of the Cellular Microenvironment. *Annu. Rev. Biophys.* **41**, 519–542 (2012).

27. Xue, X. *et al.* Mechanics-guided embryonic patterning of neuroectoderm tissue from human pluripotent stem cells. *Nat. Mater.* **17**, 633–641 (2018).

28. Warmflash, A., Sorre, B., Etoc, F., Siggia, E. D. & Brivanlou, A. H. A method to recapitulate early embryonic spatial patterning in human embryonic stem cells. *Nat. Methods* **11**, 847–854 (2014).

29. Martyn, I., Kanno, T. Y., Ruzo, A., Siggia, E. D. & Brivanlou, A. H. Self-organization of a human organizer by combined Wnt and Nodal signaling. *Nature* **558**, 132–135 (2018).

30. Martyn, I., Brivanlou, A. H. & Siggia, E. D. A wave of WNT signaling balanced by secreted inhibitors controls primitive streak formation in micropattern colonies of human embryonic stem cells. *Dev.* **146**, (2019).

31. Etoc, F. *et al.* A Balance between Secreted Inhibitors and Edge Sensing Controls Gastruloid Self-Organization. *Dev. Cell* **39**, 302–315 (2016).

32. Shao, Y. *et al.* Self-organized amniogenesis by human pluripotent stem cells in a biomimetic implantation-like niche. *Nat. Mater.* **16**, 419–427 (2017).

33. Shao, Y. *et al.* A pluripotent stem cell-based model for post-implantation human amniotic sac development. *Nat. Commun.* **8**, 1–15 (2017).

34. Simunovic, M. *et al.* A 3D model of a human epiblast reveals BMP4-driven symmetry breaking. *Nat. Cell Biol.* **21**, 900–910 (2019).

35. Anderson, K. G. V *et al.* Insulin fine-tunes self-renewal pathways governing naive pluripotency and extra-embryonic endoderm. *Nat. Cell Biol.* **19**, 1164–1177 (2017).

36. Linneberg-Agerholm, M. *et al.* Naïve human pluripotent stem cells respond to Wnt, Nodal and LIF signalling to produce expandable naïve extra-embryonic endoderm. *Development* **146**, dev180620 (2019).

37. Okae, H. *et al.* Derivation of Human Trophoblast Stem Cells. *Cell Stem Cell* **22**, 50-63.e6 (2018).

38. Gamage, T. K., Chamley, L. W. & James, J. L. Stem cell insights into human trophoblast lineage differentiation. *Hum. Reprod. Update* (2016) doi:10.1093/humupd/dmw026.

39. Yang, Y. *et al.* Derivation of Pluripotent Stem Cells with In Vivo Embryonic and Extraembryonic Potency. *Cell* **169**, 243-257.e25 (2017).

40. Yang, J., Ryan, D. J., Lan, G., Zou, X. & Liu, P. In vitro establishment of expanded-potential stem cells from mouse pre-implantation embryos or embryonic stem cells. *Nat. Protoc.* **14**, 350–378 (2019).

41. Gao, X. *et al.* Establishment of porcine and human expanded potential stem cells. *Nat.*

Cell Biol. **21**, 687–699 (2019).

42. Guo, G. *et al.* Naive Pluripotent Stem Cells Derived Directly from Isolated Cells of the Human Inner Cell Mass. *Stem Cell Reports* **6**, 437–446 (2016).

43. Guo, G. *et al.* Erratum: Correction: Epigenetic resetting of human pluripotency (doi:10.1242/dev.146811) (Development (Cambridge, England) (2017) 144 15 (2748-2763) PII: dev166397). *Development* **145**, (2018).

44. Pastor, W. A. *et al.* TFAP2C regulates transcription in human naive pluripotency by opening enhancers. *Nat. Cell Biol.* **20**, 553–564 (2018).

45. Rostovskaya, M., Stirparo, G. G. & Smith, A. Capacitation of human naïve pluripotent stem cells for multi-lineage differentiation. *Dev.* **146**, 1–15 (2019).

46. Zheng, Y. *et al.* Controlled modelling of human epiblast and amnion development using stem cells. *Nature* **573**, 421–425 (2019).

47. Kunath, T. *et al.* Imprinted X-inactivation in extra-embryonic endoderm cell lines from mouse blastocysts. *Development* **132**, 1649–1661 (2005).

