Title
Automated tracking of in vitro morphogenesis

Permalink
https://escholarship.org/uc/item/7hv446vr

Author
Joy, David Allen

Publication Date
2020

Peer reviewed|Thesis/dissertation

Automated tracking of in vitro morphogenesis

by

David Joy

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

in

Bioengineering

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO
AND
UNIVERSITY OF CALIFORNIA, BERKELEY
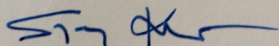
Approved:

DocuSigned by:

Todd McDevitt

Todd McDevitt

38EF3B1D3234412...

Chair

DocuSigned by:

Katherine Pollard

Katherine Pollard

DocuSigned by: 4F1...

Sanjay Kumar

1C64F66851A942B...

_____

_____

Committee Members

Copyright 2020

by

David A Joy

Acknowledgements

I was lucky enough to have the support of many mentors, mentees, collogues, well-wishers, and general fans during my PhD. Without their knowledge, encouragement, and willingness to put up with a good deal of complaining, I would not have made it this far. Thank you to everyone.

I would first like to thank my adviser, Dr. Todd McDevitt, for taking a chance on a somewhat confused computer scientist who knew he wanted to do something generally biology flavored, but had basically no idea what he was doing. He encouraged me to rotate in lab, set me up with an amazing research partner, and generally allowed me to mess around until I found my way into a project. Todd has always supported my weird ideas, occasionally suggesting that I try something a little more feasible, but generally maintaining enthusiasm for my project in the face of many half-baked plots and poorly explained analyses. He has always been happy to consult on graphic design, suggest a collaboration, or buy me the occasional emergency RNAseq run. Because of his mentorship, I have learned how to present clearly, how to write (more) concisely, and how to always keep an eye on the big picture. I'm really grateful to have had Todd as my PI.

I would like to thank my committee members Dr. Katie Pollard and Dr. Sanjay Kumar for their intellectual support of my thesis project and future career. Katie and Sanjay have both provided key feedback, helping to focus my project and keep me on track when I wanted to go off into left field. I'm especially grateful to Katie for allowing me to occasionally invade her lab and bounce ideas off of all the talented grad students

and post docs there, giving me a second home among the computational biologists when the wet lab work got too tough. I would also like to thank Dr. Bruce Conklin, who has acted as a mentor and second adviser, showing up at my bench with wild ideas that turn into productive collaborations and occasionally awkward photo shoots.

For my fellow lab mates, I first need to thank Dr. Ashley Libby. When I joined lab, Todd dropped a very green grad student in her lap, and I will always be thankful for the infinite patience she had with me as she taught me cell culture, molecular biology, every basic bench assay, and basically all of developmental biology. Ashley patiently listened every time my code didn't work, every time the cells didn't grow right, and every time I complained that we needed more robots in lab. Without her insistence on hard deadlines, I probably would have procrastinated my way into a 6 year PhD. We bonded over crazy side-projects, D&D, podcasts, and podcasts about D&D, as well as a mutual love of dinosaurs, even if hers tended to have extraneous fire-breathing powers. I would not be the scientist I am today without her guidance.

Next, I need to thank the students I've mentored over the years, especially Ivana Vasic. I spent a bus ride in my third year trying very hard to talk Ivana into joining our lab, and I can only assume she's regretted ever since. Her energy and deeply inquisitive nature kept me always having to go back to the big picture: why is biology like this? How come we don't know these things? Can't we just stop the cells from doing that? Every time I thought I had come up with an explanation, she pushed back, presenting a new angle, a new insight, or a new weird result to puzzle over. It's been amazing to see

her develop as a scientist, and (in a small way) to help her projects grow and evolve. I already can tell that Ivana will go on to change the world.

As I got a little further along in my PhD, I had the amazing and completely unexpected opportunity to help mentor two bright developmental biologists, Nick Elder and Emily Bulger. When Nick first rotated in our lab, I was expecting a pretty uneventful 6 weeks: we'd help him set up his project, I'd help with some imaging, we'd make a slideshow. Imagine my surprise when he ran in one day with the extremely worrying sentiment that "something weird" had happened. Imagine my further surprise when first he, then Emily, then Ashley and I all managed to repeat his crazy experiment, leading to the project that became the second half of my PhD. I'm thankful to Nick for always having a trick up his sleeve, ready to go whenever we run into any problems, and for being generally upbeat in the face of adversity. I'm thankful to Emily for forcing me to stop, slow down, and explain things better, and for sharing my enthusiasm for Evo Devo. It's been an honor to work with two such talented grad students.

I'd like to thank the other people I've trained, from interns (Jocelyn Cervantes) to research associates (Nik Mendoza-Camacho and Marty Krakora) for teaching me how to plan experiments well in advance, write very detailed protocols, and try to explain my code a little better. I know all of them will go on to do great things.

The rest of the lab has been crucial to my success as well, whether it was Ana's happy dances while doing lab work, or Oriane's shouts of "Eww David!" when I would say something gross, the spirit of camaraderie really helped carry me through the day. Everyone was always happy to drop their more important work to help, whether it was

Contributions

The text of Chapter 2 consists partially of reprints of the material as it appears in "Automated Design of Pluripotent Stem Cell Self-Organization" and "Optogenetic control of Wnt signaling for modeling early embryogenic patterning with human pluripotent stem cells". The texts of Chapters 3 and 4 are reprints of the material as it appears in "Deep neural net tracking of human pluripotent stem cells reveals intrinsic behaviors directing morphogenesis" and "Elongation of Caudalized Human Organoids Mimics Neural Tube Development". The co-author listed in these publications directed and supervised the research that forms the basis of this dissertation.

Anything that happens, happens.

Anything that in happening, causes something else to happen, causes something else

to happen

Anything that in happening, causes itself to happen, happens again.

It doesn't necessarily do it in chronological order, though.

<div align="right">

-Douglas Adams, Mostly Harmless

</div>

Automated tracking of *in vitro* morphogenesis

David Joy

Abstract

Embryogenesis is a critical period in the developmental life cycle of organisms, but the processes that enable multicellular coordination of germ layer formation and the subsequent dynamics of tissue structure assembly are still poorly described at the single cell level. Recent advances in microscopy and computer vision have enabled whole embryo tracking of cell migration, lineage commitment, and tissue formation in model organisms such as fly, zebrafish, frog, and chick, but it is unknown to what extent these organisms recapitulate the complex processes of early human development. Direct interrogation of the development of post-implantation human embryos carries technical and ethical limitations that render *in vivo* study of human embryogenesis largely observational, but recent advances in human induced pluripotent stem cells (hiPSCs) derived organoid models provide a powerful *in vitro* platform to elucidate developmental processes. The four studies described in this dissertation developed biological and computational pipelines to study the interdependencies between cell migration, lineage specification, and morphogenesis during early human development. First, several automated cell tracking algorithms were designed and deployed to study individual cell migration events in models of mosaic pattering and differentiation. Sparsely labeled cell populations were tracked during pluripotency, yielding measures of cell migration pre-differentiation, then cell behavior change kinetics were quantified in a mosaic Wnt differentiation model, demonstrating that automated cell tracking can non-

destructively detect signatures of cell fate transition. Second, collective cell migration was interrogated using a novel ensemble of deep neural cell networks to generate dense spatio-temporal descriptions of hiPSC colony behaviors. Dense cell tracks revealed colony position dependent cell-cell interactions, including correlated migration, changes in cell density, and coordinated multicellular structure formation. Cell behaviors were found to be dependent on substrate, media, and colony size, and striking behavioral transitions were detected during differentiation to germ lineages using multiple protocols, including divergent cell migration patterns induced by Wnt and BMP signaling that nonetheless resulted in convergent germ layer organization. These results demonstrate how robust multicellular pattern formation is achieved through integration of cell intrinsic, local, and global signals. Third, collective migration and body axis emergence was studied in an organoid model of gastrulation and neural tube formation, where initially round organoids spontaneously assembled an organizer like structure and then underwent several days of elongation. Automated tracking of multi-day time lapse videos and segmentation of culture images spanning weeks of differentiation, facilitated quantification of organoid growth, elongation, and stratification, enabling optimization of culture conditions to improve differentiation robustness and repeatability across multiple stem cell lines. Elongation was demonstrated to depend on the emergence of a subpopulation of neuromesodermal progenitors (NMPs) which formed signaling centers reminiscent of the node structures that direct early primitive streak formation *in vivo*. Ablation of the early NMP transcription factor TBXT did not abrogate tail formation, but rather caused multiple spontaneous extensions to form, suggesting a

key role for TBXT expressing cells in enabling the spontaneous symmetry breaking events underlaying the formation of a head-tail axis. Fourth and finally, a mechanistic simulation of the observed TBXT symmetry breaking and tail axis formation events was developed. Cell-cell interaction forces, cell migration parameters, and cell growth were derived from 2D measurements of re-plated organoids, shown to explain observed 3D organoid structure, growth, and elongation rates, and then applied to a 3D particle simulation of tail formation. Coalescence of structures matching the TBXT+ organizers were found to depend on both percentage of TBXT+ cells and the directionality of cell migration, demonstrating that a combination of chemotaxis and population heterogeneity is sufficient to produce organizer-like structures from initially randomly mixed populations. Overall, these studies demonstrate the critical activity of cell migration in coordinating multicellular organization during early embryogenesis, and provide a kinetic, mechanistic platform to interrogate how tissues form at multiple scales across the developmental landscape.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF ABBREVIATIONS

ABM – agent-based model

AP – anterior-posterior

AUC – area under the curve

BDNF – brain derived neurotrophic factor

BMP – bone morphogenic protein

CHIR – Wnt small-molecule agonist CHIR99021

CNN – convolutional neural network

CRISPR - clustered regularly interspaced short palindromic repeats

CRISPRi – CRISPR interference

EMT – epithelial to mesenchymal transition

FCRN – fully convolutional regression network

hESC – human embryonic stem cell

hiPSC – human induced pluripotent stem cell

IFR – inter-frame reliability

IRR – inter-rater reliability

NMP – neuromesodermal progenitor

optoWnt - optogenetically activated Wnt signaling

ROC – receiver operating characteristic

ROCK – rho-associated coiled-coil kinase

TBXT - T-Box transcription factor T, also known as brachyury

WT – wild type

1 Introduction

1.1 Overview

The goal of this dissertation was to interrogate the dynamics of collective cell migration that drive the organization of multicellular structures during embryogenesis. Descriptions of the stereotyped migration patterns of cells and their contribution to the structures of the developing embryo are among the oldest literature in biology. Cell migration is critical to the formation of the main body axes (Aman and Piotrowski, 2010), the formation and elaboration of the heart (Christiaen et al., 2010; George et al., 2020), and the innervation (Ahmed et al., 2017; Uesaka et al., 2016) and vascularization (Azad et al., 2019; Koenig et al., 2016) of the tissues of the body. In particular, the collective cell migration processes that occur during gastrulation are of critical importance to laying out the fundamental anterior-posterior body axis and positioning the three germ layers correctly within the developing embryo (Aman and Piotrowski, 2010; Chuai and Weijer, 2009; Hara et al., 2013). However, mechanistic descriptions of how the embryogenic signaling milieu is integrated to produce cell migratory behaviors and which behaviors are necessary or sufficient to enable tissue formation remain underexplored. This dissertation outlines a number of techniques for measuring and simulating individual and collective cell migration in models of germ layer specification, gastrulation, and anterior-posterior axis formation, to generate descriptions of how individual cells behave during differentiation, and how those behaviors contribute to multicellular patterning.

The goal of the first study was to generate measures of individual human induced pluripotent stem cell (hiPSC) behavior in the context of two-dimensional colonies and differentiating epithelial sheets. A time lapse imaging analysis pipeline was developed capable of tracking sparsely labeled cells within the colony, deriving baseline measures of hiPSC migration in a multicellular context. Titration studies were used to determine ideal density of labeled cells, as well as how different cell tracking algorithms performed as label density and cell number increased. Finally, the tracking algorithm was applied to a study of differentiating hiPSCs under optogenetic Wnt stimulation, revealing behavioral signatures that correspond to acquisition of mesodermal fate and the epithelial to mesenchymal transition (EMT). This study demonstrated that hiPSC colonies are constantly in motion, and that changes in migration dynamics non-destructively reveal cell state transitions during germ layer specification.

Although tracking of sparsely labeled colonies of hiPSCs revealed individual cell behaviors, the low percentage of labeled cells required by conventional cell tracking algorithms limited the study of collective cell migration and cell-cell interactions during pluripotency and differentiation. In the second study, dense, whole-colony cell tracking was achieved through development of a deep learning-based cell detection and tracking pipeline, enabling whole colony monitoring in time lapse studies ranging from hours to days. Dense tracking enabled mapping of cell behavior to position within the colony, demonstrating distinctions between cells near the edge of the colony versus those in the center. Further, by measuring cell neighborhoods, dense tracking enabled correlation of cell migration and study of the formation of multicellular structures such as rings during

mesoderm formation and neural rosettes during ectoderm specification. By measuring whole colony dynamics during differentiation, this platform was able to distinguish between pluripotent cell culture conditions, and to detect divergent cell behaviors between Wnt and BMP germ layer differentiation protocols that otherwise specified superficially similar multicellular patterning, demonstrating convergence among multicellular sorting mechanisms.

While the collective migration phenotypes in 2D formed a rich and complex dataset of cell-cell interactions, the behaviors observed in adherent colonies failed to recapitulate the directional, polarized collective cell migration observed during gastrulation. In the third study, a 3D organoid model of gastrulation and neural tube formation was developed and characterized. Organoids developed polarized cell masses by day 3 that spontaneously elongated in suspension culture, producing organoids with lengths of up to millimeters, and aspect ratios of 10 or more, resembling the "tadpole" stage of embryos *in vivo*. Time lapse imaging of organoids was used to measure organoid growth and extension, as well as multicellular flow into the tail structure, revealing that collective cell migration drove extension. Further, organoids elongated in a Wnt dose dependent manner, with a critical level of Wnt induction of a population of neuromesodermal progenitors (NMPs) identified as required for elongation. Knockdown of the early NMP transcription factor TBXT did not abrogate extension, but rather generated multiple smaller extensions, suggesting a role for TBXT-dependent chemotaxis in generating a unified single tail. This study quantified the contribution of collective cell migration to generating an *in vitro* anterior-posterior axis in

a human system, providing mechanistic insights into the assembly of the human organizer and the formation of the main body plan.

Unfortunately, 3D imaging of whole-organoid elongation was limited by technical issues, requiring a different approach to connect the single cell description of cell migration with the whole organoid description of axial extension. In the fourth study, a mechanistic simulation of organizer formation in neural organoids was developed using empirically derived cell behavior parameters from both differentiating colonies and organoids. Cell parameters for cell-cell adhesion, migration, and cell division were derived from 2D colonies of re-plated neural organoids. The 2D cell-cell measurements were then compared to static dense 3D cell-cell localization and found to be comparable. Based on this congruence, measures of organoid cell number, growth, and migration velocity were then derived from whole organoid time lapse imaging, yielding growth and migration curves that matched extrapolated 2D results. These parsimonious parameters were then employed in a mechanistic simulation of 3D organoids as randomly mixed spheres. Percentage of TBXT+ cells, mutual TBXT+ cell chemoattraction, and surface seeking behaviors were found to be the dominant terms determining the frequency of organizer coalescence, number of organizer sites, and positioning of the organizer clusters. Overall, this study suggests that human embryogenesis may utilize TBXT-dependent chemoattractive organizer assembly to produce the node driving primitive streak formation, and further defines the parametric domain where self-assembling organizers robustly form, enabling optimization of

differentiation and further avenues of study to better understand human body plan organization.

This dissertation provides a comprehensive study of human pluripotent stem cell migratory behavior as individual cells, in collective epithelial sheets, and in 3D during early differentiation and axis formation. It demonstrates that phase separation, multi-cellular structure formation, and high level morphogenic events depend on, and can be predicted by measuring cell migration parameters over time and space. These results enable a behavioral metrology approach complementary to the current focus on transcriptomic, genomic, and epigenomic profiling of single cells, enabling future mechanistic modeling of dynamic tissue formation processes over a broad range of spatiotemporal scales. This introduction provides background for the key topics and approaches applicable to this work, such as early embryogenesis and symmetry breaking, cell migration assays and measurement, *in vitro* models of gastrulation, and *in silico* models of cell organization and pattern formation.

1.2 Symmetry breaking and body plan formation

During embryogenesis, a single cell undergoes a series of stereotyped cell divisions that lead first to the division of extraembryonic, embryonic, and germ cells, then to the formation of individual germ layers (Pfeffer, 2018; Płusa and Piliszek, 2020), and finally through a complex series of migrations and lineage commitments that properly position the germ layers in relation to one another and establish the primary anterior-posterior (AP) body axis (Bénazéraf and Pourquié, 2013; McDole et al., 2018; Shah et al., 2019). A critical stage during embryo development is gastrulation, where a

radially symmetric blastocyst with no apparent AP axis undergoes a polarized symmetry breaking event (Blum et al., 2014; Schiffmann, 2006), simultaneously generating the head-tail body plan while also positioning progenitors of the three principal germ layers in a well-defined spatial arrangement (Shah et al., 2019; Winklbauer and Parent, 2017). Multiple deterministic mechanisms driving polarization have been previously demonstrated, ranging from distribution of maternal RNAs (Wieschaus, 2016), to signaling events during sperm entry (Wu and Schöler, 2016), to the mechanical environment post-implantation (Hiramatsu et al., 2013). However, the robustness of axis formation to both cell ablation (Takaoka et al., 2017) and randomization (Fulton et al., 2020) suggests that compensatory, active mechanisms must act to ensure the formation of a singular organizing axis, correctly placed.

In many model organisms, an organizer, a small pocket of pre-patterned cells which are independently capable of inducing gastrulation events, forms within the larger embryo pre-gastrulation (Anderson and Stern, 2016). Organizers secrete pro-gastrulation factors such as Nodal and are inhibited by various factors such as BMPs secreted by the trophectoderm, creating classic reaction-diffusion system where ubiquitous, rapidly diffusing inhibitors interact with localized, slowly diffusing activators to produce non-linear pattern responses reminiscent of many developmental structures (Tewary et al., 2017, 2018). This developmental pattern has elegant robustness properties, where an embryo with too few cells in the epiblast will be inhibited from undergoing gastrulation due to dominance of inhibition, while an embryo that reaches the correct size threshold spontaneously undergoes symmetry breaking, protecting

against many forms of cell ablation or misallocation (Takaoka et al., 2017; Tewary et al., 2018). Further, the ectopic axis induction properties of organizer transplants suggests that polarized organizers need only be generated from a few cells (Bertocchini and Stern, 2002), reducing the complexity of orchestrating multicellular assembly during the earliest stages of embryogenesis.

In many bilaterian organisms, the identity and position of organizer cells is defined early in embryogenesis, determined by segregation of maternal RNA in the oocyte in *drosophila* (Wieschaus, 2016), or location of sperm entry in and subsequent hemispheric polarization in amphibians (Nieuwkoop, 1985). However, in mammalian embryogenesis, organizer populations are much less well defined, with evidence that organizers control only partial body axis specification (Morgani and Hadjantonakis, 2020; Tam and Behringer, 1997), or that mammalian embryos can spontaneously regenerate organizers from mixtures of epiblast and primitive endoderm cells (Bedzhov et al., 2014). Additionally, the peri-implantation nature of gastrulation forces mammalian embryonic development to be ambivalent to initial embryo orientation relative to the uterine wall (Hiramatsu et al., 2013). These factors combine to suggest that mammalian organizer positioning must be defined spontaneously, enabling robustness against the stochastic nature of embryogenesis.

Although mammalian embryonic development has been well described, especially in murine systems, there are a number of distinctions that make extrapolation to human embryogenesis problematic. The order primates diverged from other mammals approximately 55 million years ago (Ezran et al., 2017; Luo, 2007), and the

closest extant ape genus *Pan* diverged approximately 6 million years ago (Chintalapati and Moorjani, 2020; Rolian, 2014), leading to large differences in developmental trajectory with common small-animal models (Kelly and Sears, 2011; Milinkovitch and Tzika, 2007; Selwood and Johnson, 2006), and small but significant distinctions when studying developmental processes in even close relative species (Rolian, 2014). Direct study of human embryonic development pre-implantation has been common, leading to revolutions in the efficiency and efficacy of medical procedures such as *in vitro* fertilization and embryo genetic testing, but the peri-implantation nature of gastrulation gives rise to both technical and ethical limitations to the study of human anterior-posterior patterning. To complement *in vivo* models of development, *in vitro* studies using human stem cell derived organoids can be employed to dissect which aspects of human development are evolutionarily conserved, and which processes are unique to fully human embryogenesis.

1.3 Migration and lineage tracing

Lineage tracing studies are among the oldest developmental biology assays, with the lineage of the nematode *C. elegans* among the first to be fully defined in the literature (Sulston et al., 1983). Higher organisms such as chick were first traced by embryo hybridization studies with quail or duck embryos with distinctive cell sizes (Catala et al., 1995). As genetic reporter constructs became available in models such as mouse or zebrafish, lineage tracing became much easier, with transcription of fluorescent proteins driven by cell-type specific transcription factors that marked cells of a particular fate (Buckingham and Meilhac, 2011). Further sophistication was achieved

with the introduction of Cre-activated lineage tracing systems which stochastically

activated or inhibited fluorophores when driven either by a specific transcriptional event,

or by external chemical stimulus (Cai et al., 2013; Hadjieconomou et al., 2011; Pan et

al., 2011) aiding in innumerable cell lineage tracing assays, including determining

precursors of neurogenesis, organ development, and even hair follicle growth

(Buckingham and Meilhac, 2011). While many of these cell visualization tools are

constitutively expressed and can be assayed non-destructively, limitations of optical

clarity and large scale motion have prevented the use of live reporters in studying all but

the earliest stages of *in vivo* mammalian development (McDole et al., 2018; Pantazis

and Supatto, 2014).

In contrast, study of *in vitro* mammalian cell migration was initially performed on

cells or colonies plated onto 2D surfaces and were primarily concerned with the

diffusive migratory behavior of metastatic cancer cell lines (Deryugina and Bourdon,

1996; Parent, 1999) or with the migration speed of fibroblasts and neutrophils along

one-dimensional patterned tracks (Doyle et al., 2009; Maiuri et al., 2012). These

migration assays established that cells migrate stochastically, switching between an

active directional migration mode where a cell can move up to one cell length every few

minutes to a passive mode where almost no motion occurs (Ascione et al., 2016;

Campos et al., 2010) analogous to a biased random walk (Campos et al., 2010; Patlak,

1953). Paradoxically, the diffusive and stochastic properties of individual cell migration

have been shown to contribute to the formation of ordered structures during cell-sorting

and boundary formation events (Das et al., 2017; Krupa et al., 2014; Wang et al., 2017),

suggesting that individual cell migration enhances robustness of tissue segregation during embryogenesis.

Cells undergoing collective migration in epithelial sheets exhibit much more ordered behavior. In the classic wound scratch assay, cells at the leading edge of the wound sense the absence of cell neighbors, in addition to directional sources of damage signals from across the wound gap, and hence undergo a directed migration to close the gap, dragging the cell sheet behind them through cell-cell adhesion and chemotactic signals (Arciero et al., 2011; Yarrow et al., 2004). Cells can undergo highly organized migration of hundreds to thousands of cells, each acting primarily under the influence of local migratory and adhesion signals, to produce higher-order collective organization events such as a "zippering" or "string of pearls" closure (Hashimoto et al., 2015; Pilhwa Lee and Charles W. Wolgemuth, 2011). Although the individual signaling factors and adhesion molecules differ, similar collective migration mechanisms are at work throughout embryogenesis, including during neural tube closure, heart septation, and gut tube elaboration (Anderson et al., 2003; Choi et al., 2017; Hashimoto et al., 2015). Hence, collective cell migration under chemotaxis represents an ordering principle that robustly accommodates and counteracts perturbations during development and organism damage.

During mammalian embryogenesis, individual cells within the epiblast receive a milieu of signals directing both the maintenance of their pluripotent state and instructing differentiation to each of the three major germ lineages. These contradictory signals are further complicated by the stochastic nature of both individual cell signal receptivity and

disparate cell potential dispersed among otherwise undifferentiated epiblast cells (Chazaud et al., 2006; Krupa et al., 2014). Hence, cells require a set of secondary post-commitment organizing principals to ensure that individual germ layers are organized into large scale multicellular agglomerations rather than randomly dispersed among a disordered bulk cell population. This dissertation demonstrates several modes of cell migration that enable lineage organization, with heterogeneously induced cells of each lineage migrating towards and ultimately coalescing with cells of similar cell fate, leading to a phase separation of germ layers preceding gastrulation.

1.4 Models of in vitro symmetry breaking

Because of the difficulties of studying *in vivo* human symmetry breaking, *in vitro* models of multi-cellular organization provide a window into the myriad processes involved in human embryonic development (Hayashi, 2019; Morgani and Hadjantonakis, 2020; Warmflash et al., 2014). In particular, organoids, 3D aggregates of human stem cells which undergo co-differentiation to multiple interacting cell types that form multicellular structures reminiscent of organs *in vivo,* provide a rich model of multicellular organization (Clevers, 2016; Lou and Leung, 2018). Organoids generate impressive mesoscale organization, with robust assembly of functional units such as nephrons in kidney organoids (Combes et al., 2019), villi in gut tube organoids (Silva et al., 2020), or islet structures in pancreatic organoids (Hohwieler et al., 2017), occurring at the scale from 100s of microns to occasionally millimeter scales (10s to 100s of cell lengths). However, organoids are disordered at a macroscale, lacking organization of multicellular units into functional complexes, failing to maintain boundaries and ordered

interactions between cells of different germ layer origins, and missing topological features such as chambers or tubes required for organ function (Takebe and Wells, 2019). *In vivo*, tissue formation is patterned by, and mechanically coupled with body plan elaboration, with a combination of signaling gradients, migration events, and boundary formation events enabling proper placement of organ forming cells in the correct four dimensional spatiotemporal environment (McDole et al., 2018; Sutherland, 2016). The plasticity of whole organism development prevents simple decoupling of body formation processes from the elaboration of individual organ systems, so defining which mechanical and signaling stimuli are necessary and sufficient to produce robust macroscale organ formation *ex vivo* remains an area of active research.

Generating *in vitro* global patterning phenotypes has been primarily successful through simulating the patterning events found in gastrulation. 2D hiPSC colonies under BMP4 stimulation reliably generate the three germ layers and an extraembryonic outer layer in a system of concentric rings, following what appears to be a classic reaction diffusion system under control of a mutually repressive BMP4-NOG interaction (Tewary et al., 2017; Warmflash et al., 2014). In 3D, even more striking behavior has been found in gastruloids, organoid models that spontaneously generate a polarized mass of cells and undergo elongation (Beccari et al., 2018; van den Brink et al., 2014; Marikawa et al., 2020). Gastruloid models have shown patterned expression of Hox genes that recapitulate formation of the anterior-posterior axis, polarized signaling and gene expression gradients, and segregation of germ layer compartments. Further, gastruloids directed towards central nervous system fates have been shown to generate somites

and spontaneously generate organized dorso-ventral axes, demonstrating both boundary formation and spontaneous formation of multiple body axes (Bérenger-Currias et al., 2020; van den Brink et al., 2020; Veenvliet et al., 2020). Despite these striking models of global patterning, it remains unclear which multicellular processes enable polarization and elongation, how individual cell types contribute to the elongation process, and to what extent *in vitro* organoid models recapitulate analogous body formation processes *in vivo.*

1.5 Tracking cell migration

Image analysis techniques to trace cell migration have historically employed a population of cells that are physically distinctive such as with embryo hybridization assays (Catala et al., 1995), or are marked with a reporter construct as in lineage tracing assays (Cai et al., 2013), enabling a reconstruction of the most probable migration path by linking individual cell detections over time. Modern imaging techniques such as confocal and light sheet microscopy enable non-destructive time lapse imaging of fluorescent protein expressing populations within embryos, allowing higher sampling rates and the quantification of cell migration and cell divisions over time (Keller et al., 2008; McDole et al., 2018). Imaging whole embryos is limited by both the size of the embryo and the depth of the tissue to be imaged, with early stage gastrulas relatively optically transparent while even relatively shallow tissues in late stage embryos may be inaccessible below a few millimeters in depth (Pantazis and Supatto, 2014). As the complexity of the cell migration event increases, the number of cells which need to be marked, the minimum field of view, or the temporal sampling rate

might each or all increase, rapidly outpacing the limits of manual cell annotation and necessitating computational approaches (Amat et al., 2015).

Automated cell tracking approaches utilize a number of features to first segment cells and then link those segmentations over time. Fluorescent labeling of the cell cytoplasm is the most common approach for detecting cells, with automated algorithms able to separate labeled cells from background, generating a contour or surface that gives both cell position and shape information, although dense cytoplasmic labels make it difficult to separate one cell from another (Kheradmand et al., 2017; Stegmaier et al., 2016; Zhang et al., 2015). For mononucleated cells types, cell position can also be reliably be determined by labeling the nuclei alone, typically at much higher labeling densities including 100% cell labeling, although the loss of shape information can impact both cell phenotype classification and tracking accuracy (Coelho et al., 2009; Cohen et al., 2017; Lou et al., 2014). Fusion approaches often employ both a nuclear and cytoplasmic marker, gaining shape information from cytoplasmic tracking while still being able to separate nearby cells using nucleus position (Caicedo et al., 2019; Kang et al., 2013). Depending on the biological question, sparse information about cell shape or dense information about cell position may be more relevant, leading to divergent tracking and labeling approaches.

Once a fluorophore strategy is established, cells or nuclei need to be segmented from the background and from each other. Well mixed, sparsely labeled colonies can typically be segmented with a simple background segmentation algorithm that estimates a background fluorescence threshold, followed by a labeling pass to assign pixels

above the threshold to contiguous masks representing the cell contour (Coelho et al., 2009). Densely labeled colonies or colonies where the labeled population undergoes significant clonal expansion require additional steps to calculate boundaries between cells, with various approaches exploiting information about nuclei position (if available), cell texture information, or priors about cell packing and boundary shape (Meijering et al., 2009). Dense segmentation has been enabled by recent advances in machine learning, which deep neural networks approaching and sometimes surpassing human generated cell segmentations (Chowdhury et al., 2017; Falk et al., 2019; Moen et al., 2019). In 3D volumes, further information about cell polarization and distinctions cell shapes can also be incorporated to improve boundary calculation, although computational complexity grows very rapidly with volume size, limiting the application of highly accurate segmentation methods (Amat et al., 2015; Pantazis and Supatto, 2014). Even in 2D, highly accurate cell contours are often difficult to robustly calculate for all cells in a colony due to density of cell packing, suggesting that contour and surface-based representations may not be appropriate for large image volumes.

Instead of generating complete, pixel accurate maps of the cytoplasm or nuclear shape, the center of mass of the cell or nuclei can be estimated, resulting in a simpler and computationally more efficient particle localization (Meijering et al., 2012). Fluorescence detection from round structures such as nuclei has a distinctive single peak near the center of mass of the nucleus, enabling computationally efficient approaches such as peak detection with non-maximum suppression to produce very accurate segmentations even in highly dense fields of labeled nuclei such as in

15

hematoxylin and eosin or Hoechst staining (Libby et al., 2018). Accuracy can again be enhanced with deep learning approaches, leading to state of the art cell counting and tracking, even with a single nuclear label (Cohen et al., 2017; Xie et al., 2016, 2018). Nuclear tracking scales well to 3D volumes, with density peak-detection approaches operating efficiently even on large volumes (Meijering et al., 2012). While the loss of cell shape information makes particle detectors inappropriate for assays like cell type classification, the simplicity and scalability of the center of mass representation makes it invaluable for efficient analysis of large numbers of cells.

Once cells have been detected, individual detections need to be linked into cell tracks across time series. For sufficiently rapidly sampled cells, cell contours or cell centers of mass move much less than one cell body length per frame, so greedy linkage of cells in the previous frame to their nearest neighbor in the next is both simple to implement and computationally efficient (Coelho et al., 2009; Meijering et al., 2012). Extremely rapid sampling of cell positions has a number of downsides including increased data set size, mechanical limits on the volume that can be sampled within an acquisition period, and, critically, increased fluorescent exposure time leading to cytotoxicity and abnormal cell behavior (Meijering et al., 2012). For even short term experiments of minutes to hours, sampling rates typically are too low to guarantee cell migration cannot exceed a cell body length, so more sophisticated tracking algorithms compensate by incorporating prior cell shape and size information, likely cell direction of travel, and cutoffs to prevent spurious linkages of cell tracks that cannot be positively assigned to a single detection in the new frame (Debeir et al., 2005; Gold and

Rangarajan, 1996; Peng et al., 2010). For tracking assays across longer periods of

hours to days, additional complications such as cell division, cell death, and migration in

and out of the field of view need to be considered, all of which come with additional

computational penalties to recognize and account for the event (Piltti et al., 2018).

Finally, the fidelity of the track reconstruction needs to be considered: a fine-grained

lineage trace of a few, sparsely labeled cells can be reconstructed using precise,

computationally intensive methods without incurring an analytical penalty, while mass

tracking of collective migration may be tolerant of a high rate of track breakage, cell mis-

assignments, and mis-classified cell-state transitions without impacting measurement of

overall material properties such as average speed, direction, and persistence of

migration of a cell sheet. The complexity of cell tracking in time lapse microscopy

means that no one-size-fits-all algorithm will satisfy all experimental designs, and

generating gold standard tracking data sets and metrics remains an open research

problem (Caicedo et al., 2019; Coelho et al., 2009).

1.6 Models of in silico symmetry breaking

The complexity of *in vivo* and even *in vitro* cell behavior has led to a large family

of mathematical models that attempt to provide mechanistic explanations for patterning

events. Initial models of symmetry breaking were built on non-linear phase separation

events in physical systems such as Ising spin models (Yang, 1952) and reaction

diffusion systems (Turing, 1952), which were initially used explain self-assembling

processes in coat patterning, but also have been applied to almost every aspect of

developmental patterning (Marcon and Sharpe, 2012). Continuum descriptions of

symmetry breaking invoke signaling gradients of cytokine diffusion, chemoattractive or repulsive nodes that coordinate cell migration, and anisotropic variations in the extracellular stiffness or directionality that all can provide polarizing signals to individual cells (Browning et al., 2018; González-Valverde and García-Aznar, 2018; Plank and Simpson, 2013). While continuous approaches provide simple, often computationally tractable descriptions of cell behavior, they are difficult to integrate with the fundamentally discrete nature of individual cells, and hence can only provide descriptions at coarse scales if they are not complemented by other methods.

In contrast, agent based models (ABMs) provide a cell-centric description of behavior capable of recapitulating many emergent multicellular events (Glen et al., 2019; Norfleet et al., 2020). The classic Cellular-Potts models are reminiscent of Ising models, representing cells as 4-connected regions on a 2D lattice which attempt to expand or contract from regions according to probabilistic rules (Graner and Glazier, 1992; Voss-Böhme, 2012). Even simple deterministic Cellular-Potts-like systems such as the Game of Life are Turing complete (Bak et al., 1989), while more complicated probabilistic models can simulate many of the aspects of tissue development, including cell growth and division, cell death, cell migration, and changes in cell lineage (Libby et al., 2019). The effects of continuous fields and forces can be incorporated through numerical integration, enabling Cellular-Potts models to simulate reaction-diffusion processes and other non-linear patterning events (Glen et al., 2019; González-Valverde and García-Aznar, 2018). Further extensions incorporate simulation of internal gene regulatory or cell-cell signaling networks, allowing coupling of cell state transitions to

models of internal transcription factor expression and external signal transport (Bauer et al., 2010; Mehdipour et al., 2018; Yachie-Kinoshita et al., 2018). Extensions of the Cellular-Potts framework to 3D can simulate complex morphogenic events, but the geometric increase in computational complexity places limits on the size of phenomena that can be tractably simulated (Madhikar et al., 2018; Tapia and D'Souza, 2011). Still, Cellular Potts models can provide physically plausible simulations of cell-cell interaction and pattern formation.

The final relevant class of cell simulations are particle-based models. Unlike Cellular-Potts or continuous models, cells in a particle simulation are represented by a geometric object such as a sphere or ellipse, and interactions between cells are evaluated using quasi-elastic spring force calculations between nearest neighbors (González-Valverde and García-Aznar, 2018; González-Valverde et al., 2016; Norfleet et al., 2020). Despite these simplifications, particle simulations can account for complex behaviors such as single cell and collective migration, cell division, and response to external stimuli under a single force balance framework (González-Valverde et al., 2016). Because the geometric representation of the cell is heavily simplified, operations such as directional migration, cell-cell collision detection, complex fluid flows, and stochastic cell-state transition matrices can easily be extended to simulations of 10s to 100s of thousands of cells on modern CPU or GPU hardware (Tang and Karniadakis, 2014). Further, particle simulations pay a relatively small computational scaling penalty when extended to 3D, enabling simulation of flows in organoids and developing tissues in addition to more conventional colony or sheet migrations. Hence, in cases where cell

shape information is not a dominant factor when modeling morphogenesis, particle

simulation represents a flexible tool for computational descriptions of tissue evolution.