48. Sozen, B. *et al.* Self-Organization of Mouse Stem Cells into an Extended Potential Blastoid. *Dev. Cell* **51**, 698-712.e8 (2019).

49. Posfai, E. *et al.* Defining totipotency using criteria of increasing stringency. *bioRxiv* 2020.03.02.972893 (2020) doi:10.1101/2020.03.02.972893.

50. D'Arcy, W. T. On growth and form. *Cambridge Univ. Press* **1**, 7 (1942).

51. Newcombe, H. B., McElroy, W. D. & Glass, B. The Chemical Basis of Heredity. *AIBS Bull.* **7**, 55 (1957).

52. Gierer, A. & Meinhardt, H. A theory of biological pattern formation. *Kybernetik* **12**, 30–39 (1972).

53. Oster, G. F. Lateral inhibition models of developmental processes. *Math. Comput. Model.* **12**, 1179 (1989).

54. Segel, L. A. & Jackson, J. L. Dissipative structure: an explanation and an ecological example. *J. Theor. Biol.* **37**, 545–559 (1972).

55. Vempati, P., Popel, A. S. & Mac Gabhann, F. Formation of VEGF isoform-specific spatial distributions governing angiogenesis: computational analysis. *BMC Syst. Biol.* **5**, 1–25 (2011).

56. Hashambhoy, Y. L., Chappell, J. C., Peirce, S. M., Bautch, V. L. & Mac Gabhann, F. Computational modeling of interacting VEGF and soluble VEGF receptor concentration gradients. *Front. Physiol.* **2**, 62 (2011).

57. Tveito, A. *et al.* Slow Calcium-Depolarization-Calcium waves may initiate fast local depolarization waves in ventricular tissue. *Prog. Biophys. Mol. Biol.* **110**, 295–304 (2012).

58. Liu, G., Qutub, A. A., Vempati, P., Mac Gabhann, F. & Popel, A. S. Module-based multiscale simulation of angiogenesis in skeletal muscle. *Theor. Biol. Med. Model.* **8**, 1–26 (2011).

59. Laise, P., Di Patti, F., Fanelli, D., Masselli, M. & Arcangeli, A. Deterministic and stochastic aspects of VEGF-A production and the cooperative behavior of tumoral cell colony. *J. Theor. Biol.* **272**, 55–63 (2011).

60. Glen, C. M., Kemp, M. L. & Voit, E. O. *Agent-based modeling of morphogenetic systems: Advantages and challenges. PLoS Computational Biology* vol. 15 (2019).

61. Wolpert, L. Positional information and pattern formation. *Curr. Top. Dev. Biol.* **6**, 183–224 (1971).

62. Wolpert, L. Positional information. *eLS* 1–5 (2016).

63. Wolpert, L. Positional information and the spatial pattern of cellular differentiation. *J. Theor. Biol.* **25**, 1–47 (1969).

64. Neumann, J. von. Theory of self-reproducing automata. *Ed. by Arthur W. Burks* (1966).

65. Alarcón, T., Byrne, H. M. & Maini, P. K. A cellular automaton model for tumour growth in inhomogeneous environment. *J. Theor. Biol.* **225**, 257–274 (2003).

66. Gerlee, P. & Anderson, A. R. A. The evolution of carrying capacity in constrained and expanding tumour cell populations. *Phys. Biol.* **12**, (2015).

67. Macklin, P., Edgerton, M. E., Thompson, A. M. & Cristini, V. Patient-calibrated agent-based modelling of ductal carcinoma in situ (DCIS): From microscopic measurements to macroscopic predictions of clinical progression. *J. Theor. Biol.* **301**, 122–140 (2012).

68. Scott, J. G., Hjelmeland, A. B., Chinnaiyan, P., Anderson, A. R. A. & Basanta, D. Microenvironmental Variables Must Influence Intrinsic Phenotypic Parameters of Cancer Stem Cells to Affect Tumourigenicity. *PLoS Comput. Biol.* **10**, (2014).

69. Bentley, K., Gerhardt, H. & Bates, P. A. Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation. *J. Theor. Biol.* **250**, 25–36 (2008).

70. Brown, B. N. *et al.* An agent-based model of inflammation and fibrosis following particulate exposure in the lung. *Math. Biosci.* **231**, 186–196 (2011).

71. Thorne, B. C., Hayenga, H. N., Humphrey, J. D. & Peirce, S. M. Toward a multi-scale computational model of arterial adaptation in hypertension: verification of a multi-cell agent based model. *Front. Physiol.* **2**, 20 (2011).

72. Folcik, V. A. *et al.* Using an agent-based model to analyze the dynamic communication network of the immune response. *Theor. Biol. Med. Model.* **8**, 1–25 (2011).

73. Adra, S., Sun, T., MacNeil, S., Holcombe, M. & Smallwood, R. Development of a three dimensional multiscale computational model of the human epidermis. *PLoS One* **5**, e8511 (2010).

74. Das, A., Lauffenburger, D., Asada, H. & Kamm, R. D. A hybrid continuum–discrete modelling approach to predict and control angiogenesis: analysis of combinatorial growth factor and matrix effects on vessel-sprouting morphology. *Philos. Trans. R. Soc. A Math. Phys. Eng. Sci.* **368**, 2937–2960 (2010).

75. Fallahi-Sichani, M., Flynn, J. L., Linderman, J. J. & Kirschner, D. E. Differential risk of tuberculosis reactivation among anti-TNF therapies is due to drug binding kinetics and permeability. *J. Immunol.* **188**, 3169–3178 (2012).

76. Walpole, J., Papin, J. A. & Peirce, S. M. Multiscale Computational Models of Complex Biological Systems. *Annu. Rev. Biomed. Eng.* **15**, 137–154 (2013).

77. Merks, R. M. H. & Koolwijk, P. Modeling Morphogenesis *in silico* and *in vitro* : Towards Quantitative, Predictive, Cell-based Modeling. *Math. Model. Nat. Phenom.* **4**, 149–171 (2009).

78. Holland, J. H. *Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence. Adaptation in natural and artificial systems: An introductory analysis with applications to biology, control, and artificial intelligence.* (U Michigan Press, 1975).

79. Holland, J. H. *Adaptation in Natural and Artificial Systems*. (The MIT Pres, 1992). doi:https://doi-org.proxy.lib.umich.edu/10.7551/mitpress/1090.001.0001.

80. Burke, E. K. & Kendall, G. *Search Methodologies*. (Springer US, 2005).

81. Mitchell, M. *An Introduction to Genetic Algorithms*. *An Introduction to Genetic Algorithms* (The MIT Pres, 1998). doi:https://doi.org/10.7551/mitpress/3927.001.0001.

82. Resto Irizarry, A. M. *et al.* Machine learning-assisted imaging analysis of a human epiblast model. *Integr. Biol. (Camb).* **13**, 221–229 (2021).

83. Pera, M. F. Human embryo research and the 14-day rule. *Dev.* **144**, 1923–1925 (2017).

84. Sasaki, K. *et al.* The Germ Cell Fate of Cynomolgus Monkeys Is Specified in the Nascent Amnion. *Dev. Cell* **39**, 169–185 (2016).

85. Resto Irizarry, A. M., Nasr Esfahani, S. & Fu, J. Bioengineered pluripotent stem cell models: new approaches to explore early human embryo development. *Curr. Opin. Biotechnol.* **66**, 52–58 (2020).

86. Taniguchi, K., Heemskerk, I. & Gumucio, D. L. Opening the black box: Stem cell-based modeling of human post-implantation development. *J. Cell Biol.* **218**, 410–421 (2019).

87. Fu, J., Warmflash, A. & Lutolf, M. P. Stem-cell-based embryo models for fundamental research and translation. *Nat. Mater.* **20**, 132–144 (2021).

88. Blanpain, C. & Simons, B. D. Unravelling stem cell dynamics by lineage tracing. *Nat. Rev. Mol. Cell Biol.* **14**, 489–502 (2013).

89. Chen, L. *et al.* Investigating the gene expression profiles of cells in seven embryonic stages with machine learning algorithms. *Genomics* **112**, 2524–2534 (2020).

90. Shah, N. *et al.* An Experiment on Ab Initio Discovery of Biological Knowledge from scRNA-Seq Data Using Machine Learning. *Patterns* **1**, 100071 (2020).

91. Peng, L. *et al.* Single-cell RNA-seq clustering: datasets, models, and algorithms. *RNA Biol.* **17**, 765–783 (2020).

92. Andrews, T. S. & Hemberg, M. Identifying cell populations with scRNASeq. *Mol. Aspects Med.* **59**, 114–122 (2018).

93. Liu, J., Fan, Z., Zhao, W. & Zhou, X. Machine Intelligence in Single-Cell Data Analysis: Advances and New Challenges. *Front. Genet.* **12**, (2021).