1.7 Concluding Remarks

The complex process of embryogenesis is marked by a series of symmetry

breaking events where previously homogeneous populations bifurcate into divergent

cell types, enabling formation of differentiated tissue domains, cell sorting, boundary

creation, and many of the myriad events found in development. While all developmental

processes are multifaceted, integrating gene expression patterns, signaling factors, and

physical changes to produce a final organ system, cell migration represents a key step

throughout development, enabling the formation of boundaries, the controlled mixture of

cell lineages, and the extension and elaboration of the body plan. Cell migration

processes are among the oldest described developmental events, but the rules

connecting individual and collective migration events to phenotypic outcomes remain

underexplored. In particular, the tightly regulated environment of the developing embryo

makes it difficult to determine how sensitive various multicellular patterning processes

are to perturbation of the percentages, lineages, and spatiotemporal distribution of

migratory cell types, and to what extent cell migration either ameliorates or exacerbates

stochastic variation in initial heterogeneous populations. Therefore, *in vitro* cell tracking

in models of embryogenesis offers a unique tool to explore the design space of

embryonic development, producing the necessary and sufficient conditions of pattern

formation.

In this dissertation, bioengineering, developmental biology, and computational biology are combined to produce a detailed description of cell migratory behavior in *ex vivo* models of gastrulation and axial extension. This work provides a foundational tool to interrogate patterning events during all phases of organism development, producing comprehensive descriptions of migratory cell behavior during several multicellular organization events, metrics to evaluate and compare pluripotency and differentiation protocols from live cell tracking, and an empirically derived computational model of self-assembly and convergent extension during gastruloid elongation. These tools can be employed towards deeper understanding of coupled morphogenic events *in vivo, in vitro,* and *in silico*, with the future goal of enabling rational *in vitro* organ generation for disease modeling, therapeutic generation, and tissue transplant.

1.8 Bibliography

Ahmed, M.U., Maurya, A.K., Cheng, L., Jorge, E.C., Schubert, F.R., Maire, P., Basson, M.A., Ingham, P.W., and Dietrich, S. (2017). Engrailed controls epaxial-hypaxial muscle innervation and the establishment of vertebrate three-dimensional mobility. Dev. Biol. *430*, 90–104.

Aman, A., and Piotrowski, T. (2010). Cell migration during morphogenesis. Dev. Biol. *341*, 20–33.

Amat, F., Höckendorf, B., Wan, Y., Lemon, W.C., McDole, K., and Keller, P.J. (2015). Efficient processing and analysis of large-scale light-sheet microscopy data. Nat. Protoc. *10*, 1679–1696.

Anderson, C., and Stern, C.D. (2016). Organizers in Development. In Current Topics in Developmental Biology, (Elsevier), pp. 435–454.

Anderson, R.H., Webb, S., Brown, N.A., Lamers, W., and Moorman, A. (2003). Development of the heart:(2) Septation of the atriums and ventricles. Heart *89*, 949–958.

Arciero, J.C., Mi, Q., Branca, M.F., Hackam, D.J., and Swigon, D. (2011). Continuum Model of Collective Cell Migration in Wound Healing and Colony Expansion. Biophys. J. *100*, 535–543.

Ascione, F., Vasaturo, A., Caserta, S., D'Esposito, V., Formisano, P., and Guido, S. (2016). Comparison between fibroblast wound healing and cell random migration assays in vitro. Exp. Cell Res. *347*, 123–132.

Azad, T., Ghahremani, M., and Yang, X. (2019). The Role of YAP and TAZ in Angiogenesis and Vascular Mimicry. Cells *8*, 407.

Bak, P., Chen, K., and Creutz, M. (1989). Self-organized criticality in the'Game of Life. Nature *342*, 780–782.

Bauer, A.L., Jackson, T.L., Jiang, Y., and Rohlf, T. (2010). Receptor cross-talk in angiogenesis: Mapping environmental cues to cell phenotype using a stochastic, Boolean signaling network model. J. Theor. Biol. *264*, 838–846.

Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.-C., Lutolf, M.P., Duboule, D., and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature.

Bedzhov, I., Graham, S.J.L., Leung, C.Y., and Zernicka-Goetz, M. (2014). Developmental plasticity, cell fate specification and morphogenesis in the early mouse embryo. Philos. Trans. R. Soc. B Biol. Sci. *369*, 20130538.

Bénazéraf, B., and Pourquié, O. (2013). Formation and Segmentation of the Vertebrate Body Axis. Annu. Rev. Cell Dev. Biol. *29*, 1–26.

Bérenger-Currias, N.M.L.P., Mircea, M., Adegeest, E., van den Berg, P.R., Feliksik, M., Hochane, M., Idema, T., Tans, S.J., and Semrau, S. (2020). Early neurulation recapitulated in assemblies of embryonic and extraembryonic cells (Developmental Biology).

Bertocchini, F., and Stern, C.D. (2002). The hypoblast of the chick embryo positions the primitive streak by antagonizing nodal signaling. Dev. Cell *3*, 735–744.

Blum, M., Schweickert, A., Vick, P., Wright, C.V.E., and Danilchik, M.V. (2014). Symmetry breakage in the vertebrate embryo: When does it happen and how does it work? Dev. Biol. *393*, 109–123.

van den Brink, S.C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.-K., Nowotschin, S., Turner, D.A., and Martinez Arias, A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. Development *141*, 4231–4242.

van den Brink, S.C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivié, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F., Martinez Arias, A., et al. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature.

Browning, A.P., McCue, S.W., Binny, R.N., Plank, M.J., Shah, E.T., and Simpson, M.J. (2018). Inferring parameters for a lattice-free model of cell migration and proliferation using experimental data. J. Theor. Biol. *437*, 251–260.

Buckingham, M.E., and Meilhac, S.M. (2011). Tracing Cells for Tracking Cell Lineage and Clonal Behavior. Dev. Cell *21*, 394–409.

Cai, D., Cohen, K.B., Luo, T., Lichtman, J.W., and Sanes, J.R. (2013). Improved tools for the Brainbow toolbox. Nat. Methods *10*, 540–547.

Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., et al. (2019). Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nat. Methods *16*, 1247–1253.

Campos, D., Méndez, V., and Llopis, I. (2010). Persistent random motion: Uncovering cell migration dynamics. J. Theor. Biol. *267*, 526–534.

Catala, M., Teillet, M.-A., and Le Douarin, N.M. (1995). Organization and development of the tail bud analyzed with the quail-chick chimaera system. Mech. Dev. *51*, 51–65.

Chazaud, C., Yamanaka, Y., Pawson, T., and Rossant, J. (2006). Early Lineage Segregation between Epiblast and Primitive Endoderm in Mouse Blastocysts through the Grb2-MAPK Pathway. Dev. Cell *10*, 615–624.

Chintalapati, M., and Moorjani, P. (2020). Evolution of the mutation rate across primates. Curr. Opin. Genet. Dev. *62*, 58–64.

Choi, H., Broitman-Maduro, G., and Maduro, M.F. (2017). Partially compromised specification causes stochastic effects on gut development in C. elegans. Dev. Biol. *427*, 49–60.

Chowdhury, A., Biswas, S., and Bianco, S. (2017). Active deep learning reduces annotation burden in automatic cell segmentation. BioRxiv 211060.

Christiaen, L., Stolfi, A., and Levine, M. (2010). BMP signaling coordinates gene expression and cell migration during precardiac mesoderm development. Dev. Biol. *340*, 179–187.

Chuai, M., and Weijer, C.J. (2009). Regulation of cell migration during chick gastrulation. Curr. Opin. Genet. Dev. *19*, 343–349.

Clevers, H. (2016). Modeling development and disease with organoids. Cell *165*, 1586–1597.

Coelho, L.P., Shariff, A., and Murphy, R.F. (2009). Nuclear segmentation in microscope cell images: a hand-segmented dataset and comparison of algorithms. In Biomedical

Imaging: From Nano to Macro, 2009. ISBI'09. IEEE International Symposium On, (IEEE), pp. 518–521.

Cohen, J.P., Lo, H.Z., and Bengio, Y. (2017). Count-ception: Counting by Fully Convolutional Redundant Counting. ArXiv Prepr. ArXiv170308710.

Combes, A.N., Zappia, L., Er, P.X., Oshlack, A., and Little, M.H. (2019). Single-cell analysis reveals congruence between kidney organoids and human fetal kidney. Genome Med. *11*.

Das, D., Chatti, V., Emonet, T., and Holley, S.A. (2017). Patterned Disordered Cell Motion Ensures Vertebral Column Symmetry. Dev. Cell *42*, 170-180.e5.

Debeir, O., Van Ham, P., Kiss, R., and Decaestecker, C. (2005). Tracking of migrating cells under phase-contrast video microscopy with combined mean-shift processes. IEEE Trans. Med. Imaging *24*, 697–711.

Deryugina, E.I., and Bourdon, M.A. (1996). Tenascin mediates human glioma cell migration and modulates cell migration on fibronectin. J. Cell Sci. *109*, 643–652.

Doyle, A.D., Wang, F.W., Matsumoto, K., and Yamada, K.M. (2009). One-dimensional topography underlies three-dimensional fibrillar cell migration. J. Cell Biol. *184*, 481–490.

Ezran, C., Karanewsky, C.J., Pendleton, J.L., Sholtz, A., Krasnow, M.R., Willick, J., Razafindrakoto, A., Zohdy, S., Albertelli, M.A., and Krasnow, M.A. (2017). The Mouse Lemur, a Genetic Model Organism for Primate Biology, Behavior, and Health. Genetics *206*, 651–664.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., et al. (2019). U-Net: deep learning for cell counting, detection, and morphometry. Nat. Methods *16*, 67–70.

Fulton, T., Trivedi, V., Attardi, A., Anlas, K., Dingare, C., Arias, A.M., and Steventon, B. (2020). Axis Specification in Zebrafish Is Robust to Cell Mixing and Reveals a Regulation of Pattern Formation by Morphogenesis. Curr. Biol. *30*, 2984-2994.e3.

George, R.M., Maldonado-Velez, G., and Firulli, A.B. (2020). The heart of the neural crest: cardiac neural crest cells in development and regeneration. Development *147*, dev188706.

Glen, C.M., Kemp, M.L., and Voit, E.O. (2019). Agent-based modeling of morphogenetic systems: Advantages and challenges. PLOS Comput. Biol. *15*, e1006577.

Gold, S., and Rangarajan, A. (1996). Softmax to softassign: Neural network algorithms for combinatorial optimization. J. Artif. Neural Netw. *2*, 381–399.

González-Valverde, I., and García-Aznar, J.M. (2018). Mechanical modeling of collective cell migration: An agent-based and continuum material approach. Comput. Methods Appl. Mech. Eng. *337*, 246–262.

González-Valverde, I., Semino, C., and García-Aznar, J.M. (2016). Phenomenological modelling and simulation of cell clusters in 3D cultures. Comput. Biol. Med. *77*, 249–260.

Graner, F., and Glazier, J.A. (1992). Simulation of biological cell sorting using a two-dimensional extended Potts model. Phys. Rev. Lett. *69*, 2013.

Hadjieconomou, D., Rotkopf, S., Alexandre, C., Bell, D.M., Dickson, B.J., and Salecker, I. (2011). Flybow: genetic multicolor cell labeling for neural circuit analysis in Drosophila melanogaster. Nat. Methods *8*, 260–266.

Hara, Y., Nagayama, K., Yamamoto, T.S., Matsumoto, T., Suzuki, M., and Ueno, N. (2013). Directional migration of leading-edge mesoderm generates physical forces: Implication in Xenopus notochord formation during gastrulation. Dev. Biol. *382*, 482–495.

Hashimoto, H., Robin, F.B., Sherrard, K.M., and Munro, E.M. (2015). Sequential Contraction and Exchange of Apical Junctions Drives Zippering and Neural Tube Closure in a Simple Chordate. Dev. Cell *32*, 241–255.

Hayashi, K. (2019). In vitro reconstitution of germ cell development†. Biol. Reprod. *101*, 567–578.

Hiramatsu, R., Matsuoka, T., Kimura-Yoshida, C., Han, S.-W., Mochida, K., Adachi, T., Takayama, S., and Matsuo, I. (2013). External Mechanical Cues Trigger the Establishment of the Anterior-Posterior Axis in Early Mouse Embryos. Dev. Cell *27*, 131–144.

Hohwieler, M., Illing, A., Hermann, P.C., Mayer, T., Stockmann, M., Perkhofer, L., Eiseler, T., Antony, J.S., Müller, M., Renz, S., et al. (2017). Human pluripotent stem cell-derived acinar/ductal organoids generate human pancreas upon orthotopic transplantation and allow disease modelling. Gut *66*, 473–486.

Kang, M., Xenopoulos, P., Muñoz-Descalzo, S., Lou, X., and Hadjantonakis, A.-K. (2013). Live Imaging, Identifying, and Tracking Single Cells in Complex Populations In

Vivo and Ex Vivo. In Imaging and Tracking Stem Cells, K. Turksen, ed. (Totowa, NJ: Humana Press), pp. 109–123.

Keller, P.J., Schmidt, A.D., Wittbrodt, J., and Stelzer, E.H. (2008). Reconstruction of zebrafish early embryonic development by scanned light sheet microscopy. Science *322*, 1065–1069.

Kelly, E.M., and Sears, K.E. (2011). Limb specialization in living marsupial and eutherian mammals: constraints on mammalian limb evolution. J. Mammal. *92*, 1038–1049.

Kheradmand, S., Singh, A., Saeedi, P., Au, J., and Havelock, J. (2017). Inner cell mass segmentation in human HMC embryo images using fully convolutional network. In Image Processing (ICIP), 2017 IEEE International Conference On, (IEEE), pp. 1752–1756.

Koenig, A.L., Baltrunaite, K., Bower, N.I., Rossi, A., Stainier, D.Y.R., Hogan, B.M., and Sumanas, S. (2016). Vegfa signaling promotes zebrafish intestinal vasculature development through endothelial cell migration from the posterior cardinal vein. Dev. Biol. *411*, 115–127.

Krupa, M., Mazur, E., Szczepańska, K., Filimonow, K., Maleszewski, M., and Suwińska, A. (2014). Allocation of inner cells to epiblast vs primitive endoderm in the mouse embryo is biased but not determined by the round of asymmetric divisions (8→16- and 16→32-cells). Dev. Biol. *385*, 136–148.

Libby, A.R., Joy, D.A., So, P.-L., Mandegar, M.A., Muncie, J.M., Mendoza-Camacho, F.N., Weaver, V.M., Conklin, B.R., and McDevitt, T.C. (2018). Spatiotemporal mosaic self-patterning of pluripotent stem cells using CRISPR interference. ELife *7*.

Libby, A.R.G., Briers, D., Haghighi, I., Joy, D.A., Conklin, B.R., Belta, C., and McDevitt, T.C. (2019). Automated Design of Pluripotent Stem Cell Self-Organization. Cell Syst. *9*, 483-495.e10.

Lou, Y.-R., and Leung, A.W. (2018). Next generation organoids for biomedical research and applications. Biotechnol. Adv. *36*, 132–149.

Lou, X., Kang, M., Xenopoulos, P., Muñoz-Descalzo, S., and Hadjantonakis, A.-K. (2014). A Rapid and Efficient 2D/3D Nuclear Segmentation Method for Analysis of Early Mouse Embryo and Stem Cell Image Data. Stem Cell Rep. *2*, 382–397.

Luo, Z.-X. (2007). Transformation and diversification in early mammal evolution. Nature *450*, 1011–1019.

Madhikar, P., Åström, J., Westerholm, J., and Karttunen, M. (2018). CellSim3D: GPU accelerated software for simulations of cellular growth and division in three dimensions. Comput. Phys. Commun. *232*, 206–213.

Maiuri, P., Terriac, E., Paul-Gilloteaux, P., Vignaud, T., McNally, K., Onuffer, J., Thorn, K., Nguyen, P.A., Georgoulia, N., Soong, D., et al. (2012). The first World Cell Race. Curr. Biol. *22*, R673–R675.

Marcon, L., and Sharpe, J. (2012). Turing patterns in development: what about the horse part? Curr. Opin. Genet. Dev. *22*, 578–584.

Marikawa, Y., Chen, H.-R., Menor, M., Deng, Y., and Alarcon, V.B. (2020). Exposure-based assessment of chemical teratogenicity using morphogenetic aggregates of human embryonic stem cells. Reprod. Toxicol. *91*, 74–91.

McDole, K., Guignard, L., Amat, F., Berger, A., Malandain, G., Royer, L.A., Turaga, S.C., Branson, K., and Keller, P.J. (2018). In Toto Imaging and Reconstruction of Post-Implantation Mouse Development at the Single-Cell Level. Cell *175*, 859-876.e33.

Mehdipour, N., Briers, D., Haghighi, I., Glen, C.M., Kemp, M.L., and Belta, C. (2018). Spatial-Temporal pattern Synthesis in a Network of Locally Interacting Cells. In 2018 IEEE Conference on Decision and Control (CDC), (Miami Beach, FL: IEEE), pp. 3516–3521.

Meijering, E., Dzyubachyk, O., Smal, I., and van Cappellen, W.A. (2009). Tracking in cell and developmental biology. Semin. Cell Dev. Biol. *20*, 894–902.

Meijering, E., Dzyubachyk, O., Smal, I., and others (2012). Methods for cell and particle tracking. Methods Enzym. *504*, 183–200.

Milinkovitch, M.C., and Tzika, A. (2007). Escaping the mouse trap: the selection of new Evo-Devo model species. J. Exp. Zoolog. B Mol. Dev. Evol. *308B*, 337–346.

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. Nat. Methods *16*, 1233–1246.

Morgani, S.M., and Hadjantonakis, A.-K. (2020). Signaling regulation during gastrulation: Insights from mouse embryos and in vitro systems. In Current Topics in Developmental Biology, (Elsevier), pp. 391–431.

Nieuwkoop, P.D. (1985). Inductive interactions in early amphibian development and their general nature. Development *89*, 333–347.

Norfleet, D.A., Park, E., and Kemp, M.L. (2020). Computational modeling of organoid development. Curr. Opin. Biomed. Eng. *13*, 113–118.

Pan, Y.A., Livet, J., Sanes, J.R., Lichtman, J.W., and Schier, A.F. (2011). Multicolor Brainbow imaging in zebrafish. Cold Spring Harb. Protoc. *2011*, pdb–prot5546.

Pantazis, P., and Supatto, W. (2014). Advances in whole-embryo imaging: a quantitative transition is underway. Nat. Rev. Mol. Cell Biol. *15*, 327–339.

Parent, C.A. (1999). A Cell's Sense of Direction. Science *284*, 765–770.

Patlak, C.S. (1953). Random walk with persistence and external bias. Bull. Math. Biophys. *15*, 311–338.

Peng, X., Bennamoun, M., Ma, Q., Lei, Y., Zhang, Q., and Chen, W. (2010). Drift-correcting template update strategy for precision feature point tracking. Image Vis. Comput. *28*, 1280–1292.

Pfeffer, P.L. (2018). Building Principles for Constructing a Mammalian Blastocyst Embryo. Biology *7*, 41.

Pilhwa Lee, and Charles W. Wolgemuth (2011). Crawling Cells Can Close Wounds without Purse Strings or Signaling. PLOS Comput. Biol.

Piltti, K.M., Cummings, B.J., Carta, K., Manughian-Peter, A., Worne, C.L., Singh, K., Ong, D., Maksymyuk, Y., Khine, M., and Anderson, A.J. (2018). Live-cell time-lapse imaging and single-cell tracking of in vitro cultured neural stem cells – Tools for analyzing dynamics of cell cycle, migration, and lineage selection. Methods *133*, 81–90.

Plank, M.J., and Simpson, M.J. (2013). Lattice-Free Models of Cell Invasion: Discrete Simulations and Travelling Waves. Bull. Math. Biol. *75*, 2150–2166.

Płusa, B., and Piliszek, A. (2020). Common principles of early mammalian embryo self-organisation. Development *147*, dev183079.

Rolian, C. (2014). Genes, development, and evolvability in primate evolution: Genes, Development, and Evolvability in Primate Evolution. Evol. Anthropol. Issues News Rev. *23*, 93–104.

Schiffmann, Y. (2006). Symmetry breaking and convergent extension in early chordate development. Prog. Biophys. Mol. Biol. *92*, 209–231.

Selwood, L., and Johnson, M.H. (2006). Trophoblast and hypoblast in the monotreme, marsupial and eutherian mammal: evolution and origins. BioEssays *28*, 128–145.

Shah, G., Thierbach, K., Schmid, B., Waschke, J., Reade, A., Hlawitschka, M., Roeder, I., Scherf, N., and Huisken, J. (2019). Multi-scale imaging and analysis identify pan-embryo cell dynamics of germlayer formation in zebrafish. Nat. Commun. *10*.

Silva, A.C., Matthys, O.B., Joy, D.A., Kauss, M.A., Natarajan, V., Lai, M.H., Turaga, D., Alexanian, M., Bruneau, B.G., and McDevitt, T.C. (2020). Developmental co-emergence of cardiac and gut tissues modeled by human iPSC-derived organoids (Developmental Biology).

Stegmaier, J., Amat, F., Lemon, W.C., McDole, K., Wan, Y., Teodoro, G., Mikut, R., and Keller, P.J. (2016). Real-Time Three-Dimensional Cell Segmentation in Large-Scale Microscopy Data of Developing Embryos. Dev. Cell *36*, 225–240.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Sutherland, A.E. (2016). Tissue morphodynamics shaping the early mouse embryo. In Seminars in Cell & Developmental Biology, (Elsevier), pp. 89–98.

Takaoka, K., Nishimura, H., and Hamada, H. (2017). Both Nodal signalling and stochasticity select for prospective distal visceral endoderm in mouse embryos. Nat. Commun. *8*.

Takebe, T., and Wells, J.M. (2019). Organoids by design. Science *364*, 956–959.

Tam, P.P.L., and Behringer, R.R. (1997). Mouse gastrulation: the formation of a mammalian body plan. Mech. Dev. *68*, 3–25.

Tang, Y.-H., and Karniadakis, G.E. (2014). Accelerating dissipative particle dynamics simulations on GPUs: Algorithms, numerics and applications. Comput. Phys. Commun. *185*, 2809–2822.

Tapia, J.J., and D'Souza, R.M. (2011). Parallelizing the Cellular Potts Model on graphics processing units. Comput. Phys. Commun. *182*, 857–865.

Tewary, M., Ostblom, J., Prochazka, L., Zulueta-Coarasa, T., Shakiba, N., Fernandez-Gonzalez, R., and Zandstra, P.W. (2017). A stepwise model of reaction-diffusion and positional information governs self-organized human peri-gastrulation-like patterning. Development *144*, 4298–4312.

Tewary, M., Dziedzicka, D., Ostblom, J., Prochazka, L., Shakiba, N., Woodford, C., Piccinini, E., Vickers, A., Louis, B., Rahman, N., et al. (2018). High-throughput micro-

patterning platform reveals Nodal-dependent dissection of peri-gastrulation-associated versus pre-neurulation associated fate patterning.

Turing, A.M. (1952). The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. *237*, 37–72.

Uesaka, T., Young, H.M., Pachnis, V., and Enomoto, H. (2016). Development of the intrinsic and extrinsic innervation of the gut. Dev. Biol. *417*, 158–167.

Veenvliet, J.V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Pustet, M., Heimann, S., Buschow, R., et al. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. BioRxiv.

Voss-Böhme, A. (2012). Multi-Scale Modeling in Morphogenesis: A Critical Analysis of the Cellular Potts Model. PLoS ONE *7*, e42852.

Wang, Q., Holmes, W.R., Sosnik, J., Schilling, T., and Nie, Q. (2017). Cell Sorting and Noise-Induced Cell Plasticity Coordinate to Sharpen Boundaries between Gene Expression Domains. PLoS Comput. Biol. *13*, e1005307.

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

Wieschaus, E. (2016). Positional Information and Cell Fate Determination in the Early Drosophila Embryo. In Current Topics in Developmental Biology, (Elsevier), pp. 567–579.

Winklbauer, R., and Parent, S.E. (2017). Forces driving cell sorting in the amphibian embryo. Mech. Dev. *144*, 81–91.

Wu, G., and Schöler, H.R. (2016). Lineage Segregation in the Totipotent Embryo. In Current Topics in Developmental Biology, (Elsevier), pp. 301–317.

Xie, W., Noble, J.A., and Zisserman, A. (2016). Microscopy cell counting and detection with fully convolutional regression networks. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 1–10.

Xie, Y., Xing, F., Shi, X., Kong, X., Su, H., and Yang, L. (2018). Efficient and robust cell detection: A structured regression approach. Med. Image Anal. *44*, 245–254.

Yachie-Kinoshita, A., Onishi, K., Ostblom, J., Langley, M.A., Posfai, E., Rossant, J., and Zandstra, P.W. (2018). Modeling signaling-dependent pluripotency with Boolean logic to predict cell fate transitions. Mol. Syst. Biol. *14*, e7952.

Yang, C.N. (1952). The Spontaneous Magnetization of a Two-Dimensional Ising Model. Phys. Rev. *85*, 808–816.

Yarrow, J.C., Perlman, Z.E., Westwood, N.J., and Mitchison, T.J. (2004). A high-throughput cell migration assay using scratch wound healing, a comparison of image-based readout methods. BMC Biotechnol. *4*, 21.

Zhang, X., Xing, F., Su, H., Yang, L., and Zhang, S. (2015). High-throughput histopathological image analysis via robust cell segmentation and hashing. Med. Image Anal. *26*, 306–315.

2 Sparse cell tracking reveals hiPSC behavioral signatures during symmetry breaking

2.1 Introduction

During early embryogenesis, cells must undergo a series of patterned differentiation and cell migration events to form and organize the three germ layers that compose the main body tissues (Pelegri et al, 2017). Collective cell migration is readily apparent in many core body formation events such as gastrulation (Chuai and Weijer, 2009), neural tube formation (Hashimoto et al., 2015), and the cell flows that contribute to the development of the heart (George et al., 2020), but individual cell migration events also contribute to, and enable robust formation of several different tissue types (Aman and Piotrowski, 2010; Saykali et al., 2018). While cell migration during differentiation is well described in vivo, the complex, rapidly evolving, 3D nature of developing embryos makes untangling individual cell migrations from bulk tissue motion difficult (Pantazis and Supatto, 2014). Analogous to developing tissues, human induced pluripotent stem cells (hiPSCs) and human embryonic stem cells (hESCs) have the ability to self-organize (Libby et al., 2018), differentiate to all three germ layers (Han et al., 2018), and undergo transitions from bulk endothelial cell sheets to freely migrating mesenchymal cells (Yang et al., 2020). hiPSCs provide a robust, tractable system to observe, quantify, and predict cell migration phenotypes, enabling mechanistic modeling of cell migration in pluripotency, and quantification of behavioral change during differentiation. The ability to deconvolve the interactions of heterotypic cell types with

morphological transitions during cell fate specification enables the possibility of predicting and directing differentiation via cell-intrinsic routes.

In this study, partially labeled colonies of hiPSCs with inducible constructs based on CRISPR interference (CRISPRi)(Libby et al., 2018) or optogenetically activated Wnt signaling (optoWnt)(Bugaj et al., 2013) were used to induce and subsequently track cell sorting and migration events both during pluripotency and early differentiation to multiple germ lineages. By providing the empirically derived cell behavior parameters to optimize an *in silico* model of multicellular pattern formation, this tracking pipeline enabled closed loop prediction of pattern formation and subsequent organizational control of germ lineages within a colony. Further, by tracking and analyzing cell behavior during optoWnt directed differentiation, mesodermal cell fate acquisition was detected non-destructively and migratory signatures of the epithelial to mesenchymal transition (EMT) were quantified. Through tracking of live cell migration, this pipeline generated measures of *in vitro* morphogenesis that are predictive of organization events both in pluripotency and in differentiation.

2.2 Materials and Methods

2.2.1 Cell Culture

hiPSC CRISPRi cell lines were derived from the parent line WTC (Coriell Cat. # GM25256). Optogenetic lines were derived from hESC lines (H9, WiCell) and hiPSC lines (19-9-7, WiCell). All cells were grown at 37° C and 5% CO_2 in feeder-free media conditions on Matrigel (Corning, lot # 7268012, 7275006; BD Biosciences) in mTeSR1 medium (STEMCELL Technologies)(Ludwig et al., 2006) with daily media changes.

Optogenetic cells were cultured with the hood lights off. Cells were singularized in Accutase (STEMCELL Technologies) at 37° C for 5 minutes and seeded onto Matrigel-coated plates in media containing Rho-associated coiled-coil kinase (ROCK) inhibitor Y-276932 (5-10µM depending on cell line; Selleckchem)(Watanabe et al., 2007). WTC derived cell lines were seeded at 12,000 cells per cm^2 while optoWnt cell lines were seeded at 35,000 cells cm^2.

2.2.2 Generation of cell line

CRISPRi knockdown lines were previously generated as described in (Mandegar et al., 2016), where 20 base pair guides were designed using the Broad Institute sgRNA design website (Doench et al., 2016). 20 base pair sequences were cloned into the gRNA-CNKB vector using restriction enzyme BsmBI digestions, followed by ligation with T4 DNA ligase. 200,000 cells of the CRISPRi-Gen1 hiPSC line were nucleofected with individual gRNA vectors using the Human Stem Cell Nucleofector Kit 1 solution with the Amaxa nucleofector 2b device (Lonza). Cells were then plated at increasing dilutions into 3 wells of a 6-well plate coated with Matrigel (BD Biosciences) in mTeSR1 (STEMCELL Technologies) supplemented with Y-276932 (10µM) for 2 days, then treated with blasticidin (10µg/ml) for a selection period of 7 days. Surviving colonies for each gRNA were pooled and passaged in mTeSR1 with blasticidin (10µg/ml) and Y-27632 (10µM) for a single day then transitioned to mTeSR1 media only. After stable polyclonal populations of hiPSCs were established for each gRNA, cells were karyotyped by Cell Line Genetics (Libby et al., 2018). Knockdown efficiency was tested by the addition of doxycycline (2µM) to the culture media for 6 days and subsequent

qPCR of mRNA levels of respective genes compared to time matched controls of the same line without CRISPRi induction.

Clonal knock-in Opt-Wnt cell lines were generated through CRISPR/Cas9-mediated recombination. Prior to nucleofection, hESCs were pre-treated with 10 μM ROCK inhibitor for 3 to 4 hours or 5 μM Y27632 overnight. Accutase-digested single hESCs were collected and 2.5 - 3.5 million cells were nucleofected with 2.5 μg gRNA AAVS1-T2 (Addgene # 41818), 4.5 μg pCas9-GFP (Addgene #44719), and 6 μg optoWnt donor plasmid in 200 μl room temperature PBS -/- using a NucleofectorTM 2b (Lonza) with program B-016. The resulting cells were plated onto Matrigel-coated 6-well plates containing 3 mL pre-warmed mTeSR1 with 10 μM ROCK inhibitor. Once the cells grew to confluency, they were subjected to selection with 1 μg mL-1 puromycin in mTeSR1 media for approximately 2 weeks. Clonal lines were generated by picking single-cell clones into wells of a Matrigel-coated 96-well plate that were expanded for 1-2 weeks and subjected to PCR genotyping.

2.2.3 Formation of heterotypic cell populations

Mixed population hiPSC colonies were generated using forced aggregation via PDMS microwells in a 24-well tissue culture plate (~975 400x400μm wells per well) (Hookway et al., 2016; Libby et al., 2018). hiPSCs were dissociated and singularized using Accutase (STEMCELL Technologies) and subsequently counted using an Invitrogen Countess Automated Cell Counter (Thermofisher Scientific). The proper ratios of cells to create 100 cell aggregates were then seeded into PDMS wells in mTeSR1 with Y-27632 (10μM), centrifuged at 200g for 5 minutes, and allowed to

compact overnight (~18h). Aggregates were then washed out of the PDMS wells with fresh mTeSR1 and re-plated into a growth factor reduced matrigel (BD Biosciences) coated PerkinElmer CellCarrier™-96 plates at ~10/aggregates/cm². optoWnt colonies were created by mixing WT and optoWnt cells at a 1:1 ratio and seeded onto Matrigel coated plates (Corning) at a cell density of 35,000 cells per cm². Cells in all conditions were and fed daily with mTeSR1.

2.2.4 Heterotypic colony imaging

Mixed hiPSC colonies were imaged at the basal surface on optically clear PerkinElmer CellCarrier™-96 plates on an inverted AxioObserver Z1 (Ziess) with an ORCA-Flash 4.0 digital CMOS camera (Hamamatsu) with a 20x objective, where that single plane was used for parameter estimations. Mixed colonies of wildtype and CRISPRi-Gen1 cells without knockdown guides were imaged for 6 hours every 5 minutes from hours 60-66 after plate down. Mixed colonies of wildtype, CDH1 KD, and ROCK1 KD cells lines were also generated at ratios predicted by a machine learning pipeline as described in (Libby et al., 2019). Colonies were allowed to grow for between 36 and 108 hours after plate down, and then fixed in 4% paraformaldehyde (VWR) for 25 minutes. Samples were then incubated with primary antibodies overnight at 4°C in a PBS solution with 1% bovine serum albumin (Sigma Aldrich) and 0.3% Triton-X. Samples were washed 3 times and then incubated for 1 hours at room temperature with secondary antibodies and Hoescht. Primary antibodies used were: anti-OCT4 (SantaCruz 1:400), anti-SOX2 (AbCAM 1:400), and anti-Ecadherin (AbCAM 1:200). All secondary antibodies were used at 1:1,000 and purchased from Life Technologies.

Images were taken in one focal plane on the apical surface of hiPSC colonies at 10x

magnification on a Ziess Observer.Z1 (Ziess) or an InCell Analyzer2000 (GE

Healthcare). All image analysis was done using custom code written in python using the

sckit-image package (van der Walt et al., 2014).

2.2.5 optoWnt colony imaging

Co-cultures were treated with CellTracker Red (ThermoFisher) dye diluted

5,000x in mTeSR1 media for 15 min and washed two times. A sealing membrane

(Breathe-Easy, Sigma-Aldrich) was applied to plates prior to imaging. Plates were

imaged on a Molecular Devices Image Xpress Micro (IXM) imaging system with

environmental control (37 °C, 5% CO2, and humidity control) using a 10x objective. For

single-cell tracking experiments, 4-8 sites were imaged at 18 min intervals. Optogenetic

stimulation was delivered from the fluorescence light source (SOLA Light Engine,

Lumencor) set to 5% intensity, passing through the 10x objective and GFP filter set

(472/30nm). Measured power at the sample was 2.82 mW. Optogenetic stimulation was

delivered for 3 min at each site prior to imaging of each timepoint (i.e. for 3 min every 18

min) in a sequence of short light pulses (500 ms on-pulse, 10 s off-pulse).

2.2.6 Colony Boundary Tracking

To detect the boundary of colonies imaged in phase, difference images were

calculated between subsequent frames of a 6-hour phase time series, and then the

standard deviation of pixel values calculated. A two component gaussian mixture model

was fit to the resulting anomaly image to find an optimal threshold separating non-

moving from moving pixels. Regions with motion above this threshold were segmented

and then filtered to remove both objects and holes smaller than 52 μm². The largest

contiguous region was then selected as the seed for the mask at each frame. For each

frame, the background was estimated using phase values outside of the seed mask as

definitive background and filling the resulting gap using biharmonic inpainting (Damelin

and Hoang, 2018). The contour of the resulting mask was then used to calculate

properties such as colony area, mean and max radius, and perimeter, as well as to

derive colony circularity with the formula: $Circularity = 4\pi * Area/Perimeter^2$. All

image manipulations were performed using the python package scikit-image (van der

Walt et al., 2014)

2.2.7 Contour Cell Tracking

Mixed aggregates of defined ratios of WT and CRISPRi cells without a targeting

guide were generated. With the addition of doxycycline (DOX) to the cell culture media,

the CRISPRi no guide population expressed a cytoplasmic mCherry marker which

allowed individual cells to be distinguished from the untagged WT background (Figure

2.2). 24 colonies were imaged for 6 hours at 5 minutes/image at 20X magnification

creating a time series of 73 frames. Each frame was individually normalized and

thresholded using adaptive histogram normalization. Cell migration tracks were

generated by following matching contours between frames where matching contours

share at least ten pixels overlap. Watershed segmentation was used to separate

adjacent cells. Instantaneous frame to frame velocity was calculated as:

$$v_{inst} = [(x_{cm,2} - x_{cm,1})/\Delta t, (y_{cm,2} - y_{cm,1})/\Delta t]$$

where $x_{cm,2}$ was the center of mass of each segmented cell body at the currently observed frame and $x_{cm,1}$ was the center of mass of each segmented cell body at the previous frame, and Δt was the sample rate in minutes.

2.2.8 Center of Mass Cell Tracking

To extract cell positions, each the contrast each image was corrected using adaptive histogram equalization, then the image background was removed using a difference of gaussians filter. Peaks at least 20% above background in the foreground image were then detected using non-local maximum suppression with a minimum radius of 2.6 μm. The cell counts were then fit to an exponential model of cell growth, and frames with anomalous ($R^2 > 2500$ cells^2) segmentations were discarded. This resulted in dataset of individual cell detections in 5 fields of view spanning 90 frames between approximately hours 8 and 36 of the stimulation experiment. Individual cell detections were linked to their nearest neighbor in a radius up to 32.4 μm from their previous position both forwards and backwards in time with both a global maximum velocity and neighborhood quasi-rigidity penalty(Kim and Lee, 2002). The resulting track fragments were then iteratively merged with overlapping tracks within 3.4 μm and 18 minutes of each other. This process converged after 10 iterations, generating 4,127 total tracks with mean length of 13.5 hours (standard deviation 6.4 hours).

For each track, instantaneous velocity magnitude and direction were approximated using finite differences and then smoothed with a 15 point (4.5 hour) rolling window filter. Cell migration distance was calculated by integrating over finite differences of cell position. Track turning angle was calculated by phase unwrapping

change in velocity direction. Finally, periods of persistent migration were determined as those times where a cell was both migrating at least 0.1 µm/min and turning no more than 2 degrees/min from its previous velocity direction. Both changes in instantaneous traces and binned values were assessed based on 95% confidence intervals around the mean, as determined by 1,000 iterations of bootstrap sampling.

2.3 Results

2.3.1 Uniformly mixed colonies generated from force aggregation and reattachment

To establish a system to generate sparsely labeled colonies of defined size and composition, WT cells expressing a DOX-inducible cytoplasmic mCherry construct were mixed at defined ratios with WT cells from the same genetic background without a fluorescent construct (Mandegar et al., 2016). Cells were mixed at ratios of 1:99, 1:19, 1:9, 1:3, 1:1, and 3:1 mCherry:WT cells and then formed into 100 cell aggregates using force aggregation (Hookway et al., 2016). After 18-24 hours of formation, they were removed from uWells and allowed to reattach to 96-well plates. Aggregates attached after approximately another 24 hours, after which they were imaged in both phase and 647 nm for 6 hours, sampling every 5 minutes (**Figure 2.1A**).
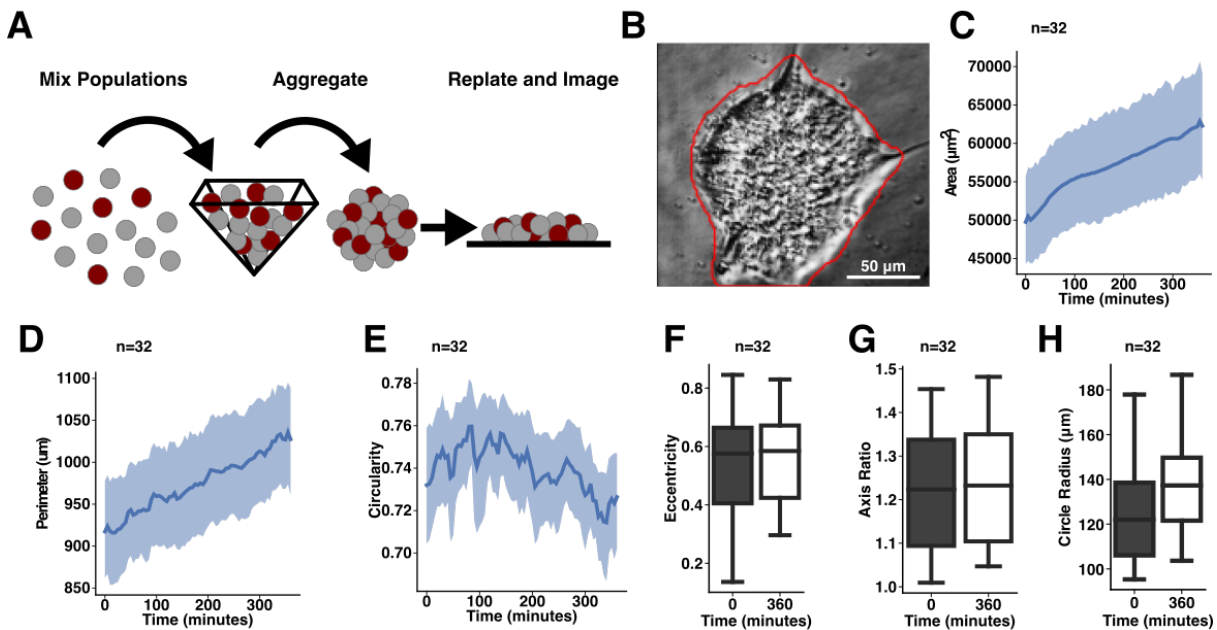
Figure 2.1: Kinetics of re-adherent hiPSC colony growth
A) Schematic of colony and cell tracking workflow B) Colony phase image with segmentation (red). C) Colony area, D) perimeter, and E) circularity changes over 6 hours. F) Colony eccentricity, G) ratio of long to short axis, and H) mean radii

Colony growth was assessed by segmenting the colony periphery in phase

(**Figure 2.1B**) and then calculating summary statistics such as area, perimeter, and

circularity. Colonies grew approximately 30% in area over 6 hours, expanding from a

mean radius of 124 µm (+/- 21 µm) to 138 µm (+/- 23 µm), while maintaining a

circularity of approximately 0.73 (1.0 is perfectly circular). Mean eccentricity of individual

colonies was stable throughout observation around 0.55 (+/- 0.16), suggesting that the

long axis of the colony was approximately 20% longer than the short axis in each

colony, with the variation in axis ratio converging over time (**Figure 2.1 C-H**). These

results demonstrate that hiPSC colonies generated from force aggregation and re-

adhesion do not diverge significantly from a circular cross section and potentially adopt

a more uniform shape over time.
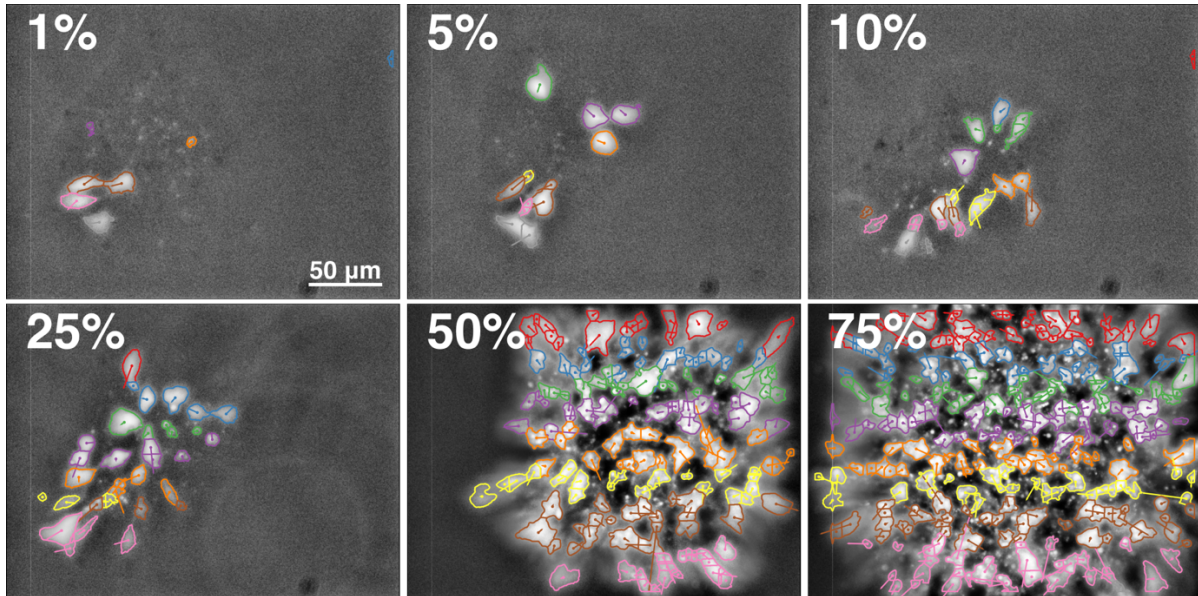
Figure 2.2: Segmentation of mixed colonies at defined ratios of mCherry:WT cells
Segmentation boundaries outlined and apparent direction of motion shown by arrows.

Cell position was assessed within a colony defined by the apparent boundary of the colony segmented in phase (**Figure 2.2**). At mixing ratios above 25%, cell boundaries were not readily apparent between mCherry cells, while colonies mixed at 1:99 cells had low incorporation of mCherry+ cells, with ~50% colonies having fewer than 2 labeled cells (n=8/16). Colonies at between 5% and 10% cell mixing were universally labeled, and had good separation between individual labeled cells, but also more than 5 cells detected per colony per frame, enabling efficient sampling. Some colonies at 25% labeling were also trackable at early stages, but clonal expansion lead to clustering of labeled cells over frames after a few hours. Based on these results, a mixing ratio of 25% or fewer labeled cells to WT cells was chosen for further analysis.

To understand how heterotopically mixed cells distribute within a given colony, average parameters for each cell, and measures of mean cell position within a colony were calculated. Individual cells occupied a median area of 500 µm^2 (IQR: 330 to 782

µm^2), and were also approximately elliptical, with median circularity 0.76 (IQR: 0.65 to 0.84; 1.0 is perfectly circular), eccentricity 0.78 (IQR: 0.69 to 0.85; 0.0 is perfectly circular) and radius ratio of 1.61 (IQR: 1.38 to 1.92; 1.0 is circular) (**Figure 2.3 A-D**). Radial distribution of the cells within a given colony was calculated by taking the distance of the center of mass of each cell from the overall colony center of mass, then dividing by the colony radius, giving a normalized distance from center. Cell distribution was not significantly different from randomly distributed cells along the unit circle, suggesting that WT cells are well mixed in an hiPSC colony, despite clonal expansion (**Figure 2.3 E**). A theoretical calculation of the number of cells per colony was made by multiplying the number of cell detections by the original mixing ratio, while a second calculation was made by assuming cells obeyed perfect random hexagonal packing within the colony segmentation (Figure 2.3.F), with the hex packing number given by:

$$n_{cells} = \frac{A_{colony}}{A_{cell}} * \frac{\pi}{\sqrt{12}} = \frac{r_{colony}^2}{r_{cell}^2} * \frac{\pi}{\sqrt{12}}$$

The two values agreed within 18% (+/- 11%) on average, suggesting that heterotypic aggregation robustly produces the anticipated mix of cell populations, and that those population ratios are stable over at least 2 days of cell culture (**Figure 2.3 G**).
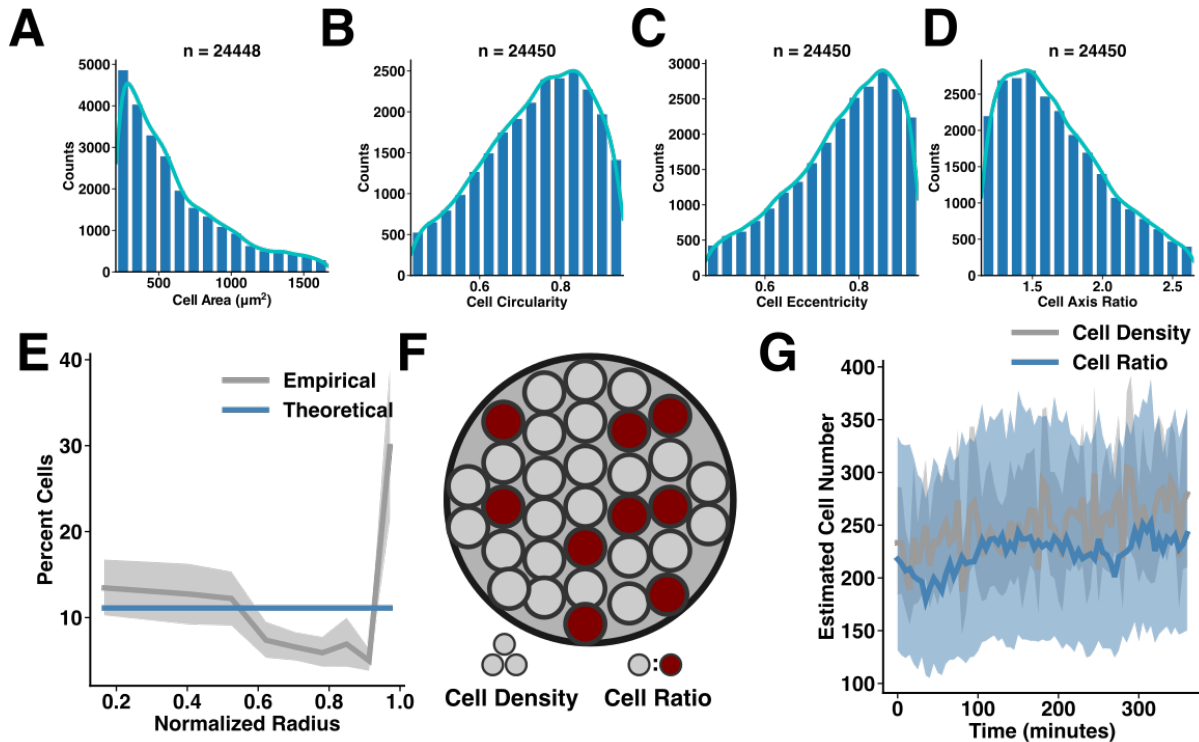
Figure 2.3 Shape and distribution of hiPSC cells in colonies
Distributions of A) cell area, B) cell circularity, C) eccentricity, and D) ratio of major to minor axis lengths.
E) Empirical and theoretical distributions of cells over the colony radius. F) Schematic of two models of
calculating total cell number, with G) estimated cell number by density (grey) and cell ratio (blue).

2.3.2 Measurement of cell migration parameters

Dynamic cell behavior in colonies was next assessed by linking individual cell

detections into cell tracks over time. Cell contours were observed to overlap from frame

to frame, so individual cell masks were linked into tracks by calculating the intersection

between all detected cells in a given frame with all the cells in the subsequent frame.

The contour with the majority overlap between two frames was assigned to be the same

cell. In cases with multiple overlapping cells, cells were assigned greedily by overlap

area, where the largest overlap was assigned first, then the second largest, and so on

until no more unlinked cells remained. Despite not accounting for cell division, shape

parameters, or frame-wise errors in cell segmentation, this method robustly produced

cell tracks spanning multiple hours that robustly followed mCherry expressing cells in

10% labeled colonies for 6 hours (**Figure 2.4 A,B**).
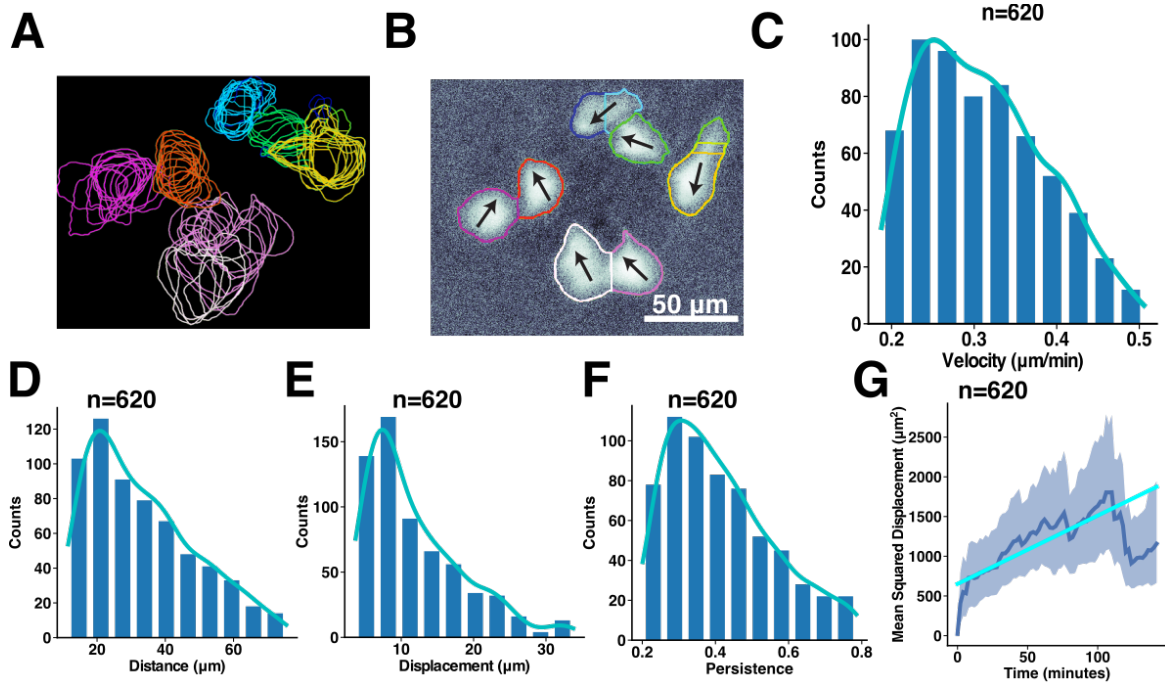
Figure 2.4 Cell behavior of hiPSCs in colonies
A) Linked cell contours colored by track identity over 6 hours. B) Single frame showing cell contour
and velocity estimate. Distributions of C) Average cell migration velocity magnitude, D) total distance
traveled, E) total distance from cell origin, and F) persistence of migration. G) Empirical (blue) and
model fit (cyan) of mean square displacement.

To calculate summary statistics, all tracks spanning at least 15 minutes of time (3

samples) were used to estimate instantaneous cell velocity, total cell distance traveled,

and total cell displacement. Cells were found to migrate with an average velocity of 0.31

µm/min (+/- 0.07 µm/min) which followed an over dispersed gamma distribution (**Figure**

2.4 C). Average track displacement, distance traveled, and persistence also followed a

heavy-right tailed distribution, with cells traveling only 36.4 µm (+/- 15.7 µm) with a

directional persistence (displacement/distance) of 0.42 (+/- 0.14; where 1.0 is travel in a

straight line) (**Figure 2.4 D-F**). Mean squared displacement was linear over a range of

approximately 2.5 hours of tracking, with a slope of 651 μm^2/min translating to a

diffusion coefficient for hiPSC stem cells of ~163 μm^2/min. These results suggest that

human stem cells are among the faster types of migrating cells, with speeds

comparable to human and murine tumor cell lines (Maiuri et al., 2012), and undergo

primarily diffusive migration in 2D colonies, producing large scale mixing that prevents

clonal expansion alone from producing heterotypic cell patterning.

2.3.3   Spatial quantification in mosaic patterned colonies using nuclei detection

Although contour-based tracking of sparsely labeled cells was effective for

analyzing hiPSC behavior within colonies, the low level of cell detection and potential

biases of clonal expansion forced selection of a different approach for characterizing

mosaic patterning within colonies. Since hiPSC and hESCs have prominent, elliptical

nuclei that compose a large fraction of the cytoplasm (Lou et al., 2014), nuclei detection

provided an attractive target to quantify cell position and expression of transcription

factors within a fixed colony. To detect cells without evaluating a contour, peak

detection was used to find the highest regional staining for Hoescht in a local area of

approximately 5x5 μm using non-maximum suppression. By visual inspection, these

peaks corresponded roughly to the center of mass of each Hoescht signal, and hence

the approximate centroid of each cell nucleus.
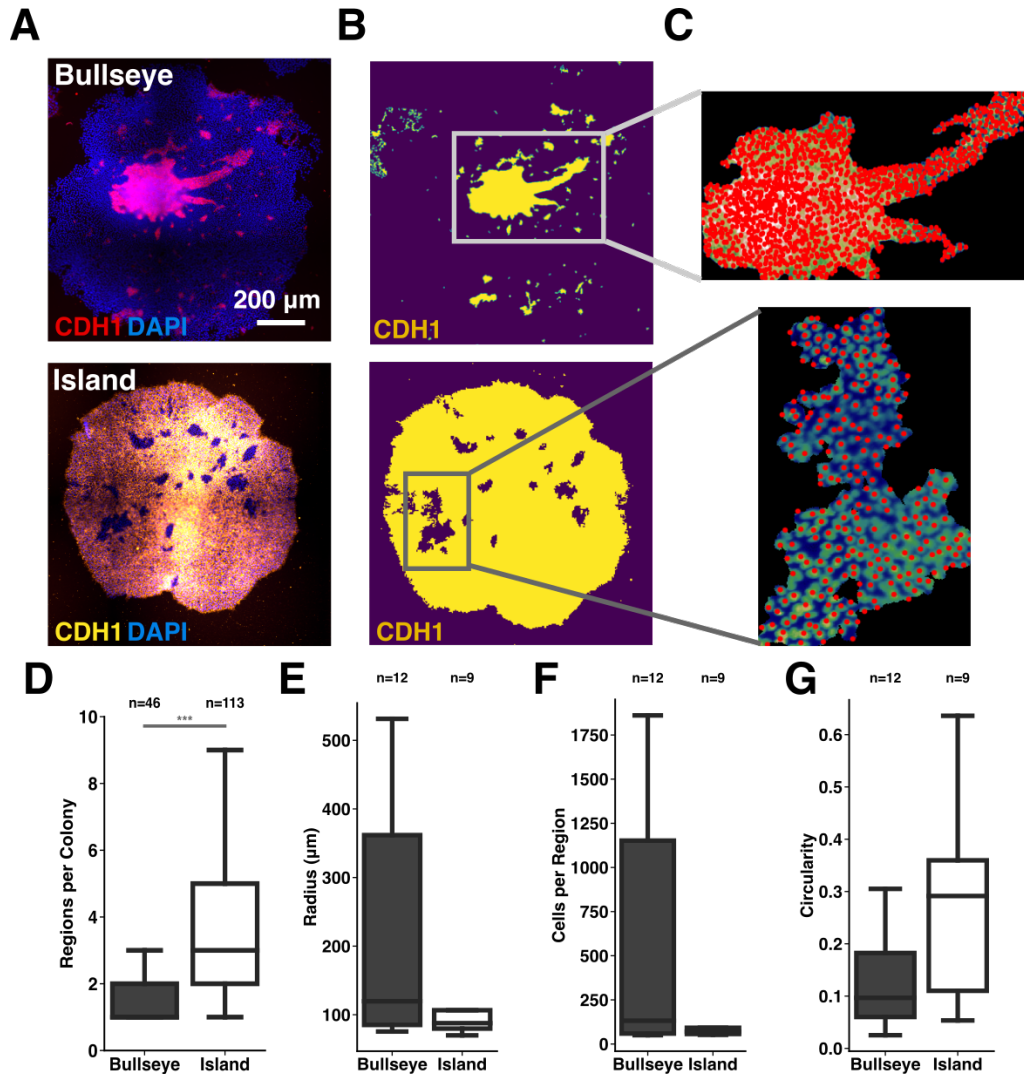
Figure 2.5 Spatial analysis of mosaic structures using nuclear tracking
A) Example bullseye (top) and island (bottom) colonies stained for CDH1 and DAPI, B) segmented for CDH1 area, C) with an example region showing nuclear localization. D) Quantification of number of regions per colony, and then for example colonies E) radii of each region, F) cells per region, and G) circularity of each region.

Previously published inducible CRISPRi cell lines inhibiting the expression of E-cadherin (CDH1) and Rho-associated coiled-coil containing protein kinase (ROCK1) were first characterized and then simulated using a machine recognition and pattern optimization framework, to generate predicted mixtures of cells that would undergo patterning events (Libby et al., 2018, 2019). Based on *in silico* predictions, a mix of 20%

CDH1 knockdown cells with 80% WT cells was used to generate colonies containing

agglomerations of CDH1-KD cells into clusters termed "islands", while a mix of 20%

ROCK1 knockdown cells to 80% CDH1 knockdown cells was used to generate a central

cluster of ROCK1-KD cells surrounded by CDH1-KD cells, termed a "bullseye".

Colonies were fixed and stained as described above, then each cell classified by the

presence or absence of CDH1 staining in a region surrounding each nucleus. The

contours of "islands" or "bullseye" structures were then calculated by finding all

connected regions of either CDH1+ (for bullseyes) or CDH1- (for islands), then

detecting the nuclei contained within each region (**Figure 2.5 A-C**). Number of regions

per colony was detected, with bullseye colonies having between a median of 1

connected region (IQR: 1-2), while island colonies had a median of 3 regions (IQR: 2-5)

(**Figure 2.5 D**). To characterize the spatial distribution of mosaic patterns within

selected colonies, shape and size parameters for each region were calculated, with

bullseyes having a median radius of 119 µm (IQR: 85 to 362 µm), a median of 132 cells

per region, and were very elongated, with median circularity of 0.1 (IQR: 0.06 to 0.18,

1.0 is circular). In contrast, islands were smaller and more round, with a median radius

of 88 µm (IQR: 80 to 106 µm) and a median circularity of 0.3 (IQR: 0.1 to 0.4) (**Figure

2.5 E-G**). These results demonstrated that nuclear segmentation of colonies could be

used to compare different modes of cell sorting in mosaic colonies and acted as a proof

of principle for using nuclear tracking to analyze time lapse studies of cell and colony

organization.

2.3.4 Cell tracking in cell sheets through nuclei detection

hESC cells expressing nuclear-localized mVenus were seeded 1:1 with WT cells and allowed to grow into ~50% confluent epithelial, then imaged in time lapse every 18 minutes for 24 hours during optogenetic stimulation. Analogous to segmenting stained colonies using nuclear Hoescht localization, peak detection with non-maximum suppression algorithm was adapted to track nuclear-localized mVenus, yielding detections of the rough center of mass of each cell at each frame. As a proxy for contour overlap, neighboring peaks in the next frame within a radius of 32.4 µm from a given peak were considered "overlapping" and ordered by distance. Peaks were then greedily matched to their closest neighbors, with ties going to cells in close proximity (Meijering et al., 2012). A second algorithm that performed a probabilistic match between close peaks was also compared (Gold and Rangarajan, 1996), but the tracks produced were not higher in quality, while the algorithm was much more computationally intensive. The resulting nearest neighbor assignment algorithm generated >1,000 cell tracks spanning 90 frames of observation (**Figure 2.6 A**).
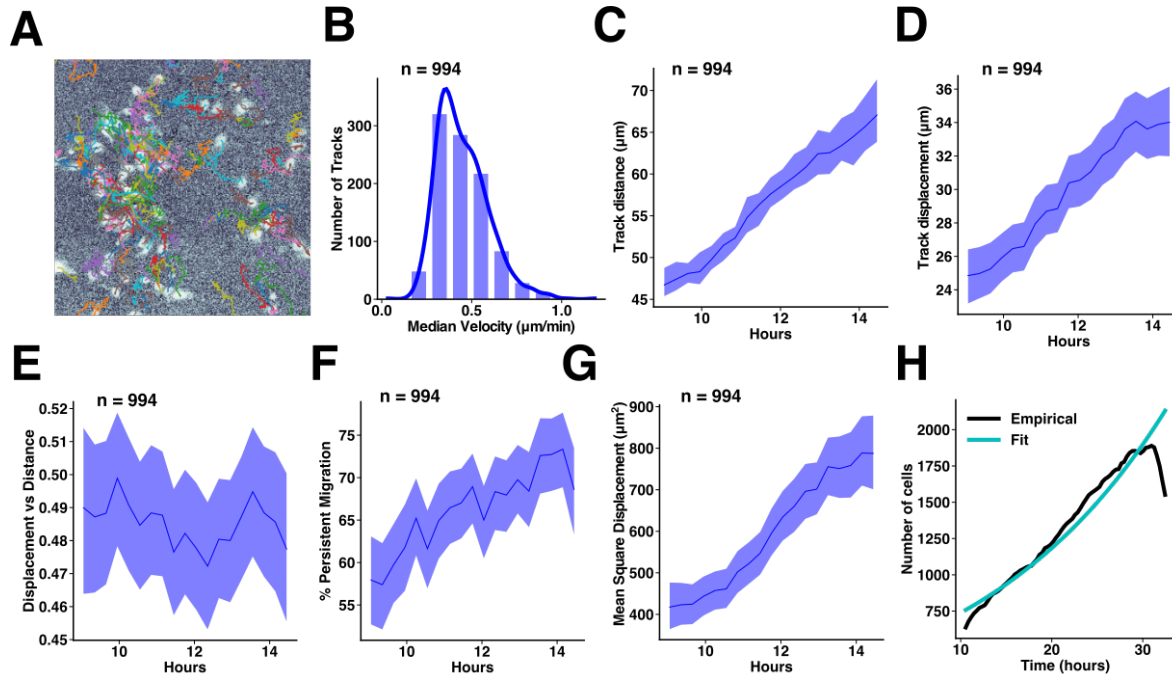
Figure 2.6 Baseline migratory behavior of optoWnt colonies
A) Image and estimated cell traces for a representative optoWnt stimulated colony. B) Distribution of median cell velocity over a 6 hour observation period. Changes in C) cell distance, D) cell displacement, E) displacement distance ratio, F) percentage travel in a uniform direction, and G) mean square displacement over 6 hours of observation. H) Cell growth over the entire 24-hour period

Distributions for cell velocity, distance, and displacement were calculated for the first

6 hours of observation and compared to analogous distributions from the contour

tracking algorithm (**Figure 2.6 B-D**). Cell median velocities were 0.35 μm/min (IQR: 0.3

to 0.42 μm/min), slightly faster than hiPSCs, while distance traveled was 72 μm and

displacement was 35 μm on average, giving an essentially constant displacement to

distance ratio of 0.49 (1.0 is directional migration) (**Figure 2.6 E**). Cells were 65%

persistent in their migration direction, with mean squared displacement 790 μm^2 over

the most linear 2 hour range, giving a diffusion coefficient of 87.7 μm^2/min, slower than

the results for hiPSC tracking (**Figure 2.6 F,G**). Despite some differences in values, the

congruence of individual cell parameters, combined with the increased computational

efficiency and higher density of cell tracking suggests that nuclear tracking is preferable when analyzing mononucleated cells if cell shape information is dispensable.

With the longer period of observation, robust estimates of other cell parameters such as growth rates, became possible. The number of cell traces was plotted and fit to an exponential model of cell growth using ordinary least squares. Cells grew at a rate of 0.7 cells/minute starting from 637 cells, giving an approximate time constant of 21 hours, in agreement with the heuristic that hESCs double in number every day (**Figure 2.6 H**). Despite obvious track breakages and migration of cells out of the field of view, these results suggest that nuclear tracking with nearest neighbor linkage produces sufficiently high-quality cell traces to evaluate colony growth and cell dynamics over extended periods.

2.3.5 EMT detection in OptoWnt stimulated cells

Wnt stimulation is commonly used as the first step in protocols that induce hiPSCs and hESCs to differentiation towards mesendodermal germ lineages. A consequence of the initial induction is a partial loss of the epithelial phenotype of stem cells as they undergo EMT, forming multilayered colonies with cells of different fates in each layer, or detaching from the colony altogether and becoming highly migratory. *In vivo*, Wnt activation represents a key signal which activates a portion of the epiblast stem cell pool to organize the primitive streak and subsequent multicellular morphogenic events, but the mechanism of heterotypic cell fate acquisition remains difficult to elucidate. While protein, small molecule, and genetic approaches all exist to stimulate aspects of the Wnt pathway in all cells, the consequences of heterotypic Wnt activation and the

subsequent cell behaviors upon mosaic acquisition of mesendodermal fates is underexplored, so an *in vitro* model to generate mosaic activation of Wnt signaling would be valuable to understanding primitive streak self-assembly.

The 1:1 population of optoWnt and WT stem cells was stimulated first for 8 hours without imaging, and then throughout the 24-hour imaging period, inducing visible changes 32 hours post stimulation including phase separation of the two populations and activation of the transcription factor TBXT in the optoWnt fraction, indicating transition out of pluripotency and towards mesendodermal fates. To determine the kinetics of optoWnt activation, rolling window smoothing was applied to all tracked cells and average velocity, displacement, and distance were calculated over time. Velocity increased from 0.30 µm/min to 0.42 µm/min, with a clear transition occurring between 22 and 30 hours (**Figure 2.7 A**). Mean squared displacement similarly increased from 600 µm^2 to a peak of over 1,000 µm^2, indicating an almost 2-fold increase in cell mixing rate (**Figure 2.7 B**). Interestingly, the ratio of displacement to distance did not change from ~0.4 indicating that cells were not traveling more directionally despite increased migration velocity (**Figure 2.7 C**). Another measure of persistent migration, percent of time spent traveling in a single direction, did increase from 50% to 65% leading to an apparent contradiction in stimulated cell migratory behavior (**Figure 2.7 D**).
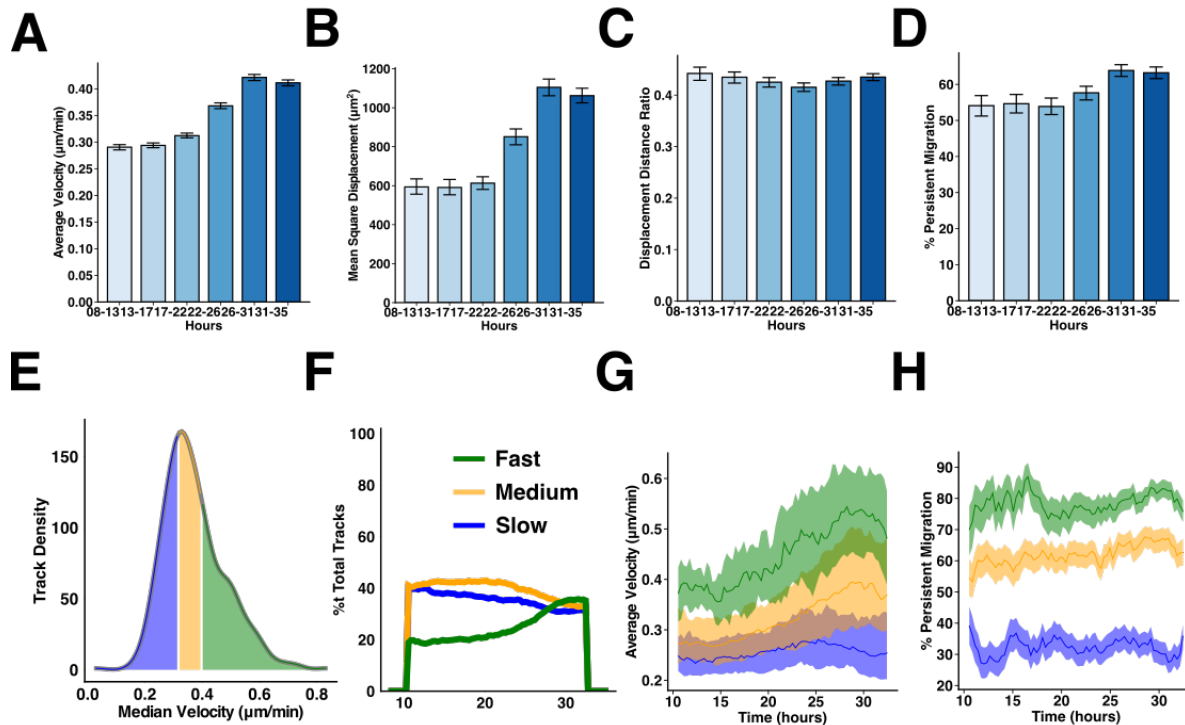
Figure 2.7 Cell behavior transitions during optoWnt stimulation
Changes during 36 hours of optoWnt stimulation in A) migration velocity, B) mean square displacement, C) displacement distance ratio, and D) persistent migration. E) Stratifying cells by mean velocity into 3 groups of slow (blue), medium (yellow), and fast (green) cells. F) Percent population membership in each group over time. G) Average population velocity in each group. H) Percent of time spent persistently migrating in each group.

To resolve this incongruity in cell migration directionality, cell tracks were fractionated into 3 groups by median track velocity, with low tracks having velocities ranging from 0-0.15 µm/min, mid speed tracks ranging from 0.15-0.18 µm/min, and the highest speed tracks averaging between 0.18-0.6 µm/min (**Figure 2.7 E**). Initially, 80% of cell tracks fell into the slow or mid bins, but between 22 and 30 hours, the number of cells in the high-speed bin increased from 20% of the population to 35% of the population. Median velocities in the high-speed group increased from 0.18 µm/min to almost 0.23 µm/min, while mid speed cells increased only from median 0.12 to 0.15 µm/min, while low speed cells were almost unchanged from 0.08 to 0.10 µm/min

(**Figure 2.7 F,G**). Cells in the highest velocity group consistently migrated directionally 80% of the time, with mid cells migrating 60% of the time and low cells migrating less than 30% of the time in a straight line (**Figure 2.7 H**). Hence, the transition in ensemble cell migration behavior during optoWnt stimulation is attributable to ~15% of cells converting from slow to fast migrating cells, simultaneous with increased cell velocity in those cells in the fastest population. These results indicate a potential new source of cell heterogeneity, where cell migration speed and persistence make cells more susceptible to Wnt stimulation and subsequent EMT transition.

2.4 Discussion

Robust estimation of multicellular state requires integration of a complex network of inputs, from internal transcriptional regulation, cell-cell signals, extracellular matrix patterning, soluble factors, as well as engineered genetic and chemical perturbations (Gerbin et al., 2020; Hackett and Surani, 2014). Feature-rich tools to bioinformatically interrogate cell genetic, epigenetic, and proteomic state have been employed to generate highly detailed description of lineage transitions during morphogenesis, but the destructive nature of these endpoint assays has limited the ability to assess the kinetics of morphogenic processes. In this chapter, I have demonstrated several image analysis approaches based on live fluorescent reporter imaging, enabling analysis of cell shape, position, and migratory behavior on the time scales of hours to days. By generating rich metrics of cell and colony shape and position, I've shown that force aggregation of heterotypic hiPSC populations followed by reattachment forms well mixed, circular colonies that maintain the engineered population ratios over a course of several days.

The features derived from cell migration demonstrate that hiPSCs are highly diffusive, preventing pattern formation through clonal expansion alone. Pointwise estimates of cell position through nuclei tracking were shown to recapitulate the same cell dynamic profiles as complete cytoplasmic segmentation, suggesting an alternate, computationally cheaper approach to generate robust cell traces. Finally, the kinetics of a mosaic Wnt differentiation were analyzed, demonstrating that cell tracking can reveal pre-differentiation stratification between cell populations, as well as providing quantitative measures of the rate and dynamic behavior changes that occur during mesendodermal differentiation in hESCs.

Live cell tracking represents a foundational approach to generating feature rich descriptions of hiPSC cell behavior complementary to traditional genomic and proteomic analyses. The interchangeability of different components, such as nuclear vs cytoplasmic detection, or different algorithms to assign tracks across frames gives a general framework for cell dynamic analysis, where each stage can be designed to answer a focused set of biological questions and trading accuracy for computational complexity as appropriate. Further, this tracking framework can be combined with constitutive labeling of subcellular components (Roberts et al., 2017), lineage tracing reporters (Cai et al., 2013; Henner et al., 2013), or label-free imaging techniques (Guo et al., 2019), to yield high resolution descriptions of subcellular, multicellular, and multilineage dynamic processes. Through a deeper understanding of the kinetics of cell behavior, live cell tracking enables future mechanistic models of morphogenesis capable of predicting the processes that underlie tissue formation.

2.5 Bibliography

Aman, A., and Piotrowski, T. (2010). Cell migration during morphogenesis. Dev. Biol. *341*, 20–33.

Bugaj, L.J., Choksi, A.T., Mesuda, C.K., Kane, R.S., and Schaffer, D.V. (2013). Optogenetic protein clustering and signaling activation in mammalian cells. Nat. Methods *10*, 249–252.

Cai, D., Cohen, K.B., Luo, T., Lichtman, J.W., and Sanes, J.R. (2013). Improved tools sfor the Brainbow toolbox. Nat. Methods *10*, 540–547.

Chuai, M., and Weijer, C.J. (2009). Regulation of cell migration during chick gastrulation. Curr. Opin. Genet. Dev. *19*, 343–349.

Damelin, S.B., and Hoang, N.S. (2018). On Surface Completion and Image Inpainting by Biharmonic Functions: Numerical Aspects. Int. J. Math. Math. Sci. *2018*, 1–8.

Doench, J.G., Fusi, N., Sullender, M., Hegde, M., Vaimberg, E.W., Donovan, K.F., Smith, I., Tothova, Z., Wilen, C., Orchard, R., et al. (2016). Optimized sgRNA design to maximize activity and minimize off-target effects of CRISPR-Cas9. Nat. Biotechnol. *34*, 184–191.

George, R.M., Maldonado-Velez, G., and Firulli, A.B. (2020). The heart of the neural crest: cardiac neural crest cells in development and regeneration. Development *147*, dev188706.

Gerbin, K.A., Grancharova, T., Donovan-Maiye, R., Hendershott, M.C., Brown, J., Dinh, S.Q., Gehring, J.L., Hirano, M., Johnson, G.R., Nath, A., et al. (2020). Cell states

beyond transcriptomics: integrating structural organization and gene expression in hiPSC-derived cardiomyocytes (Cell Biology).

Gold, S., and Rangarajan, A. (1996). Softmax to softassign: Neural network algorithms for combinatorial optimization. J. Artif. Neural Netw. *2*, 381–399.

Guo, S.-M., Yeh, L.-H., Folkesson, J., Ivanov, I., Krishnan, A.P., Keefe, M.G., Shin, D., Chhun, B., Cho, N., Leonetti, M., et al. (2019). Revealing architectural order with quantitative label-free imaging and deep learning (Biophysics).

Hackett, J.A., and Surani, M.A. (2014). Regulatory Principles of Pluripotency: From the Ground State Up. Cell Stem Cell *15*, 416–430.

Han, X., Chen, H., Huang, D., Chen, H., Fei, L., Cheng, C., Huang, H., Yuan, G.-C., and Guo, G. (2018). Mapping human pluripotent stem cell differentiation pathways using high throughput single-cell RNA-sequencing. Genome Biol. *19*.

Hashimoto, H., Robin, F.B., Sherrard, K.M., and Munro, E.M. (2015). Sequential Contraction and Exchange of Apical Junctions Drives Zippering and Neural Tube Closure in a Simple Chordate. Dev. Cell *32*, 241–255.

Henner, A., Ventura, P.B., Jiang, Y., and Zong, H. (2013). MADM-ML, a Mouse Genetic Mosaic System with Increased Clonal Efficiency. PLoS ONE *8*, e77672.

Hookway, T.A., Butts, J.C., Lee, E., Tang, H., and McDevitt, T.C. (2016). Aggregate formation and suspension culture of human pluripotent stem cells and differentiated progeny. Methods *101*, 11–20.

Kim, H.-B., and Lee, S.-J. (2002). Performance improvement of two-frame particle tracking velocimetry using a hybrid adaptive scheme. Meas. Sci. Technol. *13*, 573.

Libby, A.R., Joy, D.A., So, P.-L., Mandegar, M.A., Muncie, J.M., Mendoza-Camacho, F.N., Weaver, V.M., Conklin, B.R., and McDevitt, T.C. (2018). Spatiotemporal mosaic self-patterning of pluripotent stem cells using CRISPR interference. ELife *7*.

Libby, A.R.G., Briers, D., Haghighi, I., Joy, D.A., Conklin, B.R., Belta, C., and McDevitt, T.C. (2019). Automated Design of Pluripotent Stem Cell Self-Organization. Cell Syst. *9*, 483-495.e10.

Lou, X., Kang, M., Xenopoulos, P., Muñoz-Descalzo, S., and Hadjantonakis, A.-K. (2014). A Rapid and Efficient 2D/3D Nuclear Segmentation Method for Analysis of Early Mouse Embryo and Stem Cell Image Data. Stem Cell Rep. *2*, 382–397.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. Nat. Methods *3*, 637–646.

Maiuri, P., Terriac, E., Paul-Gilloteaux, P., Vignaud, T., McNally, K., Onuffer, J., Thorn, K., Nguyen, P.A., Georgoulia, N., Soong, D., et al. (2012). The first World Cell Race. Curr. Biol. *22*, R673–R675.

Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. Cell Stem Cell *18*, 541–553.

Meijering, E., Dzyubachyk, O., Smal, I., and others (2012). Methods for cell and particle tracking. Methods Enzym. *504*, 183–200.

Pantazis, P., and Supatto, W. (2014). Advances in whole-embryo imaging: a quantitative transition is underway. Nat. Rev. Mol. Cell Biol. *15*, 327–339.

Roberts, B., Haupt, A., Tucker, A., Grancharova, T., Arakaki, J., Fuqua, M.A., Nelson, A., Hookway, C., Ludmann, S.A., Mueller, I.M., et al. (2017). Systematic gene tagging using CRISPR/Cas9 in human stem cells to illuminate cell organization. BioRxiv 123042.

Saykali, B., Mathiah, N., Nahaboo, W., Racu, M.-L., Defrance, M., and Migeotte, I. (2018). Distinct mesoderm migration phenotypes in extra-embryonic and embryonic regions of the early mouse embryo (Developmental Biology).

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T. (2014). Scikit-image: image processing in Python. PeerJ *2*, e453.

Watanabe, K., Ueno, M., Kamiya, D., Nishiyama, A., Matsumura, M., Wataya, T., Takahashi, J.B., Nishikawa, S., Nishikawa, S., Muguruma, K., et al. (2007). A ROCK inhibitor permits survival of dissociated human embryonic stem cells. Nat. Biotechnol. *25*, 681–686.

Yang, J., Antin, P., Berx, G., Blanpain, C., Brabletz, T., Bronner, M., Campbell, K., Cano, A., Casanova, J., Christofori, G., et al. (2020). Guidelines and definitions for research on epithelial–mesenchymal transition. Nat. Rev. Mol. Cell Biol. *21*, 341–352.

(2017). Vertebrate Development (Cham: Springer International Publishing).

3 Deep neural net tracking of human pluripotent stem cells reveals intrinsic behaviors directing morphogenesis

3.1 Introduction

In the developing embryo, individual cells undergo a sequence of cell fate transitions and migration events to cooperatively form the tissues and structures of the organism. Cell tracking techniques based upon high resolution imaging have been used to trace cell lineage and describe the emergent patterns of embryogenesis across multiple model organisms (Chhetri et al., 2015; Peng et al., 2016; Sulston et al., 1983), including the early human pre-implantation embryo (Deglincerti et al., 2016; Shahbazi et al., 2016). However, automated tracking of cell migration within whole embryos *in vivo* has been limited both in size to small organisms such as *C. elegans* (Bao et al., 2006) due to the difficulty of identifying and tracking cells in a crowded multicellular environment, and in scale due to the low throughput of 3D imaging and reconstruction techniques (Stegmaier et al., 2016). Researchers frequently address the problem of density by employing sparse labeling of cells, either by only tracing cells of a single lineage (Cai et al., 2013; Henner et al., 2013), or by detecting transcriptional (Lou et al., 2014) or morphologic distinctions between cells (Stegmaier et al., 2016). Similarly, when analyzing cell behavior *in vitro*, experimental limitations such as mechanical confinement to one dimensional tracks (Maiuri et al., 2012), or sparse labeling (Libby et al., 2018) have been required to accurately track individual cells, limiting the ability of these systems to monitor multicellular tissue behaviors with comprehensive single cell resolution.

Self-organizing developmental processes are often initiated by small founder

populations within a larger population of physically inter-connected cells, as in the case

of classic Turing patterns (Turing, 1952). Similar multicellular organizational events

have been observed *in vitro* with human induced pluripotent stem cell (hiPSC),

revealing their heterogeneous differentiation potential due to global positional cues

(Warmflash et al., 2014), cell population boundaries (Libby et al., 2018), or cell-cell

interactions (Hookway et al., 2016). In particular, because cell fate and function are

strongly impacted by local interactions within multicellular networks (Malmersjo et al.,

2013; Novkovic et al., 2016; White et al., 2013), coordinated morphogenic processes

exhibit scale-free connectivity (i.e. at multiple scales, cell behavior is coordinated

through a central hub of influential cells) (Barabási et al., 1999), indicating that small

populations of cells, by establishing highly connected organizing centers, can exert a

large impact on the final composition of the developing tissue (Martinez Arias and

Steventon, 2018; Shahbazi and Zernicka-Goetz, 2018). Sparse labeling approaches

inherently under-sample these rare populations, highlighting the need for dense cell

tracking algorithms to definitively identify the origins and quantify the behaviors of

organizers.

Recent advances in machine learning, in particular in deep neural networks, have

demonstrated superhuman performance at image segmentation, revolutionizing the field

of computer vision (LeCun et al., 2015; Moen et al., 2019). Several classes of

convolutional neural nets (CNNs) have been developed specifically to perform dense

cell segmentation (Xie et al., 2016), based upon different architectures such as

autoencoders (Su et al., 2015), U-nets (Falk et al., 2019; Ronneberger et al., 2015; Xie et al., 2018), or variants of the Inception architecture (Cohen et al., 2017; Szegedy et al., 2014). Each architecture offers distinct trade-offs between cell segmentation accuracy, training efficiency, noise robustness, and computational complexity, with sub-optimal network choice leading to reduced tracking quality and poor capture of cell behavior. While cell tracking algorithms have historically been assessed through head-to-head competitions (Caicedo et al., 2019; Ulman et al., 2017), the potential advantage of combining complementary techniques for cell localization and tracking has been rarely employed.

In this study, I overcame the challenge of dense cell tracking by developing an ensemble of three neural networks (FCRN-B (Xie et al., 2016), Count-ception (Cohen et al., 2017), and a Residual U-net (Xie et al., 2018)) to localize each individual cell nucleus in an hiPSC colony. Nuclei displacements were then connected between sequential frames of a time series, enabling high spatiotemporal resolution of hiPSC behaviors over relevant developmental time scales of hours to days. This dense cell tracking pipeline revealed distinctive cell behaviors based on location within the colony, cell heterogeneity, and response to extracellular signaling molecules. Long-term cell tracking in combination with immunostaining for lineage markers, enabled tracking of the differentiation history of colonies with single-cell resolution. The whole-colony tracking and analysis pipeline revealed radially stratified shifts in cell migration speed and cell packing density in hiPSC colonies in reaction to changes in culture conditions. Changes in cellular behavior were detected at the local cell neighborhood level in

response to differentiation induced by externally applied morphogens, enabling real-time identification of local organizing centers (~10-20 cells) that precede tissue-scale morphogenic events. By detecting rare organizational events, my computational cell tracking pipeline allows for a more comprehensive dynamic understanding of the multicellular principles of morphogenesis, which can empower more refined control of organoid and engineered tissue development.

3.2 Methods

3.2.1 hiPSC culture

The hiPSC cell lines Wild-Type C11 was provided by the Conklin Lab and the Allen institute Lamin-B1 EGFP line (AICS-0013) was provided by Coriel. Both lines were cultured in feeder-free media on growth factor reduced Matrigel (BD Biosciences) and fed daily with mTeSR^{TM}-1 medium (STEMCELL Technologies)(Ludwig et al., 2006). Stem cells were dissociated to single cells using Accutase (STEMCELL Technologies) and passaged at a seeding density of 12,000 cells per cm^2. Rho-associated coiled-coil kinase 1 (ROCK-1) inhibitor, Y-276932 (10 μM; Selleckchem) was added to the media for the first 24 hours after passaging to promote survival(Watanabe et al., 2007). Starting with a base cell line cultured in mTeSR, cells were migrated to E8 (Gibco) by first culturing for one passage in a 1:1 mixture of E8 and mTeSR, followed by a minimum of two passages in E8 before evaluation.

3.2.2 Force aggregation of colonies

Cell aggregates consisting of 250 cells were generated using 400x400 μm PDMS microwell inserts in 24-well plates (975 microwells per well).(Hookway et al., 2016)

Dissociated cultures were resuspended in their respective maintenance media supplemented with Y-276932, mixed at the required cell ratio and concentration (250 cells/well), added to the microwells, and centrifuged (200 RCF). After 24 hours of formation, aggregates were transferred to ibidi slides (12 uWell format) or optically clear 96-well plates (Corning), coated with a substrate, and seeded at 10 aggregates/well or 18 aggregates/cm^2 and allowed to spread into colonies.

3.2.3 Substrate coating protocol

To promote aggregate attachment, ibidi slides were coated with growth factor-reduced Matrigel (80 $\mu g/mL$, BD Biosciences), vitronectin (10 $\mu g/mL$, Sigma Aldrich) or recombinant human laminin 521 (10$\mu g/mL$ rLaminin, Corning). Wells were uniformly coated using 125 μL/well (223 μL/cm^2) and incubated at 37 C following manufacturer's recommendations. Matrigel was incubated for 16 hours, while vitronectin and rLaminin were both incubated for 1 hour. Following manufacturer recommendation, rLaminin wells were additionally washed three times using cell culture grade water.

3.2.4 Differentiation protocol

Recombinant BMP4 (R and D Systems) was added to mTeSR at 50 ng/mL for 24 hours, starting 24 hours after aggregate seeding to induce a trilineage differentiation,(Warmflash et al., 2014) followed by 24 hours of imaging in mTeSR alone. CHIR differentiation was performed by adding 12 μM CHIR-99021 (Selleck Chemicals) to mTeSR for 24 hours, starting 24 hours after seeding, followed by 24 hours of imaging in mTeSR only. Dual SMAD inhibition was performed by adding both 10 μM SB-431542 (GlaxoSmithKline) and 0.2 μM LDN-193189 (Stemgent) to

mTeSR(Chambers et al., 2009) during force aggregation, and maintained for 72 hours through colony adhesion and imaging. CHIR pre-treatment during dual SMAD inhibition was achieved by adding 2 μM CHIR-99021 to mTeSR starting 48 hours before force aggregation, and continued through imaging (5 total days of treatment)(Libby et al., 2020).

3.2.5 Time-lapse imaging

Initial mixing studies were performed on an incubated inverted Axio Observer Z1 (Zeiss) microscope using an AxioCam MRm (Zeiss) digital CMOS camera at 20x magnification (NA 0.8, 0.323 μm x 0.323 μm per pixel). Colony positions were mapped using ZenPro software and approximately 30 colonies were imaged each experiment. Colonies were imaged over the course of 6 hours with images taken every 5 minutes. All subsequent studies were performed on an incubated spinning disk confocal Observer Z1 (Zeiss) using a motorized filter wheel (Yokogawa) and imaged using a Prime 95B (Photometrics) digital CMOS camera at 10x magnification (NA 0.45, 0.91 μm x 0.91 μm per pixel). Pluripotent colony studies were imaged over 6 hours with images taken every 3 minutes. Differentiation studies were imaged over 24 hours with images taken every 5 minutes.

3.2.6 Immunofluorescence staining

Within 15 minutes of the conclusion of imaging, slides and plates were washed once with PBS (125 μL/well, 220 μL/cm^2), then fixed for 30 minutes with 4% paraformaldehyde (100 μL/well, 178 μL/cm^2). Cells were permeabilized for 1 hour in 200 μL/well (357 μL/cm^2) of a blocking solution of 5% normal donkey serum, 0.3 %

Triton X-100 in PBS. Cells were incubated with the primary antibody for 1 hour in 100

μL/well (178 μL/cm^2) of a solution of 1% bovine serum albumin, 0.3 % Triton X-100 in

PBS. Cells were then incubated with the secondary antibody for 1 hour in 100 μL/well

(178 μL/cm^2) of a solution of 1% bovine serum albumin, 0.3 % Triton X-100 in PBS.

3.2.7 Human labeling of data set

To establish baseline human performance on labeling colonies, one annotator

labeled the first 12 frames of the time series for 8 colonies each of 9:1 (10% labeled),

7:3 (30% labeled) or 0:10 (100% labeled) wild type:GFP+ mixed colonies (336 frames

total). A power analysis was performed, indicating that 8 samples per condition in a 3-

way balanced design was required to distinguish between annotator performance when

labeling different colony mixture ratios (p <= 0.05 with 80 % confidence of rejecting the

null hypothesis).

To produce a validated human data set, 12 pairs of random sequential frames

from the original labeled data set were shuffled to obscure the order of the images (for a

total of 24 images, 8 per condition). With 50 % probability, each image was horizontally

mirrored (13 mirrored, 11 not mirrored), then with 25 % probability, each image was

randomly rotated in increments of 90 degrees (6 unrotated, 6 rotated 90 degrees, 9

rotated 180 degrees, 3 rotated 270 degrees). This data set was presented to seven

independent annotators using custom software written in Python.

A consensus segmentation was generated using k-means clustering of all

annotations on each frame with k equal to the largest number of points selected by any

individual annotator. Annotations were added to the consensus if they had at least 3

members in a cluster from unique annotators. This consensus segmentation was used as ground truth to calculate inter-rater reliability (IRR) for each annotator for each frame. Each frame to frame segmentation accuracy was measured using each annotator's inter-frame reliability (IFR) on the 12 pairs of images after inverting their transformation. IFR was compared between all pairwise two-sided t-tests with the Bonferroni Holm correction for multiple comparisons.

To evaluate the ability for this segmentation architecture to transfer learning to a different microscope, a second data set of confocal images was segmented by a single human annotator. The first two frames of each of 12 10% labeled colonies, 12 30% labeled colonies, and 8 100 % labeled colonies were segmented in sequence with no crops, flips, or rotations. Inter-frame reliability on this data set was not significantly different from the previous annotations, so this data set was used as a baseline for performance for transfer learning.

3.2.8 Initial neural net training

Each neural network architecture was implemented using Keras with the Tensorflow back end, with the input field of view for each architecture enlarged to 256x256 to enable fully convolutional segmentation of large images. Neural networks were trained using two NVidea GeForce GTX 1080 GPUs.

A training data set was generated from the initial segmentation by first splitting the images into 80% training, 10% test, and 10% validation folds. Then each fold was expanded by generating all possible 90-degree rotations and horizontal flips for each image in each data partition. Output point annotations were converted to cones

centered on the output with a radius of 4 pixels following.(Cohen et al., 2017) Input

output pairs were generated by selecting random 256x256 crops of each input image,

then further cropping the output image to match the output detector size (256x256 for

Residual U-net, FCRN-A and FCRN-B, 225x225 for Count-ception, 96x96 for U-net).

To establish the number of epochs to train each neural net before saturation, we

trained each net for 500,000 epochs and evaluated performance on the test set at the

10, 50, 100, 200, 300, 400, and 500 thousandth epoch using the Adam algorithm with a

learning rate of 1e-4. Each neural net had a slightly different loss behavior, but all nets

saturated around 100,000 epochs with highest performance on the test data at 50,000

epochs. Each architecture was then trained three times to 100,000 epochs and

evaluated at the 10, 25, 50, 75, and 100 thousandth epoch on the test data set. The top

three highest performing architectures were then ensembled to maximize train and test

score from compositing the individual net segmentations using grid search to weight

each net over the range 0.0 to 2.0 inclusive in steps of 0.1. All nets receiver operating

characteristic (ROC) and precision and recall were compared on the validation data set,

and ranked according to area under the curve (AUC).

To establish how well each network could transfer segmentations between

imaging systems, the confocal data set was also split into 80 % training, 10 % test, 10

% validation, and then each fold expanded by generating all possible 90-degree

rotations and horizontal flips for each image in each data partition. Neural nets trained

only on the inverted data set had poor performance, so each net was additionally

trained for 25,000 epochs with evaluations on the test set at the 1, 2, 5, 10, 15 and 25

thousandth epochs using Adam with a learning rate of 1e-5. Weights for the ensemble were recalculated using grid search as described above and net performance was again compared to the validation data set using AUC.

3.2.9 Cell correspondence algorithm

To detect cells in images of arbitrary size, each neural net was convolved with a zero-padded image to create a final output probability mask with the same size as the original image, with a convolution stride of 1, excluding 5 pixels around each border. To convert individual neural net predictions to cell point detections, the peak detections of each cell center were segmented using non-local maximum suppression with a minimal activation level of 0.1 and a minimal distance of 3 pixels.

Cell correspondence was found by greedily pairing the closest detected cell center in each frame to the next, taking the closest match in cases of multiple linkage, with a maximal link distance of 8 um. Since on average 5% of cells were not detected each frame, nearest neighbor linkage resulted in many short track fragments with single frame breaks that impeded long term cell tracking. To link discontinuous fragments, a dense mesh was imposed on colonies in space using Delaunay triangulation, and then holes in the mesh were detected by finding points connected to at least 3 other neighbor points in one frame, but missing in the next, then imputing their position using the average motion of the neighborhood. Finally, track fragments shorter than 15 minutes (3-5 frames) were removed from the data set as short fragments were found to not correspond to cells.

3.2.10 Track evaluation

To calculate individual track metrics, each individual track's x and y coordinates were first interpolated in time by a factor of 4 before smoothing using a rolling average with a filter width of 5 samples. Smoothed tracks were then used to calculate track length, total cell displacement, and velocity. Track persistence was calculated by analyzing the change in direction of travel at each step. A track segment was considered instantaneously persistent if the velocity was greater than 0.91 $\mu m/min$ and did not turn more than 3 degrees/min.

3.2.11 Spatial metrics

Whole colony metrics were calculated using a Delaunay triangulation after removing links longer than 50 μm (5 cell diameters away). The largest completely connected region was selected as the colony segmentation, and the perimeter and area of the whole region were calculated. To calculate estimates of density at each track point, the area of each triangle surrounding the point was calculated and the density estimated as the inverse of the average of those areas.

To map all colonies to a uniform coordinate system, the colony perimeter was projected onto the unit circle by calculating the angular position of each perimeter point, then using the distance from that perimeter point to the center as the local radius, and finally by gridding these radii onto a radially uniform 500-point grid. All interior points were then projected onto the unit circle by normalizing each point's radius with the average perimeter radius at the two nearest angular bins.

3.2.12 Statistical analysis of colony behavior

Average colony spatial behavior was assessed by dividing warped colonies into three annular rings of equal area: the center, a middle ring, and the periphery. Average density, velocity, and persistence were calculated for each bin averaging over all time and over each colony in the experimental group. All possible comparisons for each group and bin were performed using two-sided t-tests with the Bonferroni-Holm correction for multiple comparisons with significance was assessed at $p < 0.05$. Additionally, 95 % confidence intervals around the mean were calculated using 1000 iterations of bootstrap sampling. Effect size was assessed using Cohen's d using pooled standard deviations as measured using the maximum likelihood estimator.

3.3 Results

3.3.1 Manual annotation of cell migration in hiPSC colonies

Human iPSCs form dense, multilayered colonies *in vitro* with indistinctive boundaries between cells when using common phase imaging, pan-cytoplasmic, or pan-nuclear staining techniques. To establish a baseline for cell localization quality, a series of heterogeneously labeled colonies were generated by mixing wild type hiPSCs with an hiPSC-derived cell line expressing a nuclear GFP fluorescent label (Lamin-B::GFP) at ratios of 9:1 (10% labeled), 7:3 (30% labeled) or 0:10 (100% labeled). While maintaining the cells in pluripotency media, mixtures were force aggregated in microwells (Hookway et al., 2016), allowed to reattach to tissue culture plates, and then imaged every five minutes for six hours to generate a set of frames for annotation (**Figure 3.1A,B**). Seven individual human annotators selected the center of every GFP

positive cell nucleus in 12 sequential pairs of 500-cell colonies containing 10%, 30%, or 100% Lamin-B::GFP iPSCs presented in randomized order. A spatial average of all seven annotation sets was calculated using k-means clustering to generate a ground truth human consensus segmentation for each frame (**Figure 3.1 B**).

Human annotators were scored using a ratio of selected nuclei within a $5\mu m$ radius from the consensus cell, divided by the total number of expected cells, missing cells, and incorrectly selected cells (true positives divided by all positives plus any false positives). The average individual rater reliability (IRR) was 88.5% ($\pm$ 7.9 %) with a minimum of 83% and a maximum of 93% (**Figure 3.1 C**). As a second comparison, the individuals and human consensus were rated on their ability to select the same cell twice in pairs of sequential frames. Average inter-frame reliability (IFR) was 85.8% ($\pm$ 7.7%) with a minimum of 75% and a maximum of 92% (consensus 90.2% $\pm$ 4.8%) (**Figure 3.1D**). To assess how increasing label density impacted cell detection, performance metrics were stratified according to colony labeling density (**Figure 3.1E**). As expected, human annotators exhibited maximal IRR and IFR when evaluating colonies with the lowest percentage of GFP+ cells (i.e. 10%), with performance significantly declining for colonies containing higher proportions of GFP+ cells (30% and 100%).
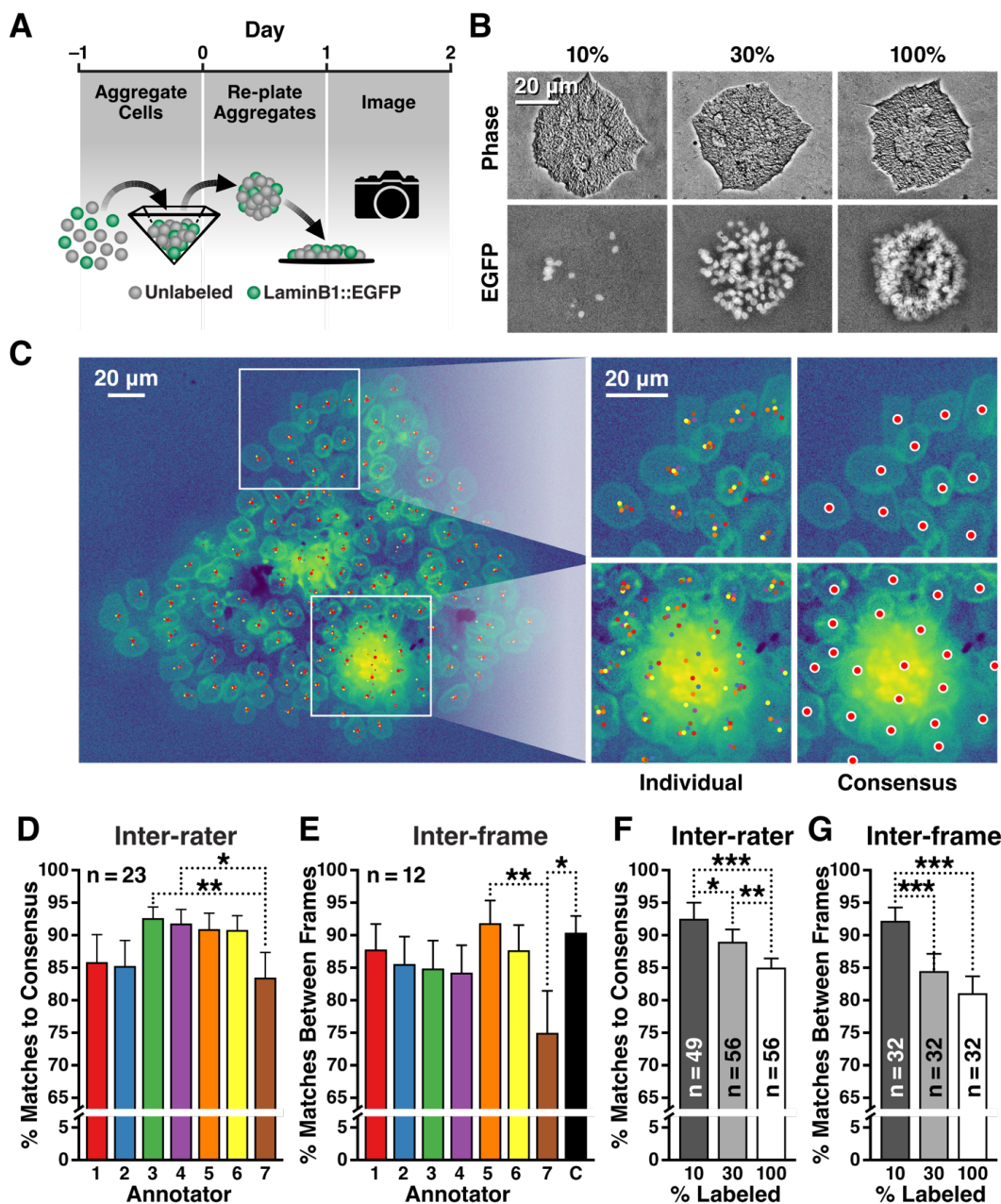
Figure 3.1 Quality of Manual Tracking Plateaus with Increasing Density of Labeled Cells
A) WTC11 and LaminB1::GFP cells were seeded into microwells, generating mixed aggregates with defined ratios of each population, then the aggregates were re-plated to form colonies. B) Annotators selected all cells in the colony, with high consistency in sparse regions, but lower agreement in dense regions. C) Individual annotator accuracy for each image was compared to the consensus for all images (n=23) with only annotator 7 different from any other rater (* $p < 0.05$, ** $p < 0.01$). D) Accuracy segmenting the same cell across sequential frames was assessed for all image pairs (n=12) but only annotator 7 was less repeatable than other annotators (* $p < 0.05$, ** $p < 0.01$). E) Colonies with 10%, 30%, and 100% LaminB1::GFP labeled cells were formed with labeled cells dispersed throughout the colony. F) Annotators were less accurate as compared to consensus on 100% labeled colonies than on 10% and 30% colonies (** $p < 0.01$, *** $p < 0.001$). G) Segmenting the same cell across sequential frames was also less repeatable in 100% labeled colonies vs 10%.

3.3.2 Ensemble deep neural network segmentation of dense hiPSC colonies

To determine how deep neural networks compare to human segmentation performance, a diverse array of independent cell segmentation network architectures was selected from recent literature (**Figure 3.2Ai**) and compared to the human annotator baseline as well as to the prediction of an ensemble of the selected architectures (**Figure 3.2Aii**). Five different neural net architectures were compared, including two networks with VGG-like architecture (FCRN-A and FCRN-B (Xie et al., 2016)), two U-net architectures (U-net (Ronneberger et al., 2015) and Residual U-net (Xie et al., 2018)), and an Inception-inspired network (Count-ception (Cohen et al., 2017)). Each neural network was trained to segment the GFP images of 10%, 30%, and 100% labeled colonies by predicting a cone-shaped probability around the human annotated center of each nucleus. Despite architectural differences, all neural networks exhibited comparable average performance, segmenting the data with a receiver operating characteristic (ROC) area under the curve (AUC) of 0.86 or better (**Figure 3.2B**). Although no individual neural network was able to equal human segmentation of 100% GFP-labeled colonies, an ensemble of the three highest performing networks surpassed human cell localization of fully-labeled colonies (**Figure 3.2Aii, B-D**). The primary variation between neural networks was due to spatial performance differences at the center or edge of individual colonies (**Figure 3.2E**).
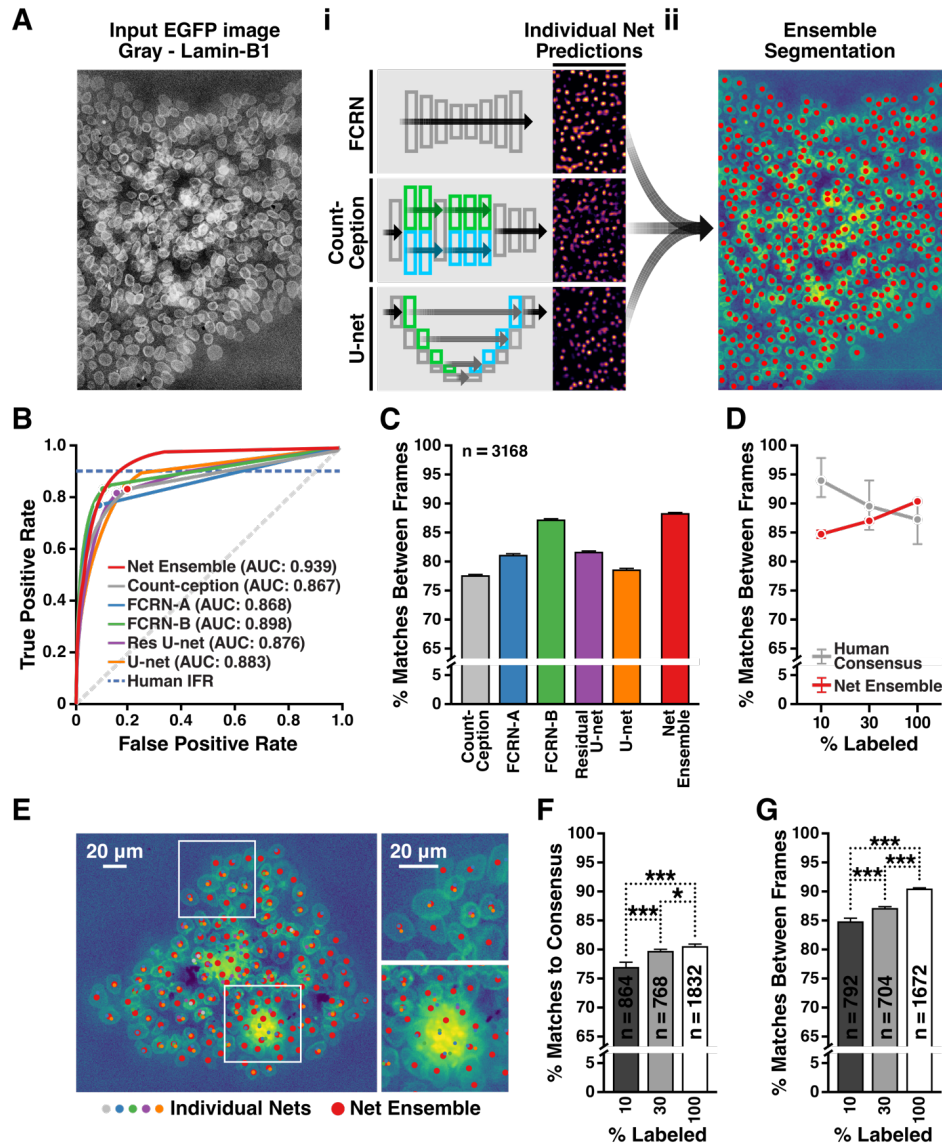
Figure 3.2 Heterotypic Neural Net Ensembles Generate Human Quality Segmentations
A) i. Individual images were segmented by one of several neural net architectures. producing probability maps localizing the center of each LaminB1::GFP labeled nuclei. ii. A weighted average of these maps from three different architectures (FCRN-B, Residual U-net, Count-ception) was used to produce the consensus segmentation. B) The true positive and false positive rate was calculated for each segmentation over the range of probability map thresholds between 0 and 1 and then the area under the curve (AUC) calculated for each architecture. C) Repeatability of cell detections between frames was calculated for the entire training set (n=3,168, $p < 0.001$ vs all single neural nets). D) Repeatability of cell detections was stratified by percent labeling and compared to the human annotator consensus, (* $p < 0.05$, 30% and 100% not significant). E) Representative image depicting individual and net ensemble detection ability where different colored dots indicate the peak probability of a cell as predicted by each neural net architecture. F) The agreement between net ensemble predicted labels and the human annotated data set was assessed for each label percentage (* $p < 0.05$, *** $p < 0.001$) G) The repeatability of the net ensemble detections over time was also compared across label percentages (*** $p < 0.001$)

80

Compared to the human annotator baseline, cell segmentation performance varied greatly between networks, with the two U-net architectures agreeing least with human annotators, whereas FCRN-B and the ensemble agreed most often (Figure 3.3A). However, ROC AUC (**Figure 3.2 B**) and effect size were often indistinguishable between similar architectures such as FCRN-A and FCRN-B (Cohen's $d = 0.11$) or U-net and residual U-net ($d = 0.02$). Segmentation speed varied widely between architectures (from 23 to 288 milliseconds per frame), but because the ensemble network was composed of several of the faster architectures, generating the composite segmentation was only 16.0% slower than using U-net only ($\pm$ 2.5% slowdown; **Figure 3.3B**). In contrast to human annotators, neural net IRR and IFR segmentation accuracy improved with increasing label density (3.6%$\pm$0.4% and 5.7%$\pm$0.4%, **Figure 3.2 F,G** respectively).
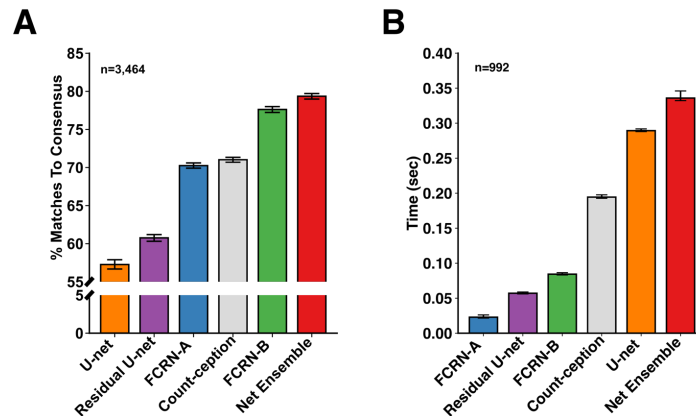
Figure 3.3 Inter-rater Reliability and Segmentation Times for Neural Nets
A) Individual rater reliability for each neural net architecture compared to human annotated dataset. B) Average time to segment one image for each neural net architecture. annotated dataset.

3.3.3 Individual cell tracking of pluripotent stem cell behavior

Individual frame segmentations were initially combined using a nearest neighbor linkage between frames to create cell tracks covering the center, middle, and edge regions of each colony (**Figure 3.4 A**, blue, gold, and red regions respectively), enabling construction of whole colony traces for all cells in 100 % GFP+ colonies over the entire time series (**Figure 3.4 B**). However, segmentation uncertainty at the individual cell level (e.g. a 95% accurate classifier will fail to detect a cell approximately once every 20th frame) led to artificially shortened tracks separated by short gaps of 1-5 frames. To reduce random breakages, a second linking step was added to combine tails of track fragments across gaps of up to 5 frames, using the motion of the local cellular neighborhood to interpolate any missing cell positions. Neighborhood interpolation significantly increased track fragment lengths (from average coverage of 21.5% of the time series length to 33.5%, **Figure 3.5 A**), bringing track fragment counts closer to the expected cell count based on cell seeding number and extrapolated growth rate - from 482 to 836 individual cells, with 1,000 expected (**Figure 3.5 B**).
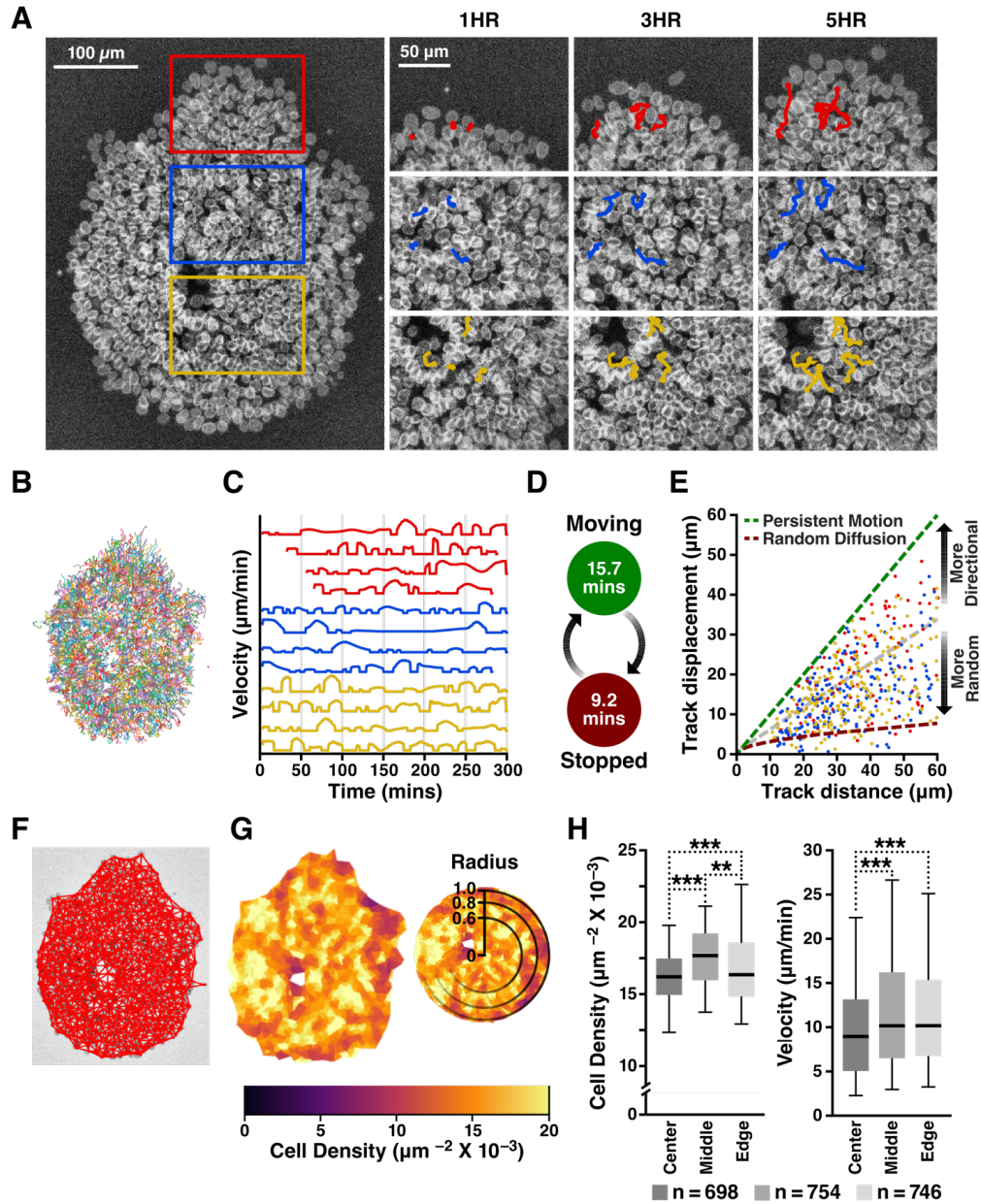
Figure 3.4 Spatio-temporal linkage of detections enables long term single cell tracking
A) Individual detections were linked across frames forming long tracks that spanned the entire time series. Depiction of example colony where regions were identified as center, middle, edge (colors). B) Dense track map created by linking detections covering the entire time series. C) Trace plot of example cell velocity tracks across colony locations (colors) D) Proposed two state model of alternating active migration and quiescence fit from average active and stopped periods. E) Distribution of the ratio of total track displacement to total track distance where colored dots represent individual tracks from the center, middle, and edge regions and dotted lines show the theoretical curves for persistent migration (dark green) and random diffusion (dark red) F) Delaunay triangulation depicted across an example colony to calculate cell neighborhoods, G) Average inverse area of Delaunay triangles around each cell (cell density) depicted on the example colony, and projected onto the unit circle. H) Quantification of cell density and velocity across the colony region identified by projecting triangulated cell position onto rings of the unit circle (** $p < 0.01$, *** $p < 0.001$).

83
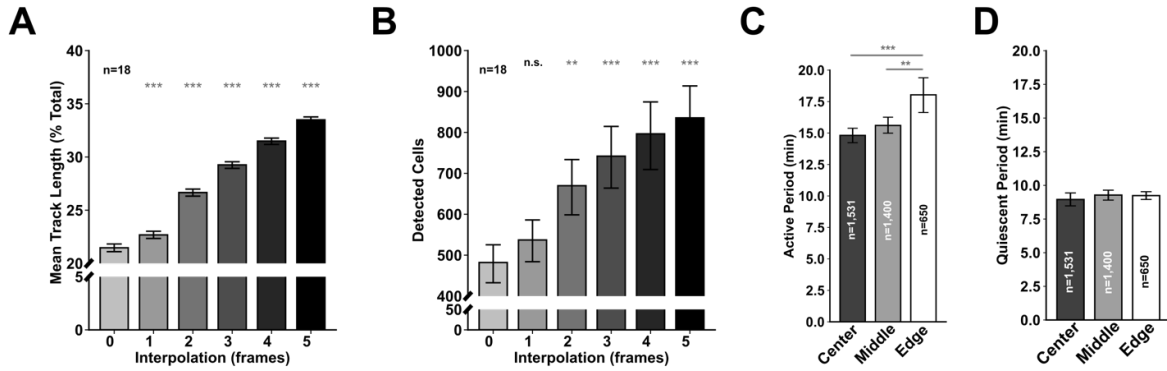
Figure 3.5 Track Interpolation and Cell Migration in Pluripotent Colonies
A) Mean track length increases with increasing number of interpolated frames (** p < 0.01, *** p < 0.001). B) Mean cell number increases with increasing number of interpolated frames (** p < 0.01, *** p < 0.001). C) Average period of active migration is higher at the edge than the center (** p < 0.01, *** p < 0.001). D) Average quiescent period between migrations is not different between the edge and center of colonies.

To understand how individual cell behavior contributes to colony spreading and density, I calculated persistence of cell migration by locating regions of each track where the direction of cell motion changed by less than 5 degrees per minute. Most cell tracks displayed clear binary switching between persistent migratory and stationary behavior (**Figure 3.4C**), with a mean active period of 15.7 minutes ($\pm$12.7 minutes) followed by a quiescent period of 9.2 minutes ($\pm$6.8 minutes), similar to the cyclic migration behavior observed in *E. coli* (Darnton et al., 2007) and eukaryotic cells (Devreotes and Janetopoulos, 2003) that can be attributed to the interaction between local polarizing cues and global inhibition of directional migration (**Figure 3.4D**). The active migration period was highest at the edge of colonies, and lowest at the center, while the quiescent period did not differ between colony regions (**Figure 3.5 C, D**).

To measure the extent to which individual cells traveled directionally or diffused randomly, we calculated the ratio of track displacement-to-distance, where a value of 1.0 represents travel in a straight line, lower values an increasingly curved trajectory,

and 0.0 a path that ultimately returns to its origin. Although cell tracks covered a broad

range between purely directional and random diffusion, there was no difference in

directionality of motion between cells at the center, middle or periphery of the colony

(**Figure 3.4 E**, blue, gold and red points respectively). Finally, to identify coordinated

movement between neighboring cells, I calculated correlation between each cell's

velocity profile and its immediate neighbors. In the center of colonies, nearest neighbors

had uncorrelated velocity profiles (Pearson's R = 0.00298, std 0.0757), whereas cells

near the periphery demonstrated much higher correlation (R = 0.118, std 0.383),

suggesting that observed peripheral spreading results from multi-cellular collective

migration, as has been shown previously in models of collective migration (Cui et al.,

2005; Pegoraro et al., 2016).

To analyze the dynamic behavior of iPSC colonies, a graph structure of each

colony was created using Delaunay triangulation (**Figure 3.4 F**). Based on the

triangulation, individual cell area was estimated using the average of all triangles

surrounding a cell within a maximum link distance threshold of 50 μm (**Figure 3.4 G**).

The entire colony mesh and all cell measurements, such as density or velocity, were

mapped onto the unit circle, then separated into three rings of equal area corresponding

to the center, middle, and periphery of the colony (**Figure 3.4 G**). In pluripotent colonies,

cells in the center region were packed more densely relative to the middle and

peripheral bins (p = 1.82×10^{-10} and 1.56×10^{-11}, respectively) (**Figure 3.4 H**), suggesting

local crowding effects contribute to radial inhomogeneities in cell packing in hiPSC

colonies. In contrast, cells in the middle and peripheral bins moved faster than cells in

the center ($p = 2.28*10^{-5}$ and 2.93^{-4} respectively, **Figure 3.4 H**), demonstrating an edge-biased cell migratory phenotype and suggesting that colony compaction may play a role in hiPSC colony spreading, as has been reported for migration of other epithelial cells (Cui et al., 2005; Pegoraro et al., 2016).

3.3.4 Packing and migratory behaviors of undifferentiated pluripotent stem cells

To interrogate the heterogeneous behavior of hiPSC colonies, I compared standard pluripotency maintenance conditions using the CNN tracking algorithm. First, I compared the effect of colony size on single cell behavior by forming colonies of either 100 or 500 cells (**Figure 3.6 Ai**). The average cell density and travel distance of 100-cell colonies were more similar to those of the edge of 500-cell colonies than to the center, suggesting that 100-cell colonies uniformly exhibit a similar phenotype to the edge of 500-cell colonies (**Figure 3.6 Aii**). At both colony sizes, cells at the edge displayed higher travel distance and migration speeds than those at the center (**Figure 3.6 Aiii,iv**). 100-cell colonies were more uniform in both density and cell distance traveled, with both measures closer to the cell density and travel values for the edge of 500-cell colonies. The transition in phenotype from edge-like to center-like cells as confluency increases may account for the observed sensitivity of hiPSC pluripotency and differentiation to cell plating density (Lian et al., 2013) and colony size (Warmflash et al., 2014).
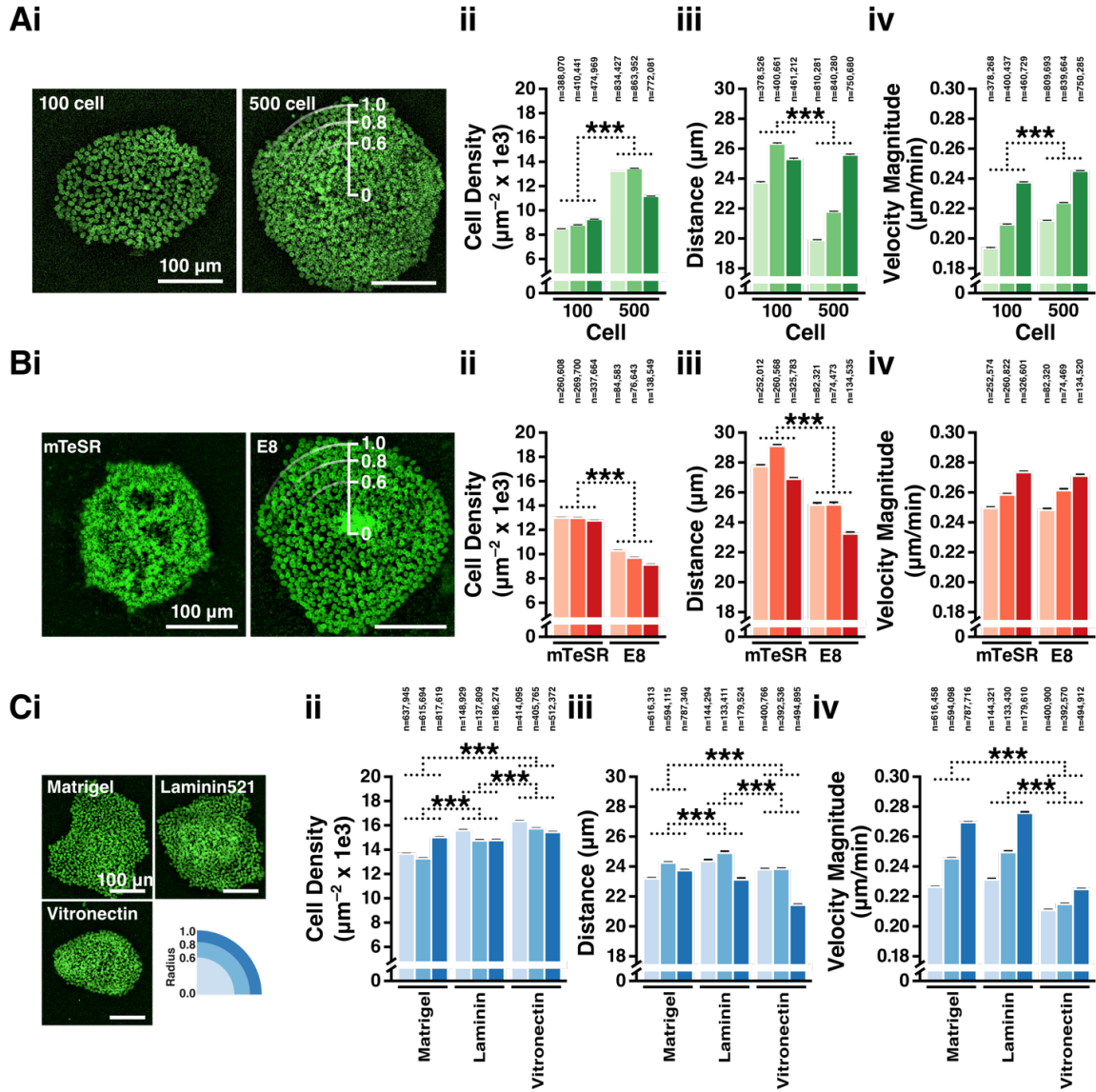
Figure 3.6 Basal culture conditions change cell packing density and migratory behavior
A) i. Example images of colonies with 100 or 500 starting cells. Comparison of 100 cell and 500 cell colonies stratified by colony region for: ii. average cell density (*** p < 0.001), iii. average total cell distance traveled in 6 hours (*** p < 0.001) and iv. average instantaneous cell velocity (*** p < 0.001). B) i. Example images of colonies generated from cells cultured in mTeSR or E8. Comparison of mTeSR and E8 colonies stratified by colony region for: ii. average cell density (*** p < 0.001), iii. average total cell distance traveled in 6 hours (*** p < 0.001) and iv. average instantaneous cell velocity (not significantly different). C) i. Example images of colonies adhered to either Matrigel, Vitronectin, or rLaminin521 coated plates. Comparison of Matrigel, Virtonectin and rLaminin521 colonies stratified by colony region for: ii. average cell density (*** p < 0.001), iii. average total cell distance traveled in 6 hours (*** p < 0.001) and iv. average instantaneous cell velocity (*** p < 0.001, Matrigel and rLaminin521 not significantly different)

Next, I explored the effect of pluripotency maintenance media on colony behavior by comparing the effect of passaging hiPSCs in mTeSR or E8 media (**Figure 3.6 Bi**). Colonies cultured in mTeSR were more compact with frequent formation of multi-layered structures and low-density regions in the center of the colony, while colonies cultured in E8 were uniformly flat with lower cell packing density (**Figure 3.6 Bii**). Individual cells within colonies cultured in E8 traveled less overall (**Figure 3.6 Biii**). Despite structural differences, cell migration velocities between the two conditions only differed slightly (p = 0.012, d = $9.44*10^{-3}$), indicating that the density shift could not be solely attributed to differences in cell motility between the two conditions (**Figure 3.6 Biv**).

Finally, I interrogated changes in colony phenotype due to commonly used adhesive extracellular matrices, which have been shown to have a cell-ECM strain-mediated effect on hiPSC morphology, behavior, and differentiation potential (Przybyla et al., 2016). hiPSC aggregates were allowed to adhere onto either Matrigel, Vitronectin, or recombinant Laminin 521 (rLaminin, **Figure 3.6 Ci**). Cell adhesion was much lower on rLaminin, with only 47.2% of aggregates adhered after 24 hours vs 91.7% on Matrigel and 97.2% on Vitronectin. Cells in adherent colonies on both rLaminin and Vitronectin had higher cell density than on Matrigel (**Figure 3.6 Cii**), while cells on Matrigel and rLaminin spread more than on Vitronectin (**Figure 3.6 Ciii**). Cells on Vitronectin had lower migration velocities, and much lower difference between center and edge migration velocities than either Matrigel or rLaminin (**Figure 3.6 Civ**). hiPSC behavior on Matrigel and rLaminin were very similar for both cell migration distance and

migration velocity, however stratifying the colonies by radius revealed that colonies plated on Matrigel were 11.1% less dense in the center. hiPSCs at the periphery of colonies grown on Vitronectin traveled only 92.7% of the distance for those on the edge of Matrigel or rLaminin colonies, and cells in Vitronectin colonies uniformly moved more slowly than those on other matrices, leading to more compact colony morphology overall. These results suggest that changes to substrate can subtly alter the local strain environment within a pluripotent stem cell colony, providing a mechanism to modulate peripheral migration and cell packing within hiPSC colonies.

Through dynamic characterization of hiPSC behavior, my tracking pipeline revealed that hiPSCs display a wide variety of heterogeneous behaviors while maintaining pluripotency. In particular, cells at the periphery of colonies exhibit a distinct phenotype from those in the center. Media environment and substrate can modulate both static and dynamic aspects of the edge and center phenotype. However, static snapshots of colony configuration, such as cell density, do not predict dynamic cell behaviors such as cell migration distance or velocity. Since both static and dynamic cell behaviors prime hiPSCs towards particular differentiation trajectories (Glen et al., 2018; Libby et al., 2019; Przybyla et al., 2016; Warmflash et al., 2014), dynamic assessment of whole colony behavior is necessary to illuminate the scope of hiPSC heterogeneity in pluripotency and predict priming during differentiation.

3.3.5 Lineage tracing of cell fate decisions during early morphogenic induction

Changes in hiPSC behavior during early lineage specification were then assessed by employing the tracking pipeline to analyze differentiation protocols used to

induce combinations of all three germ layers. Previous work has shown that multi-cellular annular ring patterns form during tri-lineage differentiation (Warmflash et al., 2014), but the dynamic changes to cell migration behavior during ring formation have not been described. In addition, protocols to induce either mesendoderm (Lian et al., 2013) or neuroectoderm (Chambers et al., 2009) have been reported, but whether those direct differentiation protocols induce similar dynamic transformations to those that occur during tri-lineage differentiation is not known. To monitor the transition from pluripotent cells to differentiating germ layers, a critical 24-hour morphogenic window was identified for each differentiation protocol for further exploration.

In the BMP4-induced trilineage protocol (**Figure 3.7Ai**), colonies adopted a round morphology 24 hours after BMP4 treatment with relatively uniform velocity and cell density, consistent with undifferentiated colonies (**Figure 3.8**). Approximately 32 hours after induction, cells across the colony slowed in migration velocity, except for a ring of cells at ~50% of the colony radius which maintained similar velocity to undifferentiated cells (**Figure 3.7Aii**, **Figure 3.8B**). In the center of the colony, cell density was constant for the entire period of observation, however, the periphery of the colony also began to rapidly decrease in cell density about 32 hours post-induction, with a dense plateau of cells forming at approximately 50% of colony radius, consistent with previous reports (Tewary et al., 2017; Warmflash et al., 2014) (**Figure 3.7Aiii, Figure 3.9A**). All three germ lineages formed by 48 hours, with OCT4+ cells in the center ring (SOX2-, EOMES-, presumptive endoderm), EOMES+ cells in the middle (presumptive mesoderm), SOX2+ cells at the colony edge (OCT4-, EOMES-, presumptive ectoderm),

90

and the periphery of the colony negative for all three markers (**Figure 3.7Aiii**). The peak

of EOMES expression corresponded with both the maximum of cell migration velocity

and the transition from high to low cell density, suggesting that the mesoderm ring acts

as a migratory barrier between ectoderm and endoderm, enabling the physical phase

separation of the colony into three distinct germ layers, analogous to gastrulation

(Shahbazi and Zernicka-Goetz, 2018; Tewary et al., 2017; Warmflash et al., 2014).
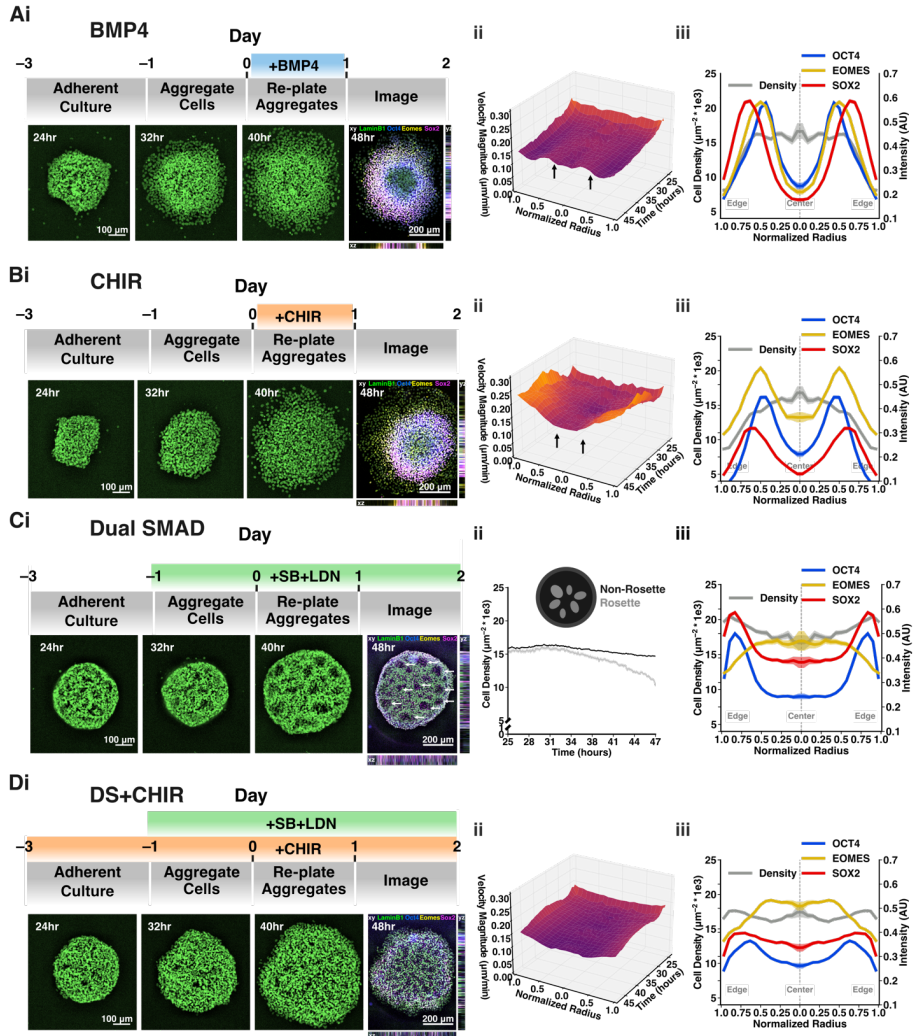
Figure 3.7 Whole Colony Analysis Reveals a Density Signature of Multi-lineage Differentiation
A) i. Treatment timeline and example time course of colony treated with BMP4 with example images at 24, 32, and 40 hours post re-seeding (HPR) and fixed and stained image of the same colony at 48 HPR. ii. Surface plot of temporal evolution of average instantaneous cell velocity over BMP4-treated colonies projected on to the unit circle (n=16 colonies). iii. OCT4, SOX2, and EOMES expression profiles and average cell density profile at 48 HPR projected onto the unit circle in BMP4-treated colonies (n=16 colonies). B) i. Treatment timeline of colony treated with CHIR with example images at 24, 32, 40, and 48 HPR. ii. Surface plot of temporal evolution of average instantaneous cell velocity over CHIR-treated colonies projected on to the unit circle (n=16 colonies). iii. OCT4, SOX2, and EOMES expression profiles and average cell density profile at 48 HPR projected onto the unit circle in CHIR-treated colonies (n=16 colonies). C) i. Treatment timeline of colony treated with Dual SMAD inhibition at 24, 32, 40, and 48 HPR with rosettes highlighted (white arrows). ii. Temporal evolution of average cell density inside and outside of rosettes (n=16 colonies). iii. OCT4, SOX2, and EOMES expression profiles and average cell density profile at 48 HPR projected onto the unit circle in Dual SMAD inhibition-treated colonies (n=16 colonies). D) i. Treatment timeline of colony treated with both Dual-SMAD inhibition and CHIR pre-treatment, at 24, 32, 40, and 48 HPR. ii. Surface plot of temporal evolution of average instantaneous cell velocity over DualSmad+CHIR-treated colonies projected on to the unit circle (n=16 colonies). iii. OCT4, SOX2, and EOMES expression profiles and average cell density profile at 48 HPR projected onto the unit circle in Dual SMAD+CHIR-treated colonies (n=16 colonies).
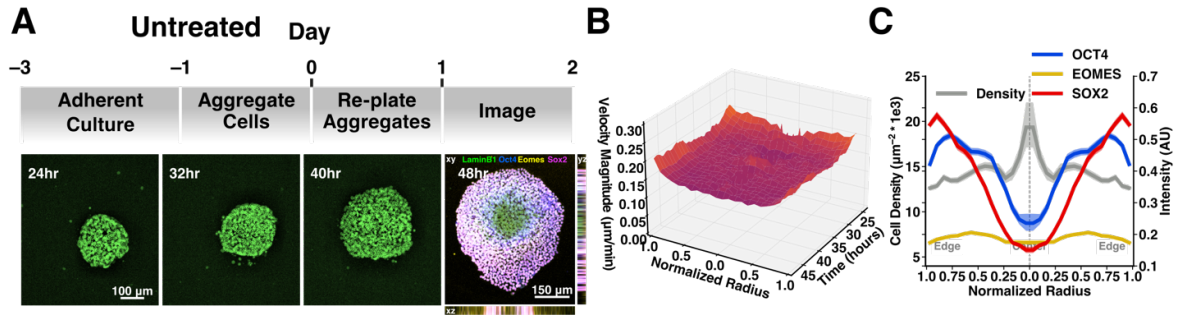
Figure 3.8 Dynamic Behaviors of Untreated Colonies over 24 hours
A) Treatment timeline and example time course of untreated colonies at 24, 32, and 40 HPR with fixed and stained image of the same colony at 48 HPR. B) Surface plot of temporal evolution of average instantaneous cell velocity over untreated colonies projected on to the unit circle (n=12 colonies). C) OCT4, SOX2, and EOMES expression profiles and average cell density profile at 48 HPR projected onto the unit circle in untreated colonies (n=12 colonies).

Treatment with the WNT activator CHIR is commonly used to induce differentiation of mesoderm (Lian et al., 2013). Tall, multilayered colonies (average $61.4 \pm 10.7 \mu m$) formed after 24 hours of 12 µM CHIR treatment, but by 48 hours a secondary flat epithelial ring expanded radially out from the colonies, ultimately forming a stratified colony similar to that induced by BMP4 (**Figure 3.7Bi**). Unlike in BMP4-treated colonies, CHIR-treated cells at the colony periphery increased in migration speed by 50%, with individual cells at the periphery of the colony undergoing EMT and traveling beyond the field of view (**Figure 3.10B**). Similar to BMP4 treatment, the central region maintained cell density similar to untreated colonies, while the middle and outer compartments rapidly decreased in density (**Figure 3.7Biii, 3.9B**). OCT4, SOX2, and EOMES were detected in all colonies, but levels of OCT4 and especially SOX2 were lower with CHIR than in BMP4 treated colonies, consistent with early CHIR induction directing differentiation towards mesoderm and away from neuroectoderm (**Figure 3.7Biii**). Again, the peak of EOMES expression occurred at ~50% of the colony radius

and corresponded spatially to the transitions between low to high velocity and high to

low density, respectively. The direct comparison of CHIR and BMP4 induced

differentiations demonstrates that limited numbers of similar static snapshots of colony

structure can mask distinctive cell behaviors that can indicate divergent differentiation

trajectories of pluripotent cells.
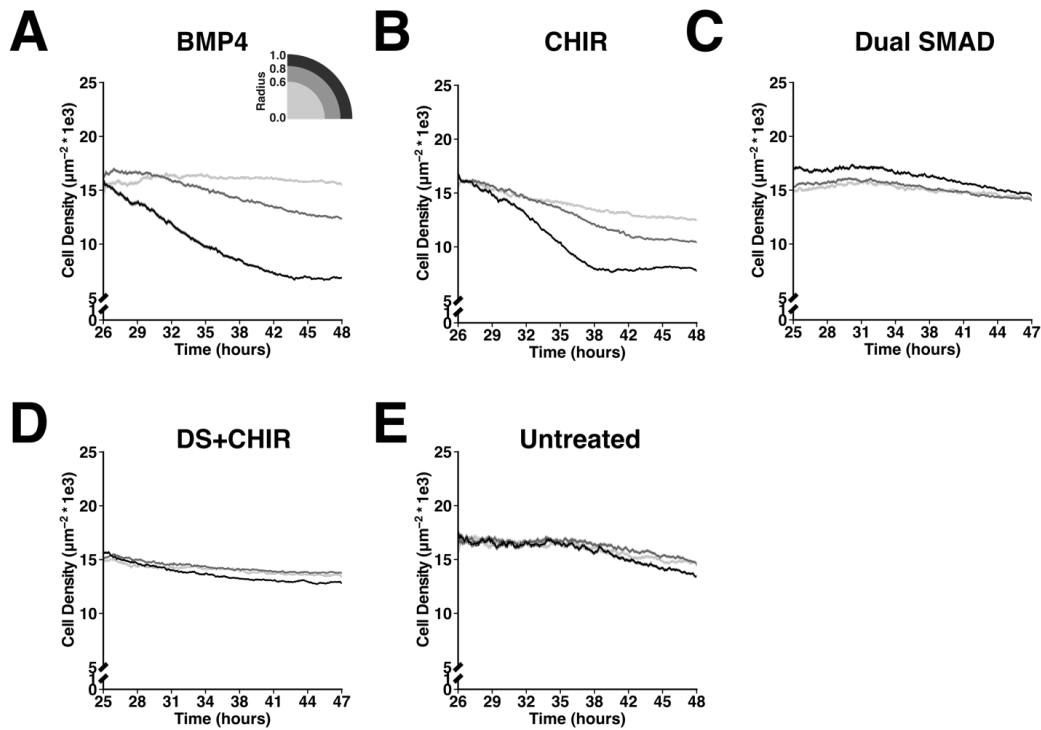
Figure 3.9 Cell Density Response to Morphogen Treatment
Temporal evolution of average cell density stratified by colony region (n=16 colonies/condition, n=12 untreated) in A) BMP4 treated colonies, B) CHIR treated colonies, C) Dual SMAD treated colonies, D) Dual SMAD colonies pre-treated with CHIR, and E) untreated colonies

Neuro-ectoderm directed colonies remained behaviorally indistinguishable from

untreated colonies through the first 48 hours of dual SMAD inhibition. However, starting

at 60 hours after treatment, small rosettes of approximately 20 cells began to form ring

structures that expanded continuously for the remaining 12 hours of imaging (**Figure**

3.7Ci). Between 6 and 18 rosettes formed per colony (mean 10.1 ± 2.3) with a mean

rosette diameter of $64.2 \pm 21.1 \mu m$. Rings consisting of regions of lower cell density

began to appear 36 hours after plating, with ring diameter expanding at a rate of $2.58 \pm$

$0.51 \mu m/hour$, and a mean center-to-center spacing between rings of $124.0 \pm 27.5 \mu m$

(**Figure 3.7Cii**). Average cell density was slightly higher at the periphery of colonies,

corresponding to higher expression of both OCT4 and SOX2 (EOMES-, potentially

undifferentiated cells), while the center of the colonies expressed high SOX2 and low

OCT4 (presumptive neuroectoderm, **Figure 3.7Ciii**). EOMES expression was slightly

elevated in the center of the colonies, but overall EOMES was rarely detected

compared to BMP4 or CHIR differentiation. None of the three lineage markers appeared

to be specifically localized to the ring structures. Addition of CHIR pre-treatment (Libby

et al., 2020) to the dual SMAD neuro-ectoderm protocol completely abrogated the

formation of rosettes (**Figure 3.7Di**). CHIR pre-treated neuro-ectoderm colonies were

indistinguishable from untreated colonies in both their uniform velocities and radial

distribution of cell densities (**Figure 3.7Dii** and **3.10D, 3.9D**, respectively). CHIR

treatment elevated expression of EOMES, and suppressed expression of both SOX2

and OCT4, likely delaying the commitment of cells to neuroectoderm fates, consistent

with its previously reported activity (Libby et al., 2020). By monitoring the trajectories of

differentiating colonies at single cell resolution, morphogenic signatures were detected

at both the local cell neighborhood and colony-wide levels, thereby enabling quantitative

measurement of the comprehensive dynamics of multicellular organization and subtle

yet distinctive differences in cell behavior that distinguish between independent

differentiation protocols.
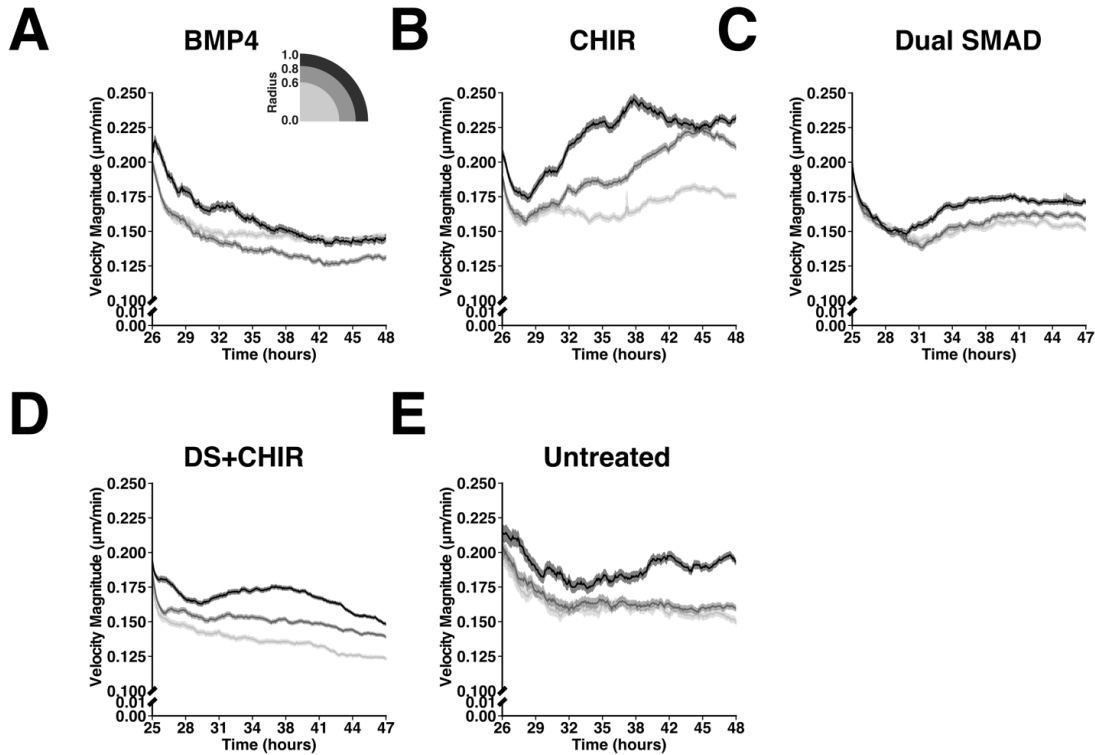
Figure 3.10 Cell Velocity Magnitude Response to Morphogen Treatment
Temporal evolution of average cell velocity magnitude stratified by colony region (n=16
colonies/condition, n=12 untreated) in A) BMP4 treated colonies, B) CHIR treated colonies, C) Dual
SMAD treated colonies, D) Dual SMAD colonies pre-treated with CHIR, and E) untreated colonies.

3.4 Discussion

Single cell analyses have highlighted the intrinsic heterogeneity present in virtually

all multi-cellular populations. Complementary approaches, such as automated cell

lineage tracing and single-cell RNA sequencing, have enabled fine-grained spatio-

temporal quantification of diverse and robust developmental processes (Bao et al.,

2006; Cai et al., 2013; Mohammed et al., 2017). Understanding the dynamic behavior(s)

of pluripotent stem cells in response to environmental factors can similarly clarify the

effects of multicellular structure and environmental factors on the behavior and ultimate fate of individual cells within developing tissues and organs *ex vivo*. To assess how organization arises from the collective action of individual cells, I developed a dense cell tracking platform to analyze time lapse imaging of hiPSC colonies with high spatiotemporal precision. Using the resulting quantitative measures of cell behaviors, I identified signatures of multicellular organization at the single cell, local neighborhood, and whole colony scale, demonstrating that hiPSC behaviors are influenced by short distance interactions between neighboring cells that propagate into global effects throughout an entire colony of 100's of cells and more. While many of the measured cell-intrinsic properties were relatively constant under pluripotent culture and early differentiation conditions, I found that the local cell neighborhood responds in characteristic ways to different external stimuli. Changes in cell-cell interactions are orthogonal to stem cell pluripotency (Libby et al., 2018; Przybyla et al., 2016), but can impact the sensitivity of hiPSCs to morphogenic cues (Libby et al., 2019), and thus may be a critical determinant in pre-patterning of cells to different cell fate decisions. The ability to specifically modulate cell-cell interactions through modification of culture conditions or genetic engineering provides new strategies to pre-pattern and control colony structure and subsequent differentiation trajectories (Libby et al., 2019). Furthermore, my live cell monitoring platform during early differentiation allows for non-destructive, high-throughput assessment of regional changes in cell fate, providing a critical first step towards feedback-control of hiPSC differentiation.

In this chapter, I applied the cell tracking system to resolve human pluripotent morphogenesis evolution at single-cell resolution to the maintenance of hiPSCs and early differentiation, but it can be used more generally to quantify multicellular structure at scale with either static or time-lapse microscopy of any cell line. Quantitative comprehensive characterization of cellular neighborhood dynamics will provide a robust approach to interrogate the effects of multicellular interactions among a broad range of cell types across many species, and will provide novel metrics to assess the fidelity of stem cell models to recapitulate developmental processes in a tissue context *ex vivo*. Ultimately, extracting unbiased cell dynamics from *in vitro* time-lapse imaging enables new insights into the complex processes underlying multicellular organization and morphogenesis.

3.5 Bibliography

Bao, Z., Murray, J.I., Boyle, T., Ooi, S.L., Sandel, M.J., and Waterston, R.H. (2006). Automated cell lineage tracing in Caenorhabditis elegans. Proc. Natl. Acad. Sci. U. S. A. *103*, 2707–2712.

Barabási, A.-L., Albert, R., and Jeong, H. (1999). Mean-field theory for scale-free random networks. Phys. Stat. Mech. Its Appl. *272*, 173–187.

Cai, D., Cohen, K.B., Luo, T., Lichtman, J.W., and Sanes, J.R. (2013). Improved tools for the Brainbow toolbox. Nat. Methods *10*, 540–547.

Caicedo, J.C., Goodman, A., Karhohs, K.W., Cimini, B.A., Ackerman, J., Haghighi, M., Heng, C., Becker, T., Doan, M., McQuin, C., et al. (2019). Nucleus segmentation across imaging experiments: the 2018 Data Science Bowl. Nat. Methods *16*, 1247–1253.

Chambers, S.M., Fasano, C.A., Papapetrou, E.P., Tomishima, M., Sadelain, M., and Studer, L. (2009). Highly efficient neural conversion of human ES and iPS cells by dual inhibition of SMAD signaling. Nat. Biotechnol. *27*, 275–280.

Chhetri, R.K., Amat, F., Wan, Y., Höckendorf, B., Lemon, W.C., and Keller, P.J. (2015). Whole-animal functional and developmental imaging with isotropic spatial resolution. Nat. Methods *12*, 1171–1178.

Cohen, J.P., Lo, H.Z., and Bengio, Y. (2017). Count-ception: Counting by Fully Convolutional Redundant Counting. ArXiv Prepr. ArXiv170308710.

Cui, C., Yang, X., Chuai, M., Glazier, J.A., and Weijer, C.J. (2005). Analysis of tissue flow patterns during primitive streak formation in the chick embryo. Dev. Biol. *284*, 37–47.

Darnton, N.C., Turner, L., Rojevsky, S., and Berg, H.C. (2007). On Torque and Tumbling in Swimming Escherichia coli. J. Bacteriol. *189*, 1756–1764.

Deglincerti, A., Croft, G.F., Pietila, L.N., Zernicka-Goetz, M., Siggia, E.D., and Brivanlou, A.H. (2016). Self-organization of the in vitro attached human embryo. Nature *533*, 251–254.

Devreotes, P., and Janetopoulos, C. (2003). Eukaryotic Chemotaxis: Distinctions between Directional Sensing and Polarization. J. Biol. Chem. *278*, 20445–20448.

Falk, T., Mai, D., Bensch, R., Çiçek, Ö., Abdulkadir, A., Marrakchi, Y., Böhm, A., Deubner, J., Jäckel, Z., Seiwald, K., et al. (2019). U-Net: deep learning for cell counting, detection, and morphometry. Nat. Methods *16*, 67–70.

Glen, C.M., McDevitt, T.C., and Kemp, M.L. (2018). Dynamic intercellular transport modulates the spatial patterning of differentiation during early neural commitment. Nat. Commun. *9*.

Henner, A., Ventura, P.B., Jiang, Y., and Zong, H. (2013). MADM-ML, a Mouse Genetic Mosaic System with Increased Clonal Efficiency. PLoS ONE *8*, e77672.

Hookway, T.A., Butts, J.C., Lee, E., Tang, H., and McDevitt, T.C. (2016). Aggregate formation and suspension culture of human pluripotent stem cells and differentiated progeny. Methods *101*, 11–20.

LeCun, Y., Bengio, Y., and Hinton, G. (2015). Deep learning. Nature *521*, 436–444.

Lian, X., Zhang, J., Azarin, S.M., Zhu, K., Hazeltine, L.B., Bao, X., Hsiao, C., Kamp, T.J., and Palecek, S.P. (2013). Directed cardiomyocyte differentiation from human

pluripotent stem cells by modulating Wnt/β-catenin signaling under fully defined conditions. Nat. Protoc. *8*, 162–175.

Libby, A., Joy, D., Elder, N., Bulger, E., Krakora, M., Gaylord, E., Mendoza-Camacho, F., Butts, J., and McDevitt, T. (2020). Elongation of Caudalized Human Organoids Mimics Neural Tube Development (Bioengineering).

Libby, A.R., Joy, D.A., So, P.-L., Mandegar, M.A., Muncie, J.M., Mendoza-Camacho, F.N., Weaver, V.M., Conklin, B.R., and McDevitt, T.C. (2018). Spatiotemporal mosaic self-patterning of pluripotent stem cells using CRISPR interference. ELife *7*.

Libby, A.R.G., Briers, D., Haghighi, I., Joy, D.A., Conklin, B.R., Belta, C., and McDevitt, T.C. (2019). Automated Design of Pluripotent Stem Cell Self-Organization. Cell Syst. *9*, 483-495.e10.

Lou, X., Kang, M., Xenopoulos, P., Muñoz-Descalzo, S., and Hadjantonakis, A.-K. (2014). A Rapid and Efficient 2D/3D Nuclear Segmentation Method for Analysis of Early Mouse Embryo and Stem Cell Image Data. Stem Cell Rep. *2*, 382–397.

Maiuri, P., Terriac, E., Paul-Gilloteaux, P., Vignaud, T., McNally, K., Onuffer, J., Thorn, K., Nguyen, P.A., Georgoulia, N., Soong, D., et al. (2012). The first World Cell Race. Curr. Biol. *22*, R673–R675.

Malmersjo, S., Rebellato, P., Smedler, E., Planert, H., Kanatani, S., Liste, I., Nanou, E., Sunner, H., Abdelhady, S., Zhang, S., et al. (2013). Neural progenitors organize in small-world networks to promote cell proliferation. Proc. Natl. Acad. Sci. *110*, E1524–E1532.

Martinez Arias, A., and Steventon, B. (2018). On the nature and function of organizers. Development *145*, dev159525.

Moen, E., Bannon, D., Kudo, T., Graf, W., Covert, M., and Van Valen, D. (2019). Deep learning for cellular image analysis. Nat. Methods *16*, 1233–1246.

Mohammed, H., Hernando-Herraez, I., Savino, A., Scialdone, A., Macaulay, I., Mulas, C., Chandra, T., Voet, T., Dean, W., Nichols, J., et al. (2017). Single-Cell Landscape of Transcriptional Heterogeneity and Cell Fate Decisions during Mouse Early Gastrulation. Cell Rep. *20*, 1215–1228.

Novkovic, M., Onder, L., Cupovic, J., Abe, J., Bomze, D., Cremasco, V., Scandella, E., Stein, J.V., Bocharov, G., Turley, S.J., et al. (2016). Topological Small-World Organization of the Fibroblastic Reticular Cell Network Determines Lymph Node Functionality. PLOS Biol. *14*, e1002515.

Pegoraro, A.F., Fredberg, J.J., and Park, J.-A. (2016). Problems in biology with many scales of length: Cell–cell adhesion and cell jamming in collective cellular migration. Exp. Cell Res. *343*, 54–59.

Peng, G., Suo, S., Chen, J., Chen, W., Liu, C., Yu, F., Wang, R., Chen, S., Sun, N., Cui, G., et al. (2016). Spatial Transcriptome for the Molecular Annotation of Lineage Fates and Cell Identity in Mid-gastrula Mouse Embryo. Dev. Cell *36*, 681–697.

Przybyla, L., Lakins, J.N., and Weaver, V.M. (2016). Tissue Mechanics Orchestrate Wnt-Dependent Human Embryonic Stem Cell Differentiation. Cell Stem Cell *19*, 462–475.

Ronneberger, O., Fischer, P., and Brox, T. (2015). U-net: Convolutional networks for biomedical image segmentation. In International Conference on Medical Image Computing and Computer-Assisted Intervention, (Springer), pp. 234–241.

Shahbazi, M.N., and Zernicka-Goetz, M. (2018). Deconstructing and reconstructing the mouse and human early embryo. Nat. Cell Biol. *20*, 878–887.

Shahbazi, M.N., Jedrusik, A., Vuoristo, S., Recher, G., Hupalowska, A., Bolton, V., Fogarty, N.M.E., Campbell, A., Devito, L.G., Ilic, D., et al. (2016). Self-organization of the human embryo in the absence of maternal tissues. Nat. Cell Biol. *18*, 700–708.

Stegmaier, J., Amat, F., Lemon, W.C., McDole, K., Wan, Y., Teodoro, G., Mikut, R., and Keller, P.J. (2016). Real-Time Three-Dimensional Cell Segmentation in Large-Scale Microscopy Data of Developing Embryos. Dev. Cell *36*, 225–240.

Su, H., Xing, F., Kong, X., Xie, Y., Zhang, S., and Yang, L. (2015). Robust Cell Detection and Segmentation in Histopathological Images Using Sparse Reconstruction and Stacked Denoising Autoencoders. In Medical Image Computing and Computer-Assisted Intervention – MICCAI 2015, N. Navab, J. Hornegger, W.M. Wells, and A.F. Frangi, eds. (Cham: Springer International Publishing), pp. 383–390.

Sulston, J.E., Schierenberg, E., White, J.G., and Thomson, J.N. (1983). The embryonic cell lineage of the nematode Caenorhabditis elegans. Dev. Biol. *100*, 64–119.

Szegedy, C., Liu, W., Jia, Y., Sermanet, P., Reed, S., Anguelov, D., Erhan, D., Vanhoucke, V., and Rabinovich, A. (2014). Going deeper with convolutions. In ArXiv Preprint ArXiv:1409.4842, p.

Tewary, M., Ostblom, J., Prochazka, L., Zulueta-Coarasa, T., Shakiba, N., Fernandez-Gonzalez, R., and Zandstra, P.W. (2017). A stepwise model of reaction-diffusion and positional information governs self-organized human peri-gastrulation-like patterning. Development *144*, 4298–4312.

Turing, A.M. (1952). The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. *237*, 37–72.

Ulman, V., Maška, M., Magnusson, K.E.G., Ronneberger, O., Haubold, C., Harder, N., Matula, P., Matula, P., Svoboda, D., Radojevic, M., et al. (2017). An objective comparison of cell-tracking algorithms. Nat. Methods.

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

White, D.E., Kinney, M.A., McDevitt, T.C., and Kemp, M.L. (2013). Spatial Pattern Dynamics of 3D Stem Cell Loss of Pluripotency via Rules-Based Computational Modeling. PLoS Comput. Biol. *9*, e1002952.

Xie, W., Noble, J.A., and Zisserman, A. (2016). Microscopy cell counting and detection with fully convolutional regression networks. Comput. Methods Biomech. Biomed. Eng. Imaging Vis. 1–10.

Xie, Y., Xing, F., Shi, X., Kong, X., Su, H., and Yang, L. (2018). Efficient and robust cell detection: A structured regression approach. Med. Image Anal. *44*, 245–254.

4 Elongation of Caudalized Human Organoids Mimics Neural Tube Development

4.1 Introduction

Development of multicellular organisms depends on the specialization and regionalization of multiple cell types to form appropriate tissue structures and organs. In human development, this process is intertwined with the initial establishment of the anterior-posterior embryonic axis (Steventon et al., 2016), which is required for neural tube and subsequent spinal cord development. Defects in early spinal cord morphogenesis result in severe congenital abnormalities, such as spina bifida, highlighting the clinical importance of understanding human axial elongation and neural tube development. However, despite the advances gleaned from vertebrate models such as mouse and chicken, many of the molecular mechanisms and cellular behaviors that regulate spinal cord development remain unknown due to the difficulty of studying these dynamic post-implantation processes (Schiffmann, 2007; Wilson et al., 2009). Experiments in model organisms have revealed that the posterior neural tube is partially generated from a unique pool of axial stem cells called neuromesodermal progenitors (NMPs). NMPs are bipotent progenitors that reside at the node-streak border within the caudal lateral epiblast and later in the chordoneural hinge of the tail bud (Beddington and Robertson, 1999; Cambray and Wilson, 2007; Wilson et al., 2009; Wymeersch et al., 2016). The lack of a model for posterior neural tube development has prevented understanding specifically how NMPs regulate and coordinate the emergence of the posterior spinal cord in humans. However, studies in model systems and 2D

differentiation models implicate WNT and FGF as critical factors regulating this process (del Corral and Storey, 2004; Ericson et al.; Liem et al., 1995; McMahon et al., 1998). Discoveries from embryonic studies have enabled the robust production of neuronal subtypes *in vitro* (Butts et al., 2017; Gouti et al., 2014). In parallel, organoids capable of recapitulating structures reminiscent of early embryonic tissues have been developed as high-throughput platforms for uncovering developmental patterns (Lancaster and Knoblich, 2014). More specifically, protocols have capitalized on NMP differentiation (Turner et al., 2014) to generate neuromuscular junctions (Martins et al., 2020), mimic spinal cord dorsal-ventral patterning (Veenvliet et al., 2020; Zheng et al., 2019), or recapitulate the spatiotemporal expression profiles of gastrulation (termed "gastruloids") (Beccari et al., 2018; van den Brink et al., 2020; Moris et al., 2020; Warmflash et al., 2014). However, organoids do not perfectly recapitulate the spatiotemporal dynamics of gene expression and differentiation that occurs within embryos. Moreover, reproducibility within differentiations, between cell lines and experiments, and across labs remains a significant technical hurdle (Ortmann and Vallier, 2017).

Building on the foundation of the limited existing 'gastruloid' models currently available, an organoid model was generated that reproduces many of the endogenous cell behaviors that generate axial elongation during human spinal cord development using both human embryonic stem cells (hESCs) and human induced pluripotent stem cells (hiPSCs). This model demonstrates extensive self-driven unidirectional growth, cell subtype specification that recapitulates NMP differentiation and neural tube morphogenesis, and regionalized gene expression profiles with distinct axial identities.

This robust model of neural tube morphogenesis and axial elongation enables direct examination of features of early human spinal cord development and patterning that have to this point remained unattainable.

4.2 Methods

4.2.1 Human Induced Pluripotent Stem Cell Line Generation and Culture

All work with human induced pluripotent stem cells (iPSCs) or human embryonic stem cells (ESCs) was approved by the University of California, San Francisco Human Gamete, Embryo, and Stem Cell Research (GESCR) Committee. Human iPSC lines were derived from the WTC11 line (Coriell Cat. #GM25256), the WTB line (Conklin Lab) (Miyaoka et al., 2014), and the Allen Institute WTC11-LaminB cell line (AICS-0013 cl.210) and the human ESCs H7 and H1 (WiCell, Madison, WI). All cell lines were karyotyped by Cell Line Genetics and reported to be karyotypically normal. Additionally, all cell lines tested negative for mycoplasma using a MycoAlert Mycoplasma Detection Kit (Lonza).

Human iPSCs were cultured on growth factor reduced Matrigel (Corning Life Sciences) and fed daily with mTeSRTM-1 medium (STEMCELL Technologies) (Ludwig et al., 2006). Cells were passaged by dissociation with Accutase (STEM CELL Technologies) and re-seeded in mTeSRTM-1 medium supplemented with the small molecule Rho-associated coiled-coil kinase (ROCK) inhibitor Y-276932 (10 μM; Selleckchem) (Ludwig et al., 2006) at a seeding density of 12,000 cell per cm^2.

The generation of the TBXT, Chordin and Noggin CRISPRi lines first involved TALEN mediated insertion of the CRIPSRi cassette pAAVS1-NDi-CRISPRi (Gen2)

(Addgene) to the AAVS1 locus of the Allen Institute WTC11-LaminB cell line. Following

antibiotic selection of clones that received the CRIPSRi cassette, CRISPRi gRNAs were

generated targeting Noggin and Chordin (**Table 4.1**) using the Broad Institute GPP Web

Portal and cloned into the gRNA-CKB (Addgene) following the previously described

protocol(Mandegar et al., 2016). Guide RNA vectors were nucleofected into the LaminB

CRISPRi iPSC line using a P3 Primary Cell 96-well NucleofectorTM Kit (Lonza) and the

4D Nucleofector X Unit (Lonza) following manufacturer's instructions. Nucleofected cells

were allowed to recover in mTeSRTM-1 medium supplemented with Y-276932 (10 µM)

and then underwent antibiotic selection with blasticidin (ThermoFisher Scientific; 10

µg/ml) following the previously published protocol (Libby et al., 2018; Mandegar et al.,

2016). Knockdown efficiency was evaluated by addition of doxycycline to the daily

feeding media over the course of 5 days, collection of mRNA, and subsequent

quantification of gene expression by qPCR.

Table 4.1 CRISPRi guide sequences

Gene Target (Symbol)	Guide sequence
Brachyury (TBXT)	CCTTGGACCGAGACCTGCGA
Noggin (NOG)	CTCCTCTCCCGGGTCTACTG
Chordin (CHRD)	AAGGAGCCGCTGCCCGTTCG

4.2.2 Organoid Differentiation

Organoid differentiations were a modified protocol of a previously published

spinal cord interneuron differentiation protocol (Butts et al., 2017). Human iPSCs were

seeded at 125000 cells/cm^2 in mTeSRTM-1 medium supplemented with the small

molecule Rho-associated coiled-coil kinase (ROCK) inhibitor Y-276932 (10 µM;

Selleckchem) and small molecule GSK inhibitor CHIR99021 (2µM, 4µM, or 6µM;

Selleckchem). Two days later, cells were singularized with Accutase (STEMCELL

Technologies), counted using a Countess II FL (Life Technologies), and seeded into

800µm X 800µm PDMS microwell inserts in a 24 well plate (~270 wells/insert)

(Hookway et al., 2016). After ~18 hours, condensed organoids were transferred to

rotary culture in 6-well plates in mTeSRTM-1 medium supplemented with Y-276932 (10

µM; Selleckchem), CHIR99021 (2µM, 4µM, or 6µM; Selleckchem), ALK5 small

molecule inhibitor SB431542 (10µM, Selleckchem), and small molecule BMP inhibitor

LDN193189 (0.2µM, Selleckchem) at an approximate density of 270 aggregates per

well unless otherwise mentioned in figure legend. Organoids were fed every other day

for up to 17 days. Y-276932 was removed from the media at day 3. At day 5 organoids

were transferred to Neural Induction Media (DMEM F:12 (Corning), N2 supplement (Life

Technologies), L-Glutamine (VWR), 2µg/ml heparin (Sigma Aldrich), non-essential

amino acids (Mediatech INC), penicillin-streptomycin (VWR), supplemented with fresh

0.4µg/ml ascorbic acid (Sigma Aldrich) and 10ng/ml brain derived neurotrophin factor

(BDNF, R&D Systems)) supplemented with CHIR99021 (2µM, 4µM, or 6µM;

Selleckchem), SB431542 (10µM, Selleckchem), and LDN193189 (0.2µM,

Selleckchem). From day 7 onwards, organoids were fed with Neural Induction Media

supplemented with retinoic acid (10nM, Sigma Aldrich), purmorphamine (300nM, EMD

Millipore) and N-[N-(3,5-difluorophenacetyl)-L-alanyl]-S-phenylglycine t-butyl ester (DAPT D5942, 1µM, Sigma-Aldrich).

4.2.3 Organoid Elongation Imaging and Quantification

Day 5 organoids from a single 10cm dish were individually transferred using wide bore pipette tips into the center 60 wells of an uncoated ultra-low attachment 96-well plate (Corning), seeding exactly one organoid per well, with the remaining organoids maintained in rotary culture through day 7. Using an inverted Axio Observer Z1 (Zeiss) microscope with incubation (Zeiss Heating Unit XL S, maintained at 37°C, 5% CO_2), all 60 wells were imaged using an AxioCam MRm (Zeiss) digital CMOS camera at 5x magnification (NA 0.16, 2.6 µm x 2.6 µm per pixel). Each well was imaged in TL Brightfield every 20 minutes for 48 hours giving a total of 145 frames. At the end of imaging (day 7), 31 organoids from the parallel rotary culture were imaged at 5x to generate a comparison image set.

To segment the organoids, all well images were first aligned by fitting a truncated quadratic curve to the average image intensity, then solving for the peak of maximum intensity, which was assumed to be the well center. Next, the average lighting inhomogeneity was calculated as the pixel-wise median of all 60 aligned well images, which was then subtracted from the individual aligned frames. After background correction, individual organoids were isolated by finding objects brighter than 0.83% of maximum intensity, but less than 3.0% of maximum intensity, with object size greater than 2,000 pixels, eccentricity greater than 0.1, and solidity greater than 40%. A bounding box 2 mm x 2 mm around the center of each of these objects was calculated

and all frames of the time series cropped to this bounding box to reduce memory usage.

To detect the region of maximum motion in the time series, the difference image

between each pair of sequential images was calculated, and then the pixel wise

standard deviation was calculated over all difference images in a given region. This

standard deviation image was then thresholded at between 0.01 and 0.03 (AU)

depending on the remaining lighting inhomogeneity in the image, producing a ring-

shaped mask around the periphery of each organoid. Finally, using the interior of the

mask as the organoid seed and the exterior as the background seed for the first frame,

organoids were segmented using anisotropic diffusion (Grady, 2006), evolving the

foreground and background seeds using the contour calculated from the previous frame

for subsequent segmentations. Segmentation, labeling, and metrology were all

performed using the python package sckit-image (van der Walt et al., 2014).

Segmentations were manually inspected for accuracy, with 45 of 60 determined

as having no or only minor flaws, with the remaining 15 excluded from automated

analysis. Using the high-quality segmentations only, each organoid time series was then

analyzed to examine geometry change over time. For each contour at each time point,

we calculated contour area, contour perimeter, minimum, maximum and mean distance

from contour center of mass to the perimeter. As additional non-dimensional measures

of shape, we calculated the ratio of maximum to minimum radius and organoid

circularity. Organoids were also manually classified as "extending", "partially extending",

or "non-extending" by examining each video. Organoids assigned to "extending"

exhibited at least one, and at most two large protuberances that extended at least

100μm from the main body. Partially extending organoids exhibited at least one, and often many protuberances, all of which failed to extend robustly past the 100μm demarcation. Non-extending organoids were any organoids that failed to generate any extensions over the observation period.

4.2.4 Dissection of Extended Organoids

Dissections were performed on a SteREO Discovery.V8 Manual Stereo Microscope (ZEISS) and images were taken with an EP50 Microscope Digital Camera (Olympus). Day 9 organoids were transferred into 12 mL of 37°C PBS in a 10 cm dish. 2 mL of warmed PBS were also added to 2 wells of a 6-well plate. Using two fine point forceps, polarized elongated aggregates were immobilized and pulled or pinched apart. Using a p200 pipette with a cut tip, the 'anterior' and 'posterior' halves of the organoids were collected and transferred to separate wells of the 6-well plate. After ~25 organoids were dissected, the collected halves were transferred to Eppendorf tubes and suspended in RLT buffer (RNAeasy Mini Kit; QIAGEN) for RNA extraction. Remaining, whole undissected organoids were also collected and suspended in RLT buffer for comparison. Standard RNA extraction was performed following manufacturer specifications.

4.2.5 Real Time Quantitative Polymerase Chain Reaction

Total RNA was isolated from organoid samples using an RNAeasy Mini Kit (QIAGEN) according to manufacturer's instructions. Subsequently, cDNA was generated using an iScript cDNA Synthesis kit (BIORAD) and the reaction was run on a SimpliAmp thermal cycler (Life Technologies). Quantitative PCR reaction using Fast

SYBR Green Master Mix (ThermoFisher Scientific) and run on a StepOnePlus Real-Time PCR system (Applied Biosciences). Relative gene expression was determined by normalizing to the housekeeping gene 18S rRNA, using the comparative threshold (CT) method. Gene expression was displayed as fold change of each sample versus control. The primer sequences were obtained from the Harvard Primer bank or designed using the NCBI Primer-BLAST website (**Table 4.2**).

Table 4.2 qPCR Primers

Target	Forward Primer	Reverse Primer
Axin 2 (AXIN2)	GGC GGG ATC ACT GGC TC	GGG CTC ATC TGA ACC TCC TC
Brachyury (TBXT)	TTT CCA GAT GGT GAG AGC CG	CCG ATG CCT CAA CTC TCC AG
Brachyury (TBXT)	TTG GCC TTG GAC CGA GAC CTG CGA	AAA CAG CGT CCA GAG CCA GGT TCC
Caudal type homeobox 2 (CDX2)	GCA GCC AAG TGA AAA CCA GG	TTC CTC TCC TTT GCT CTG CG
Chordin (CHRD)	TATGCCTTGGACGAGACGTG	ATGTTCTTGCAGCTGACCCT
Cytochrome P450 family 26 subfamily A member 1 (CYP26A1)	ATG AAG CGC AGG AAA TAC GG	AGG AGT CGT GCA GGT TAG AGA
Engrailed Homeobox 1 (EN1)	CGCCCAGTTTCGTTTTCGTT	GCAGAACAGACAGACCGACA
Fibroblast growth factor 8 (FGF8)	CCTTCGCAAAGCTCATCGTGG	CACAATCTCCGTGAAGACGCAG
GATA Binding Protein 6 (GATA6)	TCT CCA TGT GCA TTG GGG AC	AAG GAA ATC GCC CTG TTC GT

Table 4.2 continued

Target	Forward Primer	Reverse Primer
Hes Famil y BHLH Transcription Factor 1 (HES1)	TCAACACGACACCGGATAA AC	GCCGCGAGCTATCTTTC TTCA
Homeobox A1 (HOXA1)	CTA CCA GAC TTC CGG GAA CC	CCC ACC ACT TAC GTC TGC TT
Homeobox A11 (HOXA11)	GGA AGA GGG CTG CAA ATC CT	CAC CTC AGG GAA CAG TCC AC
Homeobox A3 (HOXA3)	ATG CAA AAA GCG ACC TAC TAC G	TAC GGC TGC TGA TTG GCA TTA
Homeobox A6 (HOXA6)	TCC CGG ACA AGA CGT ACA C	CGC CAC TGA GGT CCT TAT CA
Homeobox A9 (HOXA9)	GTC CAA GGC GAC GGT GTT T	CCG ACA GCG GTT CAG GTT TA
LIM Homeobox 2 (LHX2)	AAGTTCAGGCGCAACCTCT T	AAGACGGACGTCACAGT TGG
LIM Homeobox 5 (LHX5)	GTGCAAAGACGACTACCTG AG	CGGTCCGTACAGGATGA CAC
Noggin (NOG)	GCTGCGGAGGAAGTTACAG A	ACGAGCGCTTACTGAAG CAG
Oligodendrocyte Transcription Factor 2 (OLIG2)	CGCATCCAGATTTTCGGGT C	AAAAGGTCATCGGGCTC TGG
RNA, 18S Ribosomal 5 (RNA18S5, 18S)	CTC TAG TGA TCC CTG AGA AGT	ACT CGC TCC ACC TCA TCC TC
RNA, 18S Ribosomal 5 (RNA18S5, 18S)	CTTCCACAGGAGGCCTACA	CTTCGGCCCACACCCTT AAT
SRY-box 2 (SOX2)	AAC CAG CGC ATG GAC AGT TA	CGA GCT GGT CAT GGA GTT GT
SRY-box 2 (SOX2)	CCG TTC ATC GAC GAG GCT AA	TAA CTG TCC ATG CGC TGG TT

Table 4.2 continued

Target	Forward Primer	Reverse Primer
T-box transcription factor 6 (TBX6)	CATCCACGAGAATTGTACCCG	AGCAATCCAGTTTAGGGGTGT
Visual System Homeobox 2 (VSX2, CHX10)	CGGCGACACAGGACAATCTT	CCTGTATCCTGTCTTCCGGC
Wnt family member 3 (WNT3)	CAC AAC ACG AGG ACG GAG AA	GCT TCC CAT GAG ACT TCG CT

Due to low RNA abundance in the dissected organoids, PreAmp Master Mix (FLUIDIGM) was used to amplify cDNA. Briefly, 20ng of cDNA were amplified per 5μL reaction volume for 15 cycles on a SimpliAmp thermal cycler (Life Technologies). Amplified cDNA was diluted 5-fold using nuclease free water. 1μL of amplified diluted cDNA was used for each 20μL quantitative PCR reaction using Fast SYBR Green Master Mix (ThermoFisher Scientific) and run on a StepOnePlus Real-Time PCR system (Applied Biosciences), following normal qPCR methods described above.

4.2.6 Histology, Immunocytochemistry and Imaging

Organoids were fixed with 4% paraformaldehyde (VWR) for 40 minutes, washed three times with PBS. Organoids to be used for histology were embedded in HistoGel Specimen Processing Gel (Thermo Fisher) prior to paraffin processing. Parafin embedded samples were sectioned in 5m sections, and subsequently stained for H&E. For immunofluorescent staining, slides were deparaffinized as for H&E staining. Epitope retrieval was performed by submersing slides in Citrate Buffer pH 6.0 (Vector Laboratories) in a 95°C water bath for 35 minutes. Samples were permeabilized in 0.2%

Triton X-100 (Sigma-Aldrich) for 5min, blocked in 1.5% normal donkey serum (Jackson

Immunoresearch) for 1 hour, and probed with primary antibodies against SOX2, PAX6,

TBXT, NES, TUBB3, and CDH2 (Table 4.3) overnight at 4°C and secondary antibodies

for 1 hour at room temperature. Nuclei were stained with a 1:10000 dilution of Hoechst

33342 (Thermo Fisher) included with secondary antibodies. Coverslips were mounted

with anti-fade mounting medium (ProlongGold, Life Technologies) and samples were

imaged on a Zeiss Axio Observer Z1 inverted microscope equipped with a Hamamatsu

ORCA-Flash 4.0 camera.

Table 4.3 Antibodies

Gene Target (Symbol)	Species	Company (cat. #)	Dilution
SRY-box 2 (SOX2)	mouse	Abcam (ab79351)	1:400
Paired Box 6 (PAX6)	rabbit	ThermoFisher Scientific (42-6600)	1:400
Brachyury (TBXT)	goat	ThermoFisher Scientific (PA5-46984)	1:400, 1:300
Nestin (NES)	mouse	Santa Cruz (SC-23927)	1:400, 1:200
Hoescht DNA stain	NA	ThermoFisher Scientific (62249)	1:10000
Tubulin Beta 3 Class III (TUBB3/TUJ1)	rabbit	Biolegend (802001)	1:500
N-cadherin (CDH2)	rabbit	Abcam (ab76057)	1:400
Collagen type IV	goat	EMD Millipore (AB769)	1:75
Tight junction protein ZO-1 (TJP1/ZO1)	Mouse	Invitrogen (33-9100)	1:200
Caudal-Type Homeobox 2 (CDX2)	Goat	R&D Systems (AF-1979)	1:100

Table 4.3 continued

Gene Target (Symbol)	Species	Company (cat. #)	Dilution
phospho-SMAD1/5 (pSMAD1/5)	Rabbit	Cell Signaling Technology (41D10)	1:800
Oligodendrocyte Transcription Factor-2 (OLIG2)	Rabbit	Abcam (ab9610)	1:400
Left-Right Determination Factor-1 (Lefty1)	Rabbit	ThermoFisher Scientific (PA5-19507)	1:1000
NODAL	Rabbit	ThermoFisher Scientific (PA5-23084)	1:200
Mesenchyme Homeobox MEOX1	Mouse	ThermoFisher Scientific (TA-804716)	1:400
Fibronectin	Mouse	Sigma-Aldrich (F6140)	1:400
Phospho-Histone H3 (Ser10)	Rabbit	Cell Signaling Technology mAb #3377	1:1600

4.2.7 Whole Mount Lightsheet Imaging

4% paraformaldehyde-fixed paraffin-embedded samples (see "Histology, Immunocytochemistry, and Imaging") were permeabilized with 1.5% Triton X-100 (Sigma-Aldrich) for 1 hour, blocked in 5% normal donkey serum (Jackson Immunoresearch) for 1 hour, and probed with primary and secondary antibodies (Table 4.3) overnight. Nuclei were stained with a 1:10000 dilution of Hoechst 33342 (Thermo Fisher) included with secondary antibodies. Samples were then embedded in 1.5% low melt agarose (BioReagent) and drawn up into ~1mm imaging capillaries and

subsequently imaged on the Zeiss Z.1 Light sheet Microscope equipped with two PCO.edge SCMOS cameras at 5X and 20X (NA 1.34, aqueous objective).

4.2.8 Flow cytometry

WTC and H1 cells were pretreated with either 2, 4, or 6 µM CHIR for 2 days, then dissociated from tissue culture plates with Accutase (STEMCELL Technologies) and washed with PBS. Similarly, the LBC-TBXT knockdown cells were pretreated with between 0 and 5 days of doxycycline concurrent with a final 2 days of 4 µM CHIR treatment, then dissociated and washed as described above. Cells were fixed for 20 minutes with 4% paraformaldehyde and washed 3x for 3 minutes with PBS. Samples were permeabilized in 0.5% Triton-X-100 (Sigma-Aldrich) for 30 minutes, then blocked in 1% normal donkey serum (Jackson Immunoresearch) for 1 hour, and probed with primary and secondary antibodies (Table 4.3) overnight. Samples were run on a LSR-II analyzer (BD Biosciences). Singlets were first identified by gating on forward scatter to side scatter ratio, then samples were gated into SOX2+/- and TBXT+/- using single stained controls for each cell line, then samples were assessed for percent SOX2+/TBXT+ cells. Analysis was conducted with a minimum of 20,000 events per sample.

4.2.9 Bulk RNA Sequencing Sample and Library Preparation

Whole organoids differentiated in either 0µM CHIR or 4µM CHIR at days 1, 3, 5, 7, and 10 of the differentiation protocol (n=3 per condition per day) were lysed with RPE buffer with 5 mM 2-mercaptoethanol, and RNA was extracted using the RNeasy Mini Kit (Qiagen) and quantified using the NanoDrop 2000c (ThermoFisher Scientific). RNA-seq

libraries were generated using the SMARTer Stranded Total RNA Sample Prep Kit (Takara Bio) and sequenced using NextSeq500/550 High Output v2.5 kit to a minimum depth of 25 million reads per sample. The sequences were aligned to GRCh37 using HiSat2 (Kim et al., 2015), reads were quantified using the featureCounts tool in the subread package (Liao et al., 2014), and differential expression between 0µM and 4µM CHIR-treated organoids was assessed at each day using the edgeR differential expression pipeline with limma/voom normalization (Law et al., 2014). Longitudinal differential expression was assessed for each CHIR condition using the "l" normalization algorithm (Zhang et al., 2019). Raw data is available at Geo under the accession number GSE155382.

4.2.10 Single Cell RNA Sequencing Sample and Library Preparation

Multiple organoid samples were combined and processed together using the MULTI-Seq technology (McGinnis et al., 2018). Organoids were singularized using Accutase (STEMCELL Technologies) and washed with cold PBS. Cells were resuspended in PBS with lipid-modified Anchor and Barcode oligonucleotides (gift from Zev Gartner, UCSF) and incubated on ice for 5 minutes. A co-Anchor oligo was then added in order to stabilize membrane retention of the barcodes incubated for an additional 5 minutes on ice. Excess lipid-modified oligos were quenched with 1% BSA in PBS, washed with cold 1% BSA solution, and counted using a Countess II FL (Life Technologies). Single cell GEMs and subsequent libraries were then prepared using the 10X Genomics Single Cell V2 protocol with an additional anchor specific primer during cDNA amplification to enrich barcode sequences. Short barcode sequences (approx.

65-100bp determined by Bioanalyzer) were purified from cDNA libraries with two sequential SPRI bead cleanups. Barcode library preparation was performed according to the KAPA HiFi Hotstart (Kapa Biosystems) protocol to functionalize with the P5 sequencing adapter and library-specific RPIX barcode. Purified ~173bp barcode fragments were isolated with another SPRI bead cleanup and validation by Bioanalyzer. Raw data is available at Geo under the accession number GSE155383.

The sample library was sequenced on an Illumina NovaSeq yielding an average of 41,112 reads per cell and 6,444 cells. The MULTI-Seq barcode library was sequenced on an Illumina NextSeq yielding an average of 9,882 reads per barcode and enabling sample assignment for 4,681 of 6,124 unique UMIs detected (76.4% recovery), using the demultiplexing code provided by the MULTI-Seq protocol (McGinnis et al., 2018).

4.2.11 Genome Annotation, RNA-seq Read Mapping, and Estimation of Gene and Isoform Expression

The sample library was aligned to the human GRCh38 reference genome using Cell Ranger v1.2.0 (10x Genomics). Gene expression levels were assessed using the Seurat v3.0.0 analysis pipeline (Butler et al., 2018). First cells were removed with fewer than 200 detected genes, fewer than 1,000 total detected transcripts, or which had greater than 10% mitochondrial gene expression. Next, expression levels were log normalized, and the top 2,000 variable genes calculated using the VST algorithm. The top 20 principal components were used to group cells into 2 clusters using a resolution of 0.3. Finally, top markers were detected for each cluster by detecting the top differentially expressed genes between both clusters, where at least 25% of cells in the

cluster expressed the gene and the gene was expressed at least 0.25 log2 fold-change different from the remaining population. Clusters and gene expression were visualized on a two-dimensional UMAP projection of the first 15 principal components.

4.2.12 Cluster Analysis

To assign cluster identity, the top markers for each cluster were tested for GO term enrichment using the biological process "enrichGO" function in the R package "clusterProfiler" v3.12 (Yu et al., 2012). In addition, differentiation maturity in each cluster was assessed by examining expression level of panels of early neuroectoderm markers, proliferation markers, markers of neuron fate commitment, and markers of cell types present in neural tube formation and axial extension (Tanabe and Jessell, 1996). Finally, to assess anterior-posterior position of each, panels of HOX genes were examined to assign rough position of each cluster along the head-tail axis (Bel-Vialar et al., 2002; Carpenter, 2002; del Corral and Storey, 2004).

4.2.13 Quantification of EdU and PH3 Localization

Sections stained for DAPI, EdU and ph3 were segmented by detecting the peaks of DAPI staining for each section using non-maximum suppression. The exterior of each section was then segmented by grouping DAPI+ cells into spatially contiguous clusters consisting of at least 5,000 pixels and which touch the image border in less than 1% of their total area. The exterior contour was then calculated for each region and fit to an elliptical model by total least squares (Halır and Flusser, 1998). All detected cells were projected onto the ellipse major axis, which was then normalized by total length. Where more detections fell on the left half of the major axis, the coordinates were reversed

such that the right half of each projected section always contained the majority of detections. Detected cells that corresponded to ph3 or EdU fluorescence at 20% or more above the background level outside of any section were counted as positive detections. A 15-bin histogram was then used to calculate percentage of projected cells for each day at each point along the semi-major axis, and then gaussian kernel density estimation was used to produce the empirical distribution. Linear and quadratic models were fit using ordinary least squares over the total population.

4.2.14 RNAScope

In situ hybridization for HOXB1, HOXC6, HOXB9 (probe information in Table 4.4) was performed on sections of 4% paraformaldehyde-fixed paraffin-embedded samples (see "Histology, Immunocytochemistry, and Imaging") using the RNAscope Multiplex Fluorescent Reagent Kit v2 (Advanced Cell Diagnostics) and following the protocol outlined in User Manual 323100-USM. Sections were imaged on a Zeiss Axio Observer Z1 inverted microscope equipped with a Hamamatsu ORCA-Flash 4.0 camera.

Table 4.4 RNAscope Probes

Gene Target (Symbol)	Channel
HOXB1	C2
HOXC6	C1
HOXB9	C3

4.2.15 Statistical Analysis

Each experiment was performed with at least three biological replicates. Multiple comparisons were used to compare multiple groups followed by unpaired T-tests (two tailed) between two groups subject to a post-hoc Bonferroni correction. In gene expression analysis, three replicates were used for each condition, and all gene expression was normalized to control wildtype populations followed by unpaired T-tests (two tailed). Significance was specified as P-values< 0.05 unless otherwise specified in figure legends. All error bars represent standard error of the mean (SEM) unless otherwise noted in the figure legend.

4.3 Results

4.3.1 Wnt Agonism Induces Emergence of Axial Extension of Neuronal Organoids

Because caudalization of the spinal cord and generation of NMPs are critically dependent on Wnt signaling (Henrique et al., 2015; Wilson et al., 2009; Yamaguchi, 2001), increased canonical Wnt signaling was paired with a previously-described hindbrain differentiation protocol (Butts et al., 2017) to attempt to produce spinal cord cell populations. Human induced pluripotent stem cells (hiPSCs) were pretreated for 48 hours with the Wnt small-molecule agonist CHIR99021 (CHIR; 4 µM) before organoid formation (**Figure 4.1 A,B**). Organoids were generated by aggregating singly-dissociated cells in non-adherent pyramidal inverted wells (3,000 cells/well) followed by rotary suspension culture in 6-well plates (Hookway et al., 2016). Aggregation was considered day 0 of differentiation (**Figure 4.1 A**). After 5 days in suspension culture,

pronounced singular extensions emerged from the CHIR-pretreated organoids (**Figure 4.1 C**). Elongated organoids were continuous, with linear striations extending longitudinally from a radially symmetric anterior aggregate. Histological analysis revealed multiple internal elongated epithelial compartments separated by regions devoid of cells (**Figure 4.1 C**). To ensure elongation was not due to organoid fusion in bulk culture, organoids were imaged continuously over 48 hours which revealed robust elongation in 83% of imaged wells as measured by axis ratios above 1.6 at the end of the time-course (**Figure 4.1 D, Figure 4.2 A-B**). Because NMPs can be generated *in vitro* in as little as 48 hours (Gouti et al., 2014), CHIR treated organoids were examined to determine if they generated a detectable population of NMPs immediately following aggregation. Indeed, SOX2(+) TBXT(+) NMPs were present in CHIR organoids at day 0, whereas in the absence of CHIR treatment, organoids lacked NMPs and did not extend (**Figure 4.3 A,B**). To ascertain whether extensions resulted from polarized populations of proliferating cells, EdU incorporation and phospho-histone H3 presence were measured on days 6, 7 and 9 of differentiation (**Figure 4.4 A**). A two-hour EdU pulse labeled >50% of cells in extending organoids, where proliferation was observed increasingly at the outer edges of extending aggregates, however polarized cell-division was not recapitulated when examining actively dividing cells via phospho-histone H3 (**Figure 4.4 B**). Furthermore, extensions occurred in organoids of various sizes (**Figure 4.5 B**) and extended more robustly when cultured at low density (**Figure 4.5 A**), suggesting that density-mediated signaling parameters, such as paracrine effects or nutrient availability, likely play a role in inducing and maintaining axial extensions,

similar to the documented role of glycolysis in vertebrate axial elongation(Oginuma et al., 2020). Overall, an organoid model was generated that robustly extends across a variety of culture conditions due to the early presence and size of SOX2(+) TBXT(+) NMPs that seem to account for the generation of the elongation phenotype.
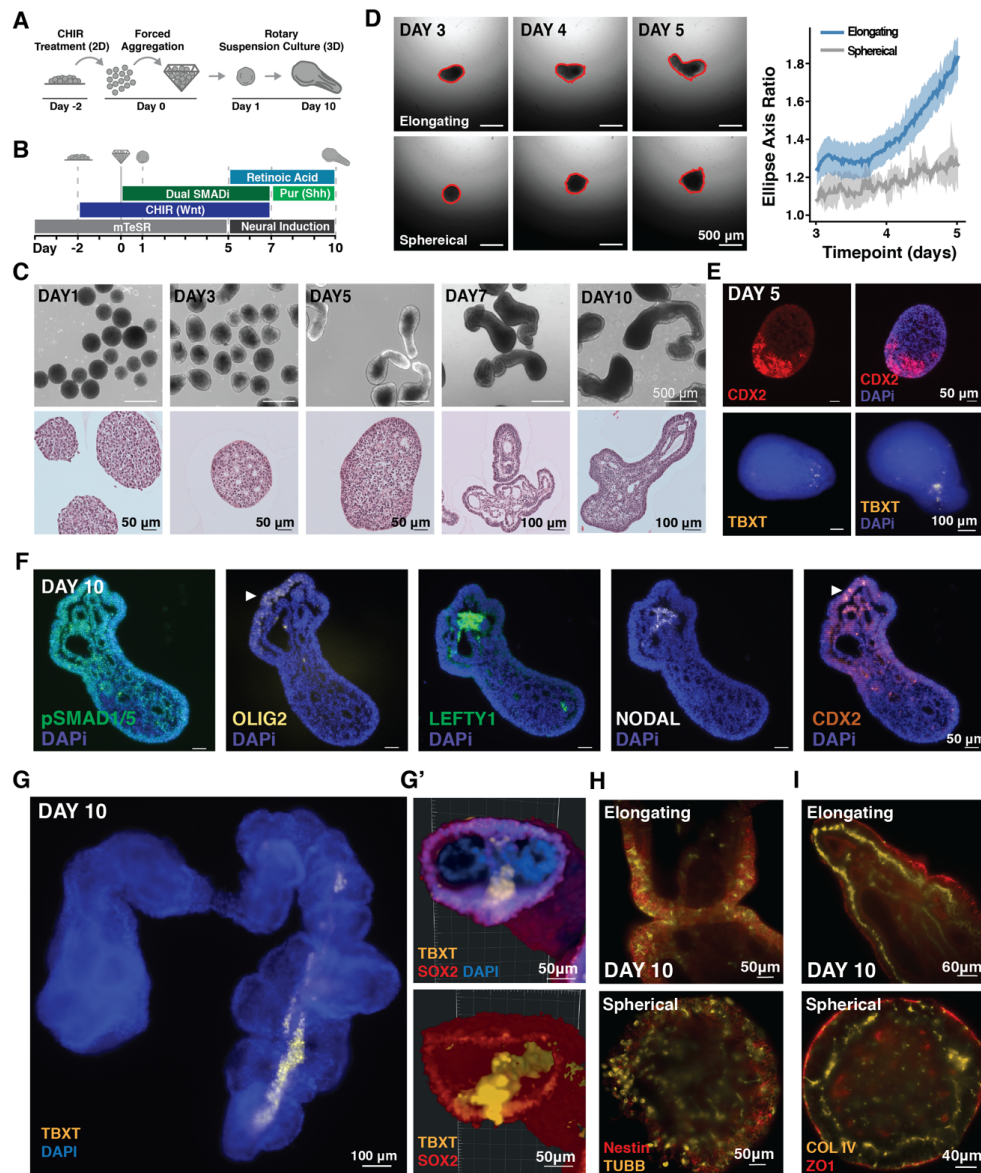
Figure 4.1 Spatiotemporal characterization of elongating organoids

A-B) Schematic of experimental set up and differentiation protocol. C) Brightfield and histological images of extending organoids. D) (LEFT) Frames from video time-course tracking organoid extensions in static 96 well plate culture where red outline defines measurements taken. (RIGHT) Quantification of ellipse axis ratios from day 5 to 7 of extending (blue) and non-extending (grey) organoids. Solid lines represent the mean ratio (dark color) with 95 percent confidence interval (light color); n= 2 non-elongating, n= 15 elongating. E) Immunofluorescence images at day 5 of differentiation from paraffin sections in extending organoids examining CDX2 and TBXT localization. F) Immunofluorescence images at day 10 of differentiation from paraffin sections in extending organoids examining axis and developmental patterning markers. G) Fluorescence image of TBXT streak traveling down the length of organoid G') 3D reconstruction cross section of the posterior of extending organoids based on light-sheet microscopy displaying TBXT+ streak. H-I) Optical sections from light-sheet microscopy of elongating (top) and spherical (bottom) organoids stained for Nestin and bIII-Tubulin (TUBB) or COL IV and ZO1.

126
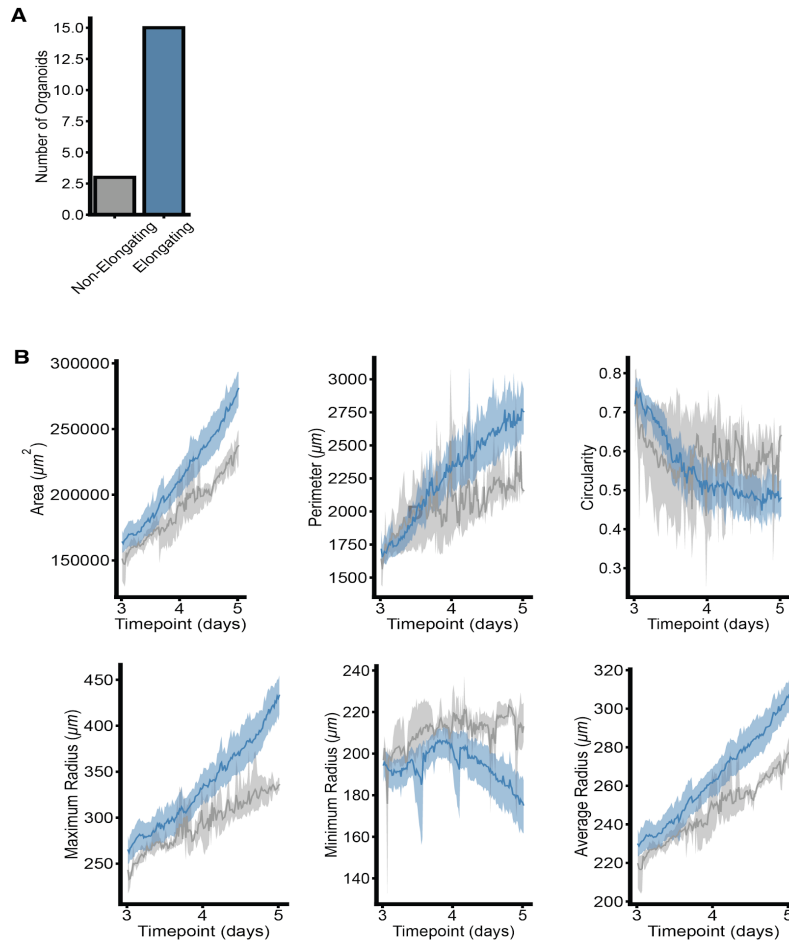
Figure 4.2 Time lapse quantification of organoid elongation
A) Number of extending verse non-extending aggregates in 4uM CHIR conditions (n=18). B) Quantification of area, perimeter, circularity, maximum, minimum, and average radius lengths of extending (blue) and non-extending (grey) aggregate videos. (mean value (dark line) with 95 percent confidence interval (light color)).
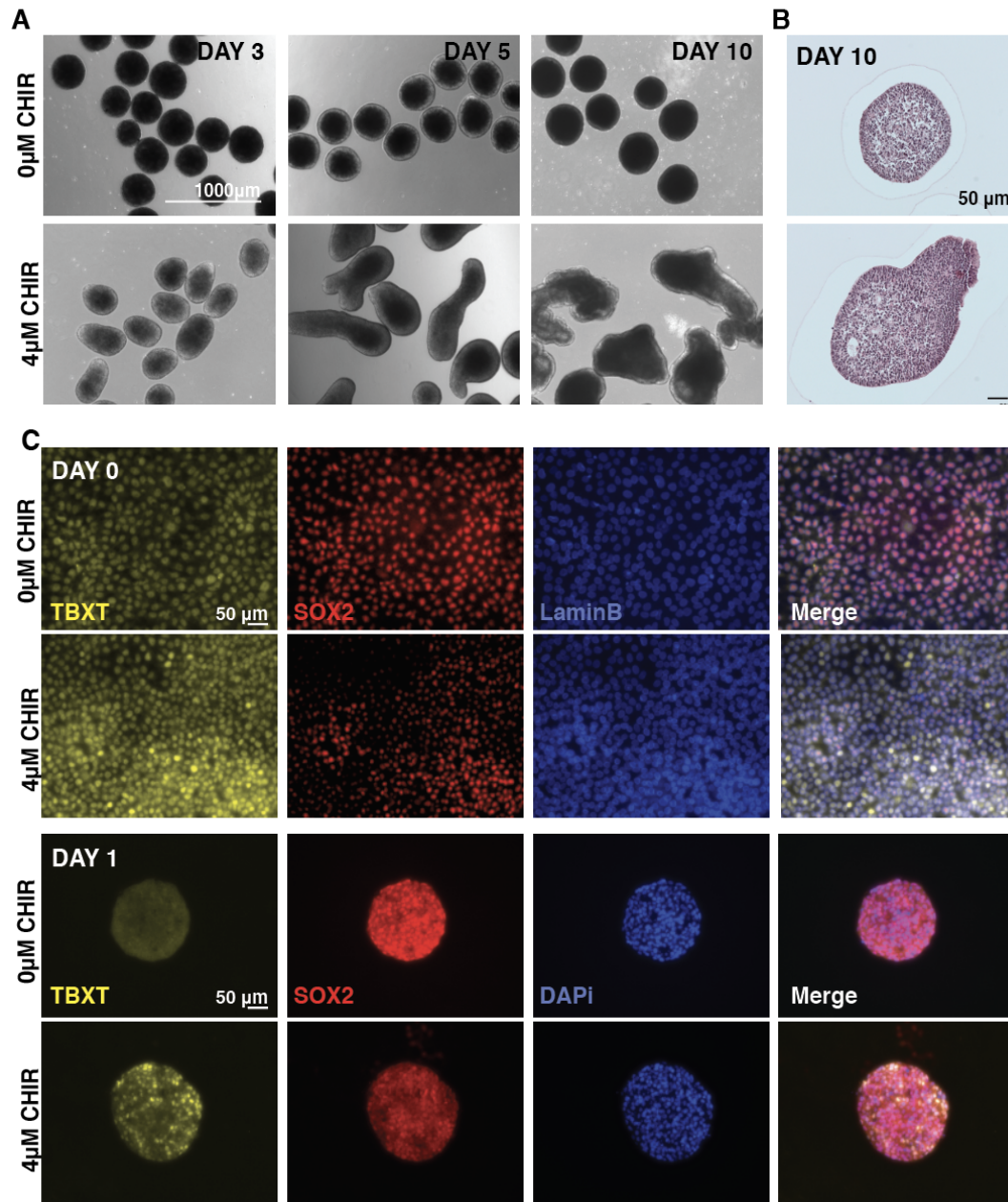
Figure 4.3 CHIR treatment results in extensions and increase of TBXT expression
A) Stereoscope images of differentiation performed with and without CHIR in the WTC hiPSC line. B) Histological sections of WTC organoids exposed to 0uM CHIR and 4uM CHIR C) Immunofluorescence images of TBXT and SOX2 in CHIR-treated and non-treated organoids.
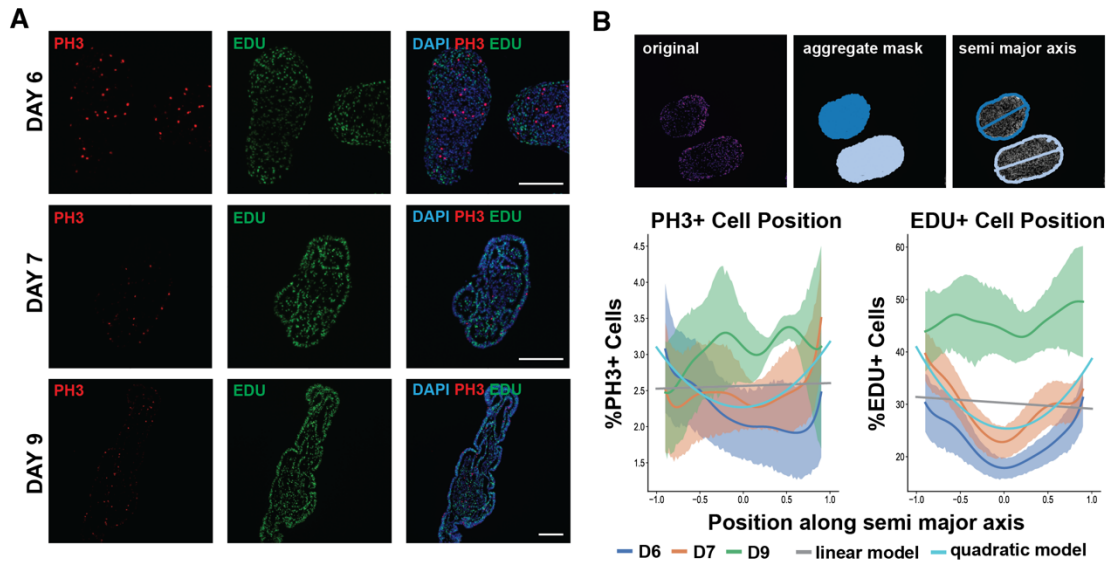
Figure 4.4 Cell proliferation occurs throughout the extending organoid
A) Immunofluorescence images of EdU incorporation and phospho-histone 3 (PH3) localization in extending organoids taken from paraffin sections. B) TOP: Segmentation of organoid sections showing LEFT) ph3 staining, MIDDLE) segmentation of nuclei by DAPI staining, RIGHT) a total least squares fit of the segmentation contour with the major axis superimposed. BOTTOM: Lineplots for PH3 or EdU expression along the normalized semi-major axis of aggregates at day 6, 7, and 9. Linear and quadratic least squares fits of PH3 and EdU expression along the normalized semi-major axis for all days.
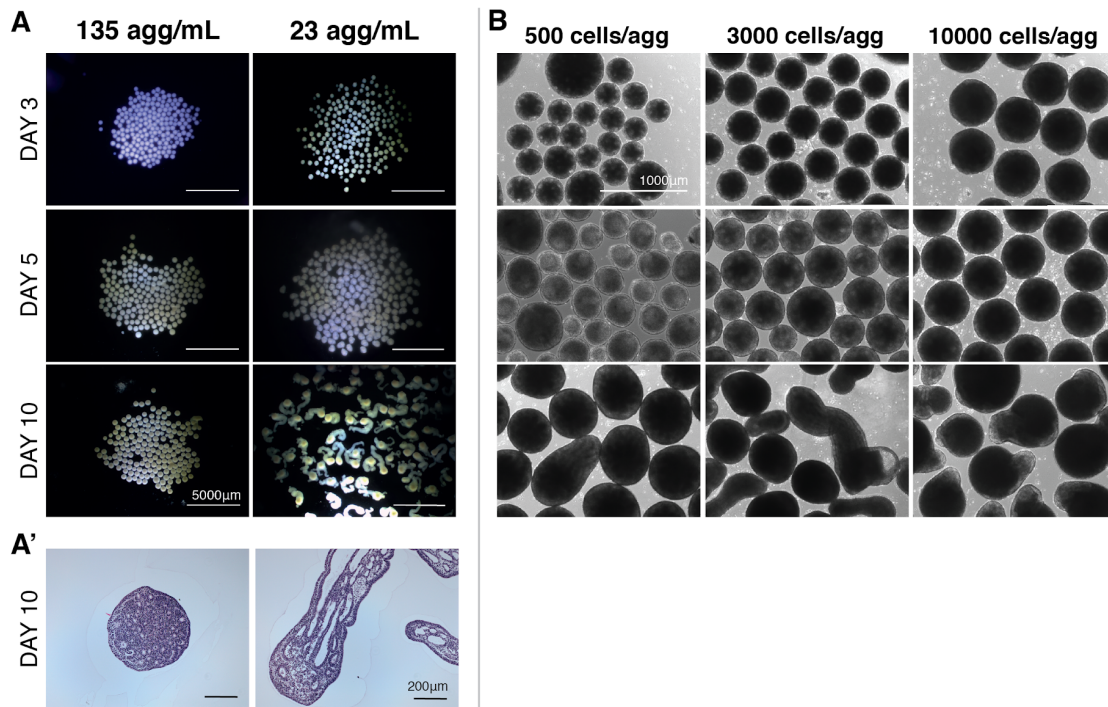
Figure 4.5 Organoid culture density changes morphology of extensions
A) Stereoscope images of organoids grown at high (135 aggregates/ml) or low (23 aggregates per ml) densities. A') Histological stains on paraffin sections of organoids at day 10 of differentiation in high (left) vs low (right) density culture. B) Differentiations conducted in the WTC hiPSC line starting with different aggregate sizes.

4.3.2 Organoids Display Markers of Tissue and Cellular Polarity

Next, it was determined whether extending organoids generated an anterior-posterior (AP) axis. The posteriorly expressed gene CDX2 (Beck et al., 1995) was detected in a polarized manner from day 5 to day 10 of differentiation (**Figure 4.1 E,F**). Furthermore, within extending organoids, clusters of NODAL- and LEFTY-positive cells were observed at day 10, suggesting the presence of an organizer-like population (**Figure 4.1 F, Figure 4.6 B**). In fact, multiple clusters of NODAL-positive cells were present in organoids with multiple extensions (**Figure 4.6 A**). In day 10 extending organoids, compartments of cells expressing the ventral spinal cord marker, OLIG2,

130

lined the epithelial cysts, and a polarization of pSMAD1/5 expression was detected

across the organoid (**Figure 4.1 F**). These observations suggest that the extending

organoids exhibit characteristics of the emergent populations that establish the

embryo's anterior-posterior axis.
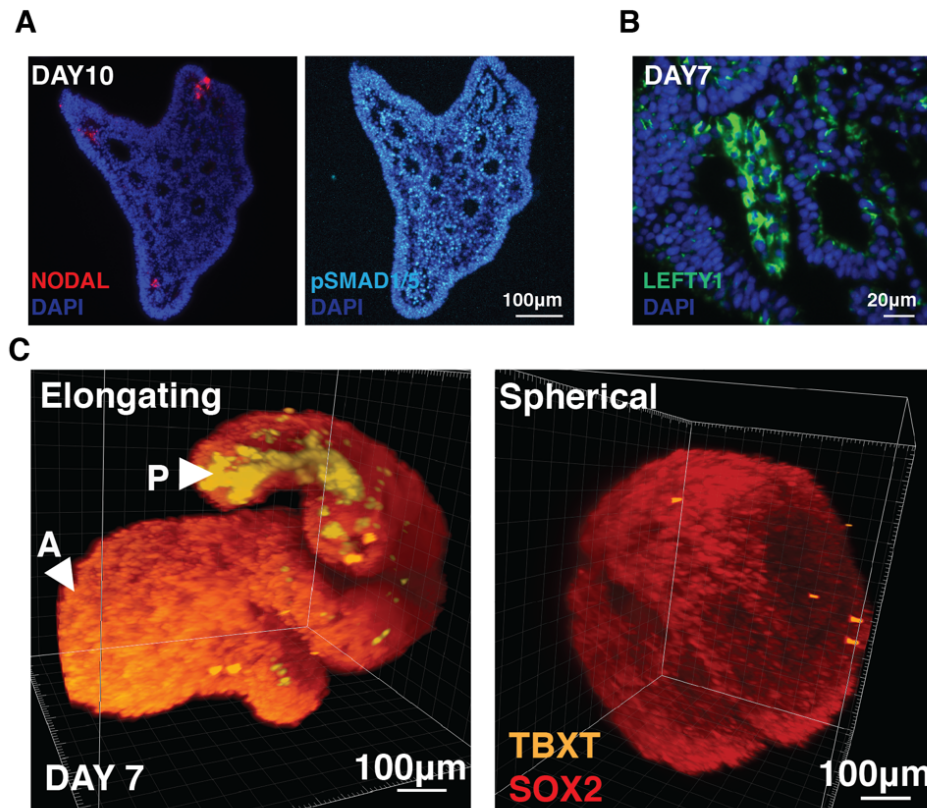
Figure 4.6 Gene expression distribution within extensions.
A) Immunofluorescence staining of an extending organoid at day 10 of differentiation for NODAL and pSMAD1/5. B) Immunofluorescence staining of an extending organoid at day 7 of differentiation for LEFTY1. C) 3D reconstructions of light-sheet microscopy images of day 7 extending and non-extending organoids stained for TBXT and SOX2.

To understand how differences in the identity and spatial distribution of NMP

populations in extending organoids contributed to differential organoid morphologies,

extending and non-extending organoids were imaged via light-sheet microscopy at day

7 of differentiation (**Figure 4.1 G-G', Figure 4.6 C**). In most extending organoids, the

TBXT(+) population formed a streak along the extension and accumulated at the end of the extension, reminiscent of the notochord and previous gastruloid reports (Beccari et al., 2018). Additionally, regions of SOX2(+)TBXT(+) NMPs were adjacent to the TBXT(+) streak (**Figure 4.1 G**). In contrast, TBXT(+) cells in non-extending organoids were sparsely distributed across the surface of the spherical organoid with no clearly identifiable regions of SOX2(+)TBXT(+) co-expression. At day 10, both extending and non-extending organoids contained cells expressing the neuronal progenitor markers Nestin and beta 3-tubulin, indicating that the majority of cells differentiated to a neural fate (**Figure 4.1 H, Figure 4.7 B**). However, in the extending organoids, the neural-committed cells organized into distinct compartmentalized layers surrounding lumens within the extensions.
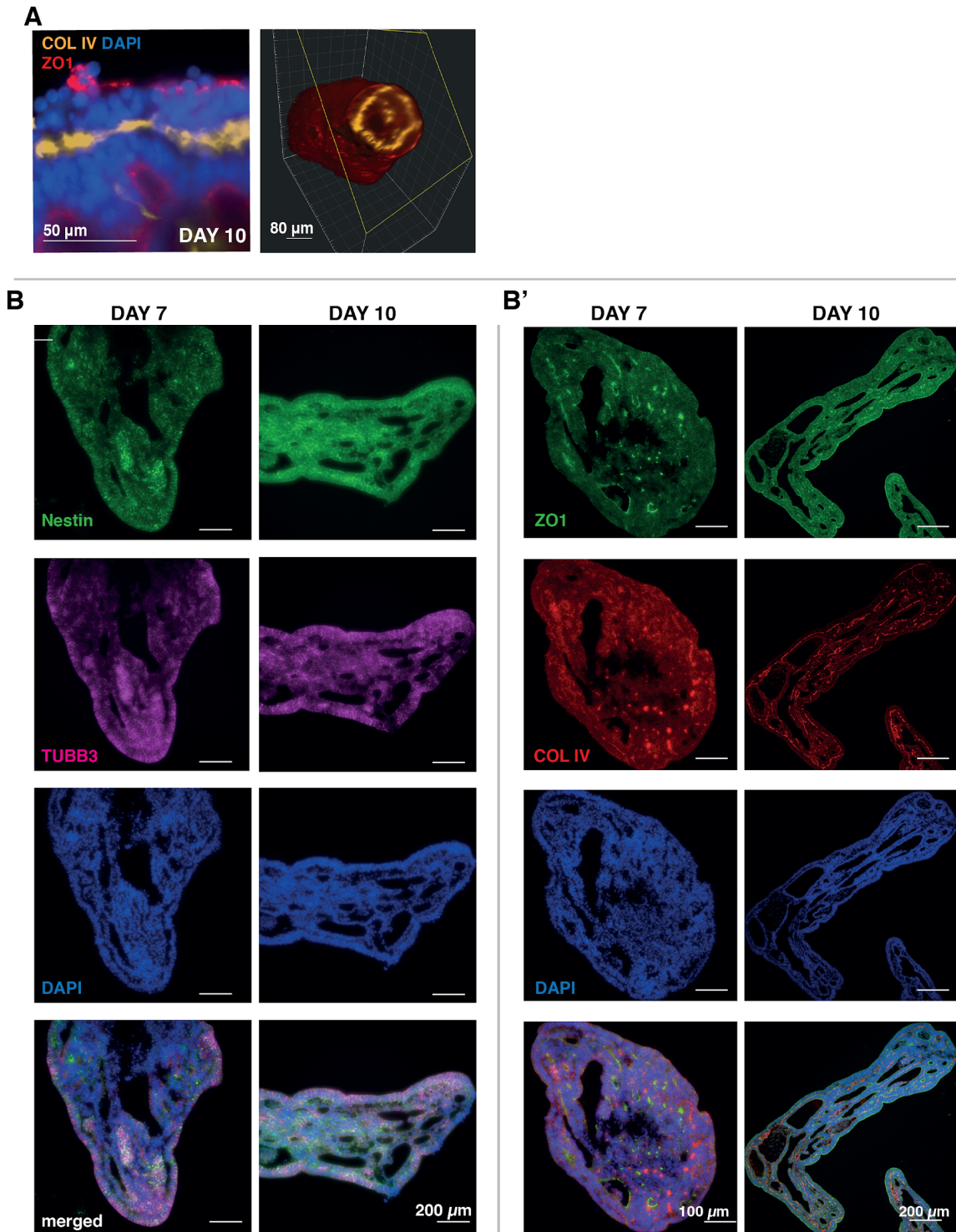
Figure 4.7 Apical-Basal polarity in extending organoids over time
A) Optical section from light-sheet imaging of extending stained for COL IV and ZO1 (left). 3D reconstruction showing internal ring of basement membrane within the posterior extensions (right). B) Immunofluorescence of paraffin sections of extending organoids at day 7 and day 10 of differentiation. B') Immunostaining of paraffin sections of extending organoids at day 7 and 10 of differentiation for markers of basement membrane (COL IV) and tight junctions (ZO1).

Cell polarity was examined by staining for collagen IV (COL IV), a marker of the basement membrane, and zona occluden-1 (ZO-1), a marker of tight junctions at cellular apical domains (**Figure 4.1 I, Figure 4.7 A**). An increase in basement membrane deposition and cell polarization was observed between days 7 and 10 of differentiation in extending organoids (**Figure 4.7 B'**). After 10 days, ZO-1 was apparent at the apical domain of cells in the outer epithelial layers of the extensions, as well as in cells facing the internal lumens. Distinct layers of basement membrane marked by collagen IV were observed basally to ZO1, forming a bilayer between lumens and organoid exterior (**Figure 4.1 I, Figure 4.7 A,B'**). Bilayers of cells were also apparent in non-extending organoids, but to a lesser extent since the internal lumens were much smaller. Overall, these patterns indicate that extending organoids undergo tissue polarization via axial extensions that consist of polarized sheets of neuroepithelial cells that generate organized basement lamina.

4.3.3 Organoids Display Regionalized HOX Patterning

To examine transcriptional differences between the two poles of extending organoids, the "anterior", or main organoid body was manually dissected from the "posterior", or extending portion, of day 9 organoids (**Figure 4.8 A,B**) and compared via qPCR. Of the genes tested, only HOXA11, a marker of the lumbar region, displayed a significant difference in expression between the anterior and posterior fragments, likely because of high variability between replicates (**Figure 4.8 C**). To gain a more nuanced understanding of HOX expression in extending organoids, transcripts marking hindbrain

(HOXB1), brachial (HOXC6), and thoracic (HOXB9) regions of the neural tube were examined via RNAscope at days 7 and 10 (**Figure 4.9 A, 4.10 A**). While non-extending organoids expressed both hindbrain and brachial HOX genes, their expression was distributed radially, and the expression of HOXB9 was delayed relative to extending organoids (**Figure 4.10 B**). In contrast, HOXB1 (hindbrain) was enriched in the central mass of extending organoids, and HOXC6 (brachial) and HOXB9 (thoracic) were enriched in the extensions (**Figure 4.9 A**). Interestingly, HOXC6 and HOXB9 often overlapped at day 7 within extending organoids. However, by day 10 HOXB9 was more pronounced than HOXC6 in the extension regions, suggesting a transition to a more posterior fate. These data indicate that while organoid extension is not a requirement for posterior HOX expression, extension enables the stratification of distinct HOX domains.
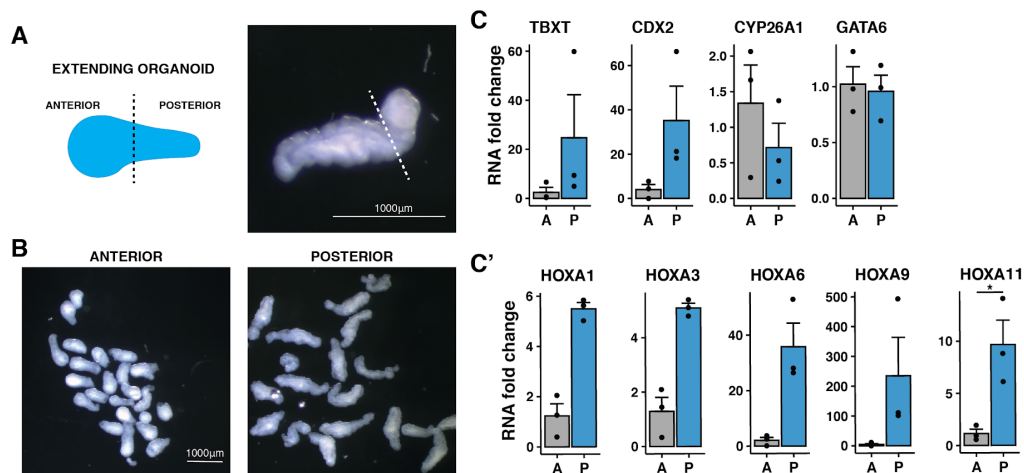
Figure 4.8 Dissections of extending organoids to isolate anterior and posterior regions.
A) Schematic and example of anterior vs posterior regions. B) Stereoscope images of organoids after isolation of anterior or posterior region. C-C') qPCR data of HOX genes and region-specific genes in anterior vs. posterior of organoids ($p < 0.05$).
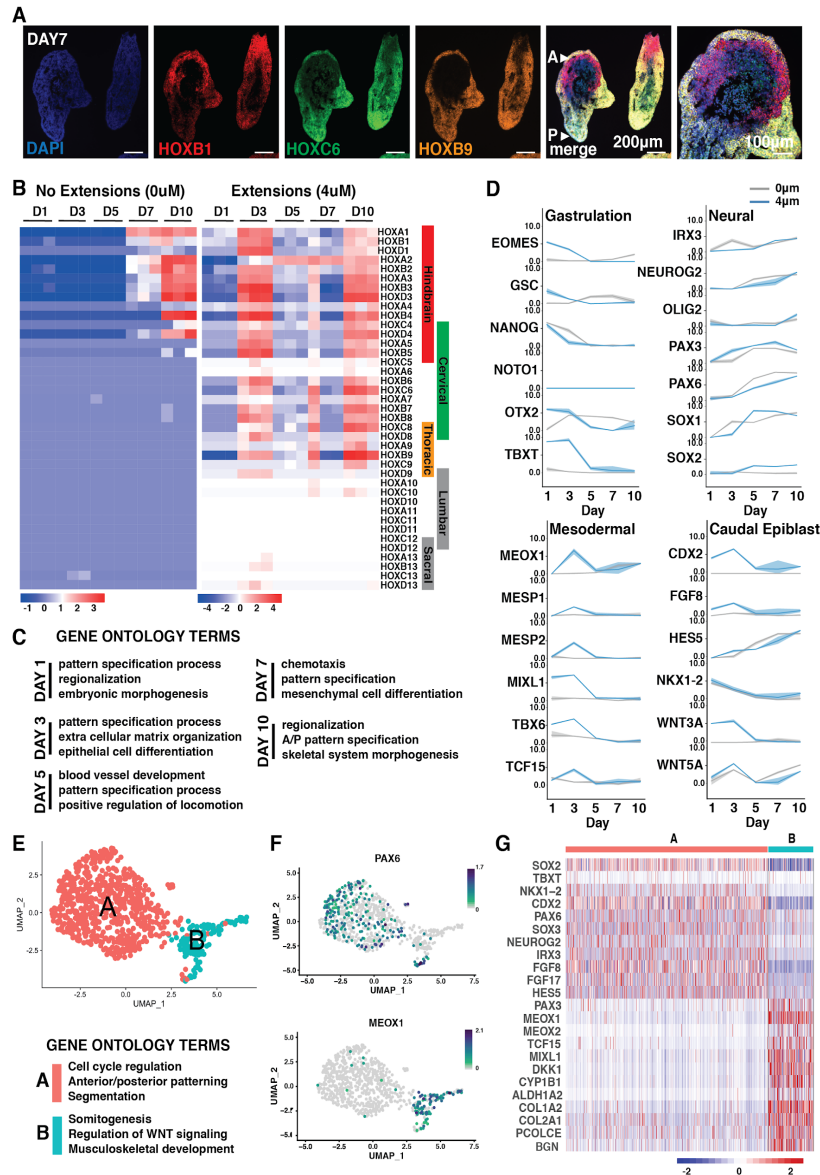
Figure 4.9 Gene expression differences within extending organoids.
A) RNAscope of sectioned organoids cultured at low density with probes for HOX genes marking different regions of the spine at day 7 of differentiation (arrows mark anterior (A) and posterior (P) axis of organoids). B) Heatmap of HOX gene expression in organoids with and without CHIR treatment from bulk RNAseq C) Line graphs of gene expression change in markers associated with Gastrulation, Neural, Mesodermal, and Caudal Epiblast fates (grey line indicates 0uM CHIR and blue line indicates 4uM CHIR; solid line indicates mean, shading represents 95% confidence bounds; axis displayed are $\log_2$ counts). D) Significant ($p < 0.05$) Gene Ontology classifications derived from the most differentially upregulated genes in extending organoids from bulk RNAseq. E) UMAP of n=789 cells from extending organoids exposed to a lower concentration of CHIR (2µM) at day 10 of differentiation from single-cell RNAseq. Gene Ontology terms assigned to identified clusters. F) UMAPs showing cells expressing PAX6 and MEOX1. Color scale indicates a normalized increase in $\log_2$ fold change from min expression to max expression of the respective gene. G) Heat map showing normalized expression levels of genes associated with spinal neuron specification, caudalizing factors, and mesoderm specification from single-cell RNAseq.
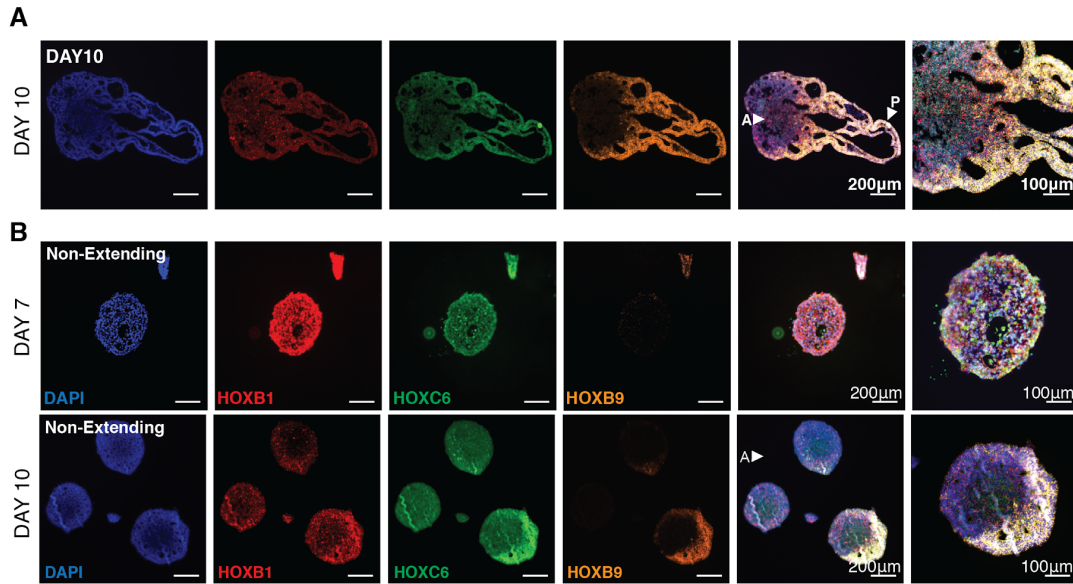
Figure 4.10 Limited segregation of HOX genes in non-extending organoids
A) RNAscope of sectioned extending organoids with probes for HOX genes marking different regions
of the spine at day 10 of differentiation. B) RNAscope of sectioned non-extending organoids with
probes for HOX genes marking different regions of the spine at day 7 and 10 of differentiation.

4.3.4 Developmental Timing and Maturation of Organoids

To obtain a more detailed picture of HOX gene expression over time, bulk RNA

sequencing was performed across 5 timepoints, comparing organoids with and without

CHIR treatment. As previously observed, timing and extent of HOX expression varied

between the two conditions. Non-extending organoids began to express detectable

HOX transcripts at Day 7 and adopted a hindbrain identity upon the experiment's

conclusion at day 10 (**Figure 4.9 B**). In contrast, extending organoids expressed both

hindbrain and cervical HOX transcripts as early as day 3, with a notable upregulation in

thoracic HOX genes by day 10. This, in conjunction with the RNA-scope data, suggests

that the spatiotemporal onset of HOX expression in extending organoids recapitulates

the progression of anterior-to-posterior elongation of *in vivo* development.

Analysis of the top divergent genes between the extending and non-extending

organoids (**Figure 4.11 A**) revealed unique components of pattern specification in the

extending condition, as indicated by Gene Ontology (**Figure 4.9 C**). Day 1 extending

organoids were enriched for genes involved in pattern specification and embryonic

morphogenesis, and by day 10 included hallmarks of anterior-posterior pattern

specification. Extending organoids upregulated WNT3 and WNT5B signaling at day 1,

whereas the non-extending organoids upregulated WNT5B at day 5 (**Figure 4.9 D, 4.11

B**). Extending aggregates were enriched for the caudal marker CDX2, while non-

extending organoids were enriched for the forebrain marker OTX2 at later timepoints

(**Figure 4.9 D, 4.11 A**). Further, elongating organoids passed through a goosecoid

(GSC) and TBXT-high stage at day 1 and more specific lineage markers arose at day 3.

Finally, neural markers remained high as the differentiation progressed. Together, this

analysis indicates that early upregulation of Wnt signaling induces pattern specification

in organoids that mirrors many aspects of *in vivo* morphogenesis and favors caudal

neurogenesis.
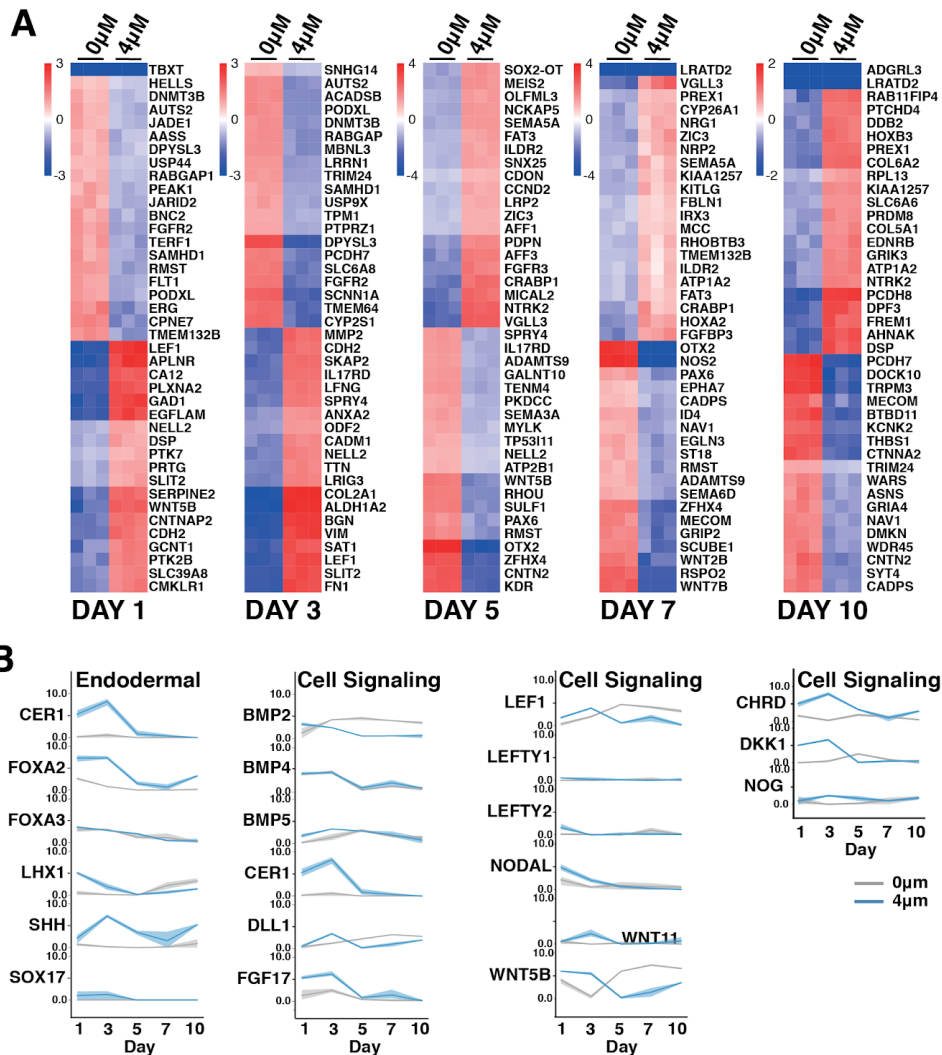
Figure 4.11 Bulk RNA sequencing of organoids with or without CHIR treatment.
A) Top differentially expressed genes. B) Line plots showing gene expression of genes associated with signaling morphogens (grey line: 0μM CHIR; blue line: 4μM CHIR).

4.3.5 Extending Organoids Consist Exclusively of Neural and Mesodermal Lineages

To examine cell diversity in extended organoids, single-cell RNA sequencing was performed on extending day 10 organoids. The single cell transcriptomic data recapitulated HOX gene expression profiles previously observed within extending organoids (**Figure 4.12 A**). Shared nearest neighbor computation revealed two clusters

139

(**Figure 4.9 E**); one (cluster A) largely distinguished by cell cycle regulation and the other (cluster B) by somitogenesis, according to Gene Ontology. Neural markers, such as SOX2, were expressed in both clusters (**Figure 4.12 B,C**), but, cluster A had higher expression levels of genes associated with neural tube fates (SOX2, IRX3, PAX6, NEUROG2) and the caudal lateral epiblast (CDX2, NKX1.2 and FGF8) (**Figure 4.9 F,G**). In contrast, MIXL1, MEOX1, CYP1B1, MEOX2, and PAX3 were almost exclusively expressed in cluster B, indicating that this cluster contained mesodermal cells (**Figure 4.9 F,G; 4.12 B,C**). Expression of endoderm-specific markers such as SOX17 was not observed, and only very low expression of FOXA1 and PAX9 in <1% of cells (**Figure 4.12 C**), indicating that, in contrast to previously reported gastruloid models(Beccari et al., 2018; van den Brink et al., 2020; Moris et al., 2020), these extending organoids consisted entirely of neural and mesodermal cell types. Overall, these data indicate that extending organoids transcriptionally mimic the caudal epiblast and contain cell populations associated with both mesoderm and neural tube, similar to the embryo during axial extension.
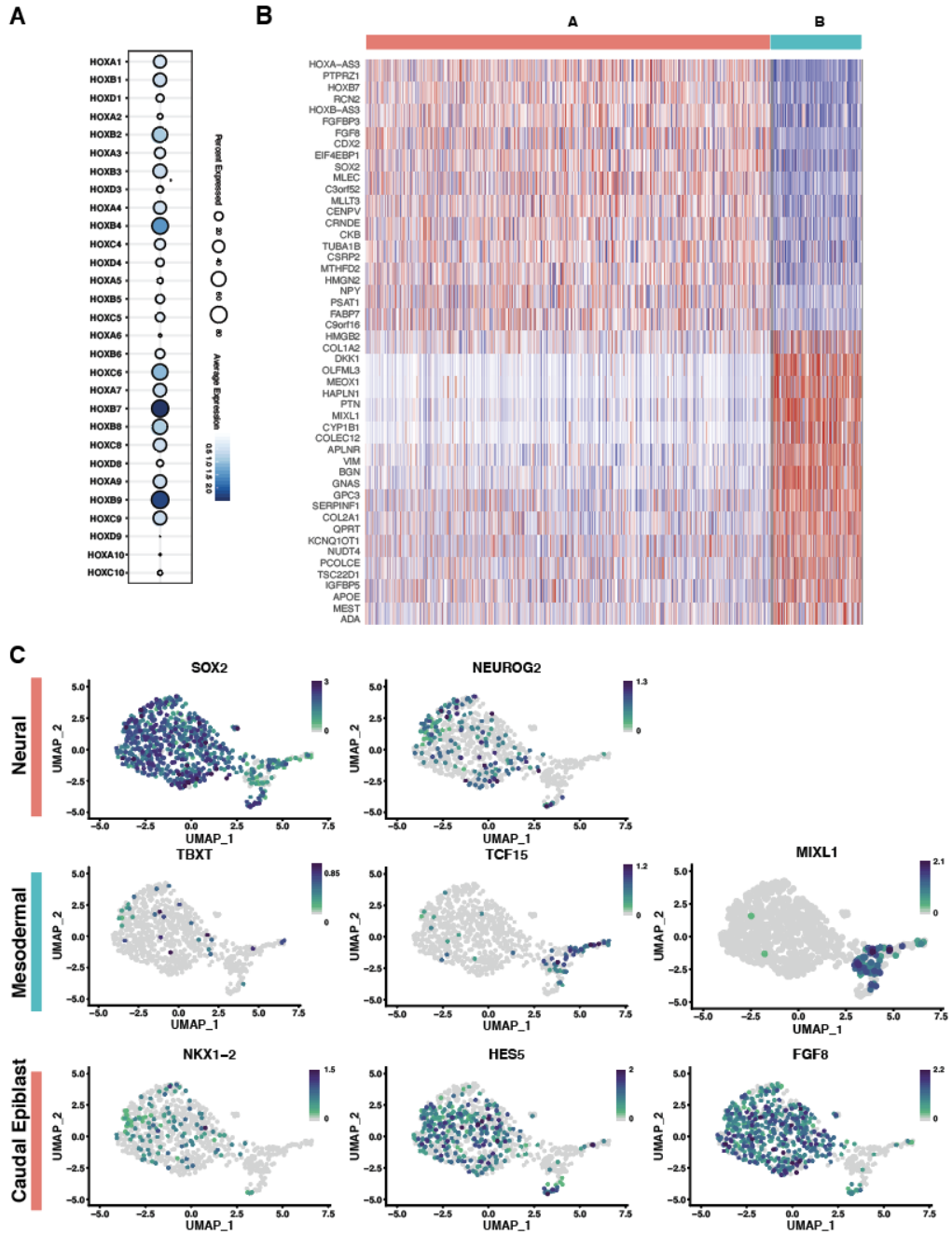
Figure 4.12 Single cell gene expression within extending organoids
A) Heat map of top differentially expressed genes between cluster A and B. B) Gene expression distribution plots. C) Dot plot of HOX gene expression within organoids.

4.3.6 Wnt Signaling Influences Extension and Axis Identity Across Stem Cell Lines

To determine the robustness of the effect of WNT signaling on axial elongation, two iPSC lines (WTC and WTB) and two hESC lines (H1 and H7) were differentiated into neural organoids in the presence of varying amounts of CHIR. The threshold of CHIR required to reliably generate extensions varied between lines (**Figure 4.13 A**), highlighting cell-line-specific responsivity to Wnt agonism. Whereas the WTC line produced elongations at a 2-4µM dose of CHIR, all other lines tested only began to elongate when the CHIR dose was increased to 6µM. Further, increasing CHIR dosage increased both the degree of extension (**Figure 4.13 B,C',D'**) and the frequency of SOX2(+)TBXT(+) cells at the start of the differentiation (**Figure 4.13 C,D**). Interestingly, the WTC line demonstrated a maximum threshold for CHIR-induced extension, as 6µM CHIR treatment reduced extensions (**Figure 4.13 A,C**). Importantly, all cell lines displayed polarized organization with internal epithelialization and cavitation at their respective condition that permitted extension (**Figure 4.14 A**). To further characterize this polarization, we found that the H1 ESC line displayed similar regionalized expression of axial markers as seen in the WTC iPSC line (**Figure 4.14 B**).
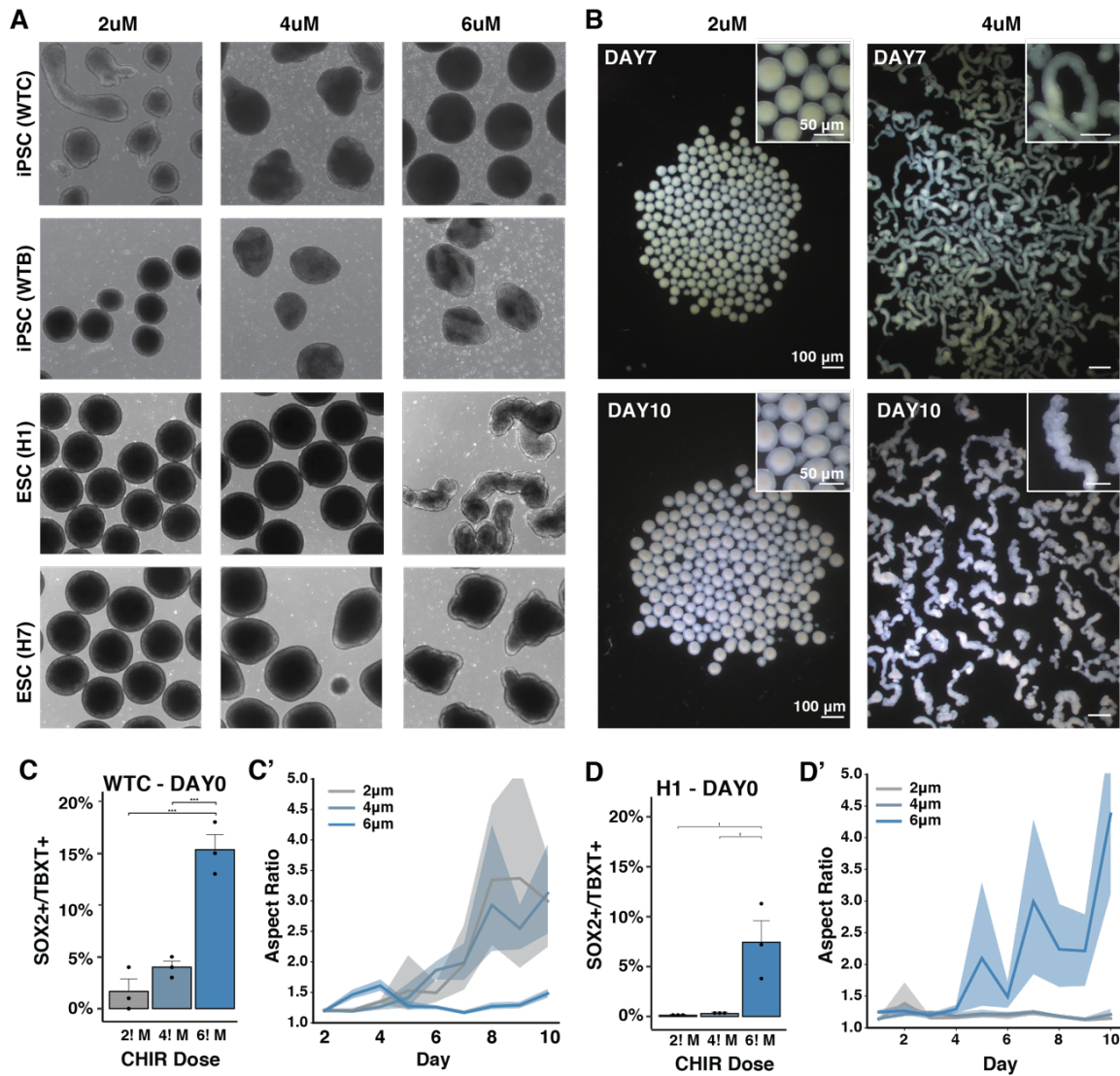
Figure 4.13 Wnt mediated increase in extensions across stem cell lines.
A) Brightfield images at day 7 of differentiation with increased CHIR doses conducted in the hiPSC and ESC lines. B) Stereoscope images of extending and non extending organoids in two different CHIR doses. C) Quantification of SOX2(+)TBXT(+) cells at increasing CHIR doses by FLOW cytometry in the WTC hiPSC or H1 ESC line at day 0. D) Quantification of the length of extensions in WTC or H1 organoids shown as radius ratio (ie. major to minor) with increasing doses of CHIR. (Solid line indicates mean, shading represents 95% confidence bounds).
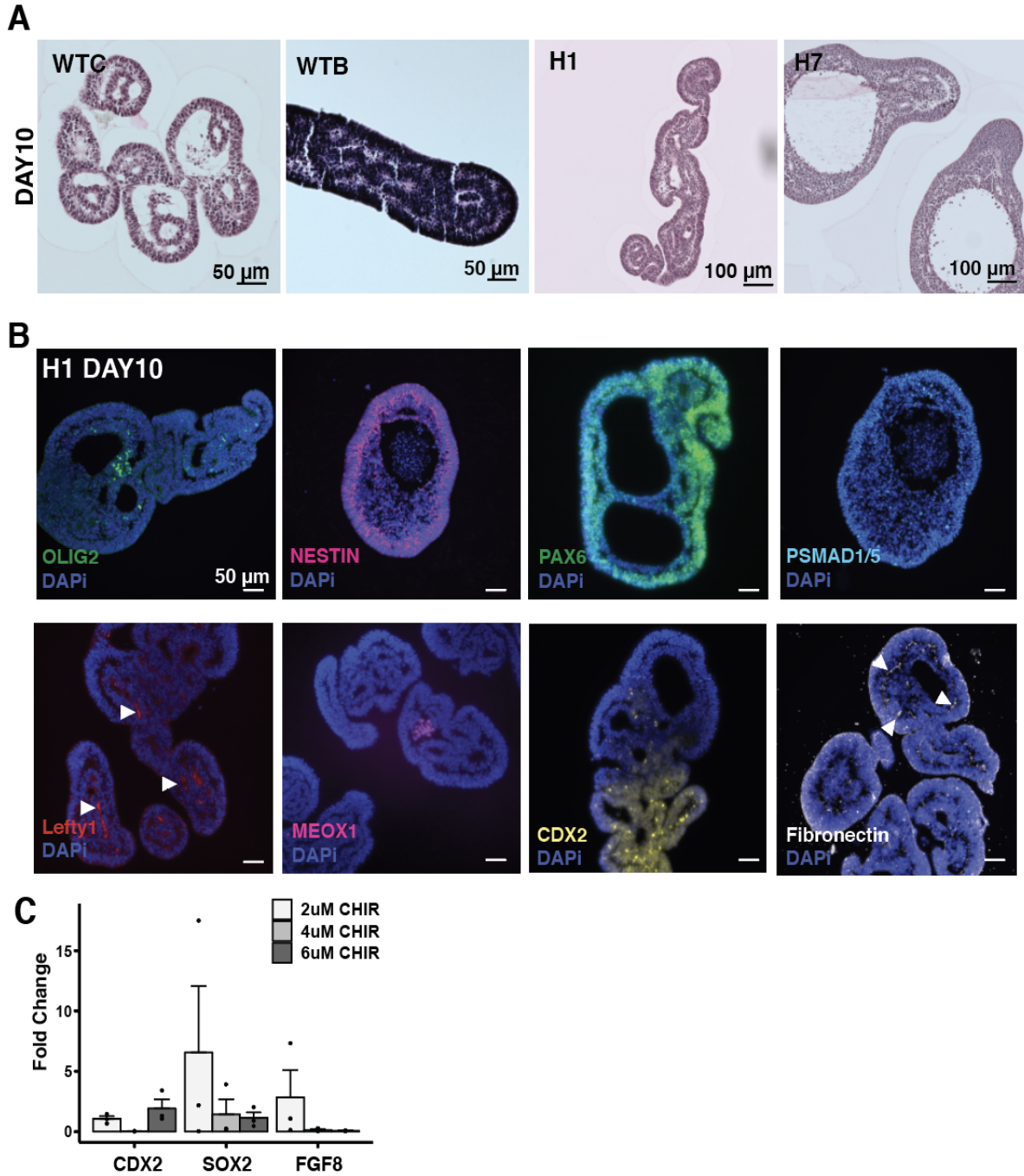
Figure 4.14 Extension reproducibility across hiPSC and hESC lines
A) Histology of the WTC and WTB hiPSC lines and H1 and H7 ESC line at extending (6uM) CHIR concentrations at day 10 of differentiation. B) Immunofluorescence staining of paraffin sections for OLIG2, Nestin, PAX6, pSMAD1/5, LEFTY1, MEOX1, CDX2, and Fibronectin at day 10 in H1 organoids (arrows indicate regions of positive staining). C) qPCR of caudal epiblast markers in the H1 ESC line across CHIR doses.

To examine differences in expression across Wnt conditions, sections of WTC organoids were stained at day 10 treated with 2μM-, 4μM- or 6μM-CHIR (**Figure 4.15**). Organoids exposed to 4μM and 6μM CHIR expressed SOX2 and PAX6 in distinct regions of the organoids, whereas organoids exposed to 2μM CHIR displayed SOX2 and PAX6 expression throughout, suggesting a non-neuronal population emerging in higher Wnt signaling environments. All organoids displayed N-cadherin and beta III-tubulin expression, indicating the emergence of maturing populations of neurons; however, regions without N-cadherin or beta III-tubulin were more prevalent in organoids exposed to higher doses of CHIR, suggesting a Wnt-dependent reduction in the overall neural population. Overall, the level of Wnt pathway activation that allows for robust emergence of an NMP progenitor pool and organoid extension varies across both hiPSC and ESC lines, as commonly reported for most differentiation protocols (Ortmann and Vallier, 2017).
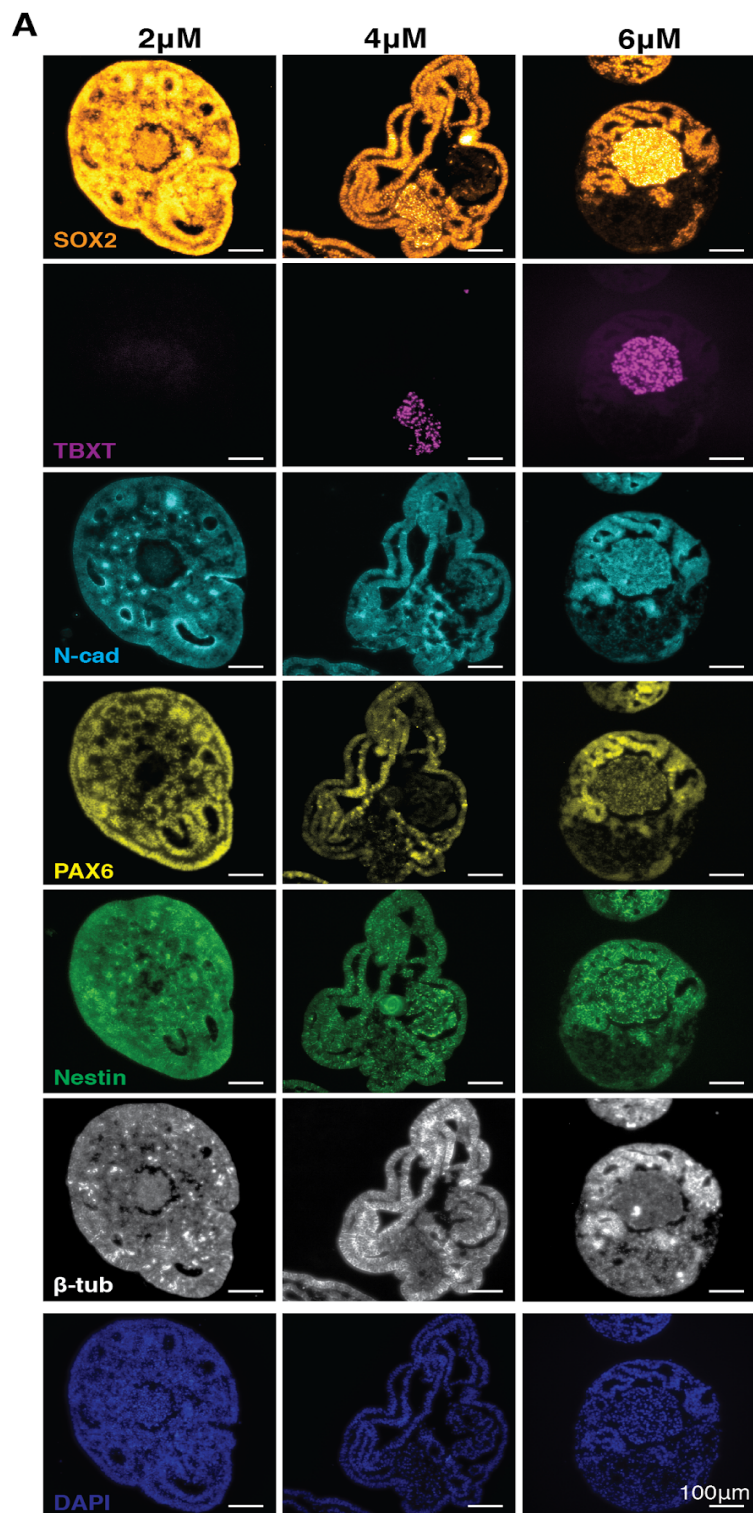
Figure 4.15 Neural cellular identities are maintained across CHIR dosage
A) Immunofluorescence