94. Guo, J. *et al.* Machine learning-assisted high-content analysis of pluripotent stem cell-derived embryos in vitro. *Stem Cell Reports* **16**, 1331–1346 (2021).

95. Chen, K. *et al.* Branching development of early post-implantation human embryonic-like tissues in 3D stem cell culture. *Biomaterials* **275**, 120898 (2021).

96. Irshad, H. *et al.* Automated mitosis detection using texture, SIFT features and HMAX biologically inspired approach. *J. Pathol. Inform.* **4**, 12 (2013).

97. Liu, A., Hao, T., Gao, Z., Su, Y. & Yang, Z. Nonnegative mixed-norm convex optimization for mitotic cell detection in phase contrast microscopy. *Comput. Math. Methods Med.* **2013**, (2013).

98. Huang, C. H. & Lee, H. K. Automated mitosis detection based on eXclusive Independent Component Analysis. *Proc. - Int. Conf. Pattern Recognit.* 1856–1859 (2012).

99. Yang, F., Mackey, M. A., Ianzini, F., Gallardo, G. & Sonka, M. Cell segmentation, tracking, and mitosis detection using temporal context. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **3749 LNCS**, 302–309 (2005).

100. Gallardo, G. M., Yang, F., Ianzini, F., Mackey, M. & Sonka, M. Mitotic cell recognition with hidden Markov models. *Med. Imaging 2004 Vis. Image-Guided Proced. Disp.* **5367**, 661 (2004).

101. Liu, A. A., Li, K. & Kanade, T. Mitosis sequence detection using hidden conditional random fields. *2010 7th IEEE Int. Symp. Biomed. Imaging From Nano to Macro, ISBI 2010 - Proc.* 580–583 (2010) doi:10.1109/ISBI.2010.5490279.

102. Su, Y. T., Lu, Y., Chen, M. & Liu, A. A. Spatiotemporal Joint Mitosis Detection Using

CNN-LSTM Network in Time-Lapse Phase Contrast Microscopy Images. *IEEE Access* **5**, 18033–18041 (2017).

103. Harder, N. *et al.* Automatic analysis of dividing cells in live cell movies to detect mitotic delays and correlate phenotypes in time. *Genome Res.* **19**, 2113–2124 (2009).

104. Nie, W. Z., Li, W. H., Liu, A. A., Hao, T. & Su, Y. T. 3D Convolutional Networks-Based Mitotic Event Detection in Time-Lapse Phase Contrast Microscopy Image Sequences of Stem Cell Populations. *IEEE Comput. Soc. Conf. Comput. Vis. Pattern Recognit. Work.* 1359–1366 (2016) doi:10.1109/CVPRW.2016.171.

105. Li, K. *et al.* Computer vision tracking of stemness. *2008 5th IEEE Int. Symp. Biomed. Imaging From Nano to Macro, Proceedings, ISBI* 847–850 (2008) doi:10.1109/ISBI.2008.4541129.

106. Nasr Esfahani, S. *et al.* Microengineered human amniotic ectoderm tissue array for high-content developmental phenotyping. *Biomaterials* **216**, 119244 (2019).

107. Weng, S., Shao, Y., Chen, W. & Fu, J. Mechanosensitive subcellular rheostasis drives emergent single-cell mechanical homeostasis. *Nat. Mater.* **15**, 961–967 (2016).

108. Fu, J. *et al.* Mechanical regulation of cell function with geometrically modulated elastomeric substrates. *Nat. Methods* **7**, 733–736 (2010).

109. Lacoste, A., Berenshteyn, F. & Brivanlou, A. H. An Efficient and Reversible Transposable System for Gene Delivery and Lineage-Specific Differentiation in Human Embryonic Stem Cells. *Cell Stem Cell* **5**, 332–342 (2009).

110. Watanabe, K. *et al.* A ROCK inhibitor permits survival of dissociated human embryonic stem cells. *Nat. Biotechnol.* **25**, 681–686 (2007).

111. Peng, J. Y., Hsu, C. N. & Lin, C. C. Adaptive image enhancement for fluorescence microscopy. *Proc. - Int. Conf. Technol. Appl. Artif. Intell. TAAI 2010* 9–16 (2010) doi:10.1109/TAAI.2010.13.

112. Farhan, M., Yli-Harja, O. & Niemistö, A. A novel method for splitting clumps of convex objects incorporating image intensity and using rectangular window-based concavity point-pair search. *Pattern Recognit.* **46**, 741–751 (2013).

113. Schindelin, J. *et al.* Fiji - an Open platform for biological image analysis. *Nat. Methods* **9**, (2009).

114. McQuin, C. *et al.* CellProfiler 3.0: Next-generation image processing for biology. *PLoS Biol.* **16**, 1–17 (2018).

115. ICY : A NEW OPEN-SOURCE COMMUNITY IMAGE PROCESSING SOFTWARE Unité d ' analyse d ' images quantitative – CNRS URA 2582 – Institut Pasteur - France. *Development* 234–237 (2011).

116. Sasai, Y. Cytosystems dynamics in self-organization of tissue architecture. *Nature* **493**, 318–326 (2013).

117. Eiraku, M. *et al.* Self-organizing optic-cup morphogenesis in three-dimensional culture. *Nature* **472**, 51–58 (2011).

118. Johnston, S. T., Simpson, M. J., McElwain, D. L. S., Binder, B. J. & Ross, J. V. Interpreting scratch assays using pair density dynamics and approximate Bayesian computation. *Open Biol.* **4**, 140097 (2014).

119. Vo, B. N., Drovandi, C. C., Pettitt, A. N. & Simpson, M. J. Quantifying uncertainty in parameter estimates for stochastic models of collective cell spreading using approximate Bayesian computation. *Math. Biosci.* **263**, 133–142 (2015).

120. Buzbas, E. O. & Rosenberg, N. A. AABC: approximate approximate Bayesian

computation for inference in population-genetic models. *Theor. Popul. Biol.* **99**, 31–42 (2015).

121. Lambert, B. *et al.* Bayesian inference of agent-based models: a tool for studying kidney branching morphogenesis. *J. Math. Biol.* **76**, 1673–1697 (2018).

122. Bengio, Y., Courville, A. & Vincent, P. Representation learning: A review and new perspectives. *IEEE Trans. Pattern Anal. Mach. Intell.* **35**, 1798–1828 (2013).

123. Wang, Z. *et al.* Deep reinforcement learning of cell movement in the early stage of C.elegans embryogenesis. *Bioinformatics* **34**, 3169–3177 (2018).

124. Schmidhuber, J. Deep Learning in neural networks: An overview. *Neural Networks* **61**, 85–117 (2015).

125. Lecun, Y., Bengio, Y. & Hinton, G. Deep learning. *Nature* **521**, 436–444 (2015).

126. Ribeiro, M. T., Singh, S. & Guestrin, C. 'Why Should I Trust You?' Explaining the Predictions of Any Classifier. *NAACL-HLT 2016 - 2016 Conf. North Am. Chapter Assoc. Comput. Linguist. Hum. Lang. Technol. Proc. Demonstr. Sess.* 97–101 (2016) doi:10.18653/v1/n16-3020.

127. Burrell, J. How the machine 'thinks': Understanding opacity in machine learning algorithms. *Big Data Soc.* **3**, 1–12 (2016).

128. Wilensky, U. NetLogo. *Center for Connected Learning and Computer-Based Modeling, Northwestern University, Evanston, IL* http://ccl.northwestern.edu/netlogo/ (1999).

129. Deb, K., Agrawal, S., Pratap, A. & Meyarivan, T. A fast elitist non-dominated sorting genetic algorithm for multi-objective optimization: NSGA-II. *Lect. Notes Comput. Sci. (including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics)* **1917**, 849–858 (2000).

130. Jaxa-Rozen, M. & Kwakkel, J. H. Pynetlogo: Linking netlogo with python. *J. Artif. Soc. Soc. Simul.* **21**, (2018).

131. Dobreva, M. P., Pereira, P. N. G., Deprest, J. & Zwijsen, A. On the origin of amniotic stem cells: of mice and men. *Int. J. Dev. Biol.* **54**, 761–777 (2010).

132. Luckett, W. P. The development of primordial and definitive amniotic cavities in early Rhesus monkey and human embryos. *Am. J. Anat.* **144**, 149–167 (1975).

133. Shao, Y. *et al.* A pluripotent stem cell-based model for post-implantation human amniotic sac development. *Nat. Commun.* **8**, 208 (2017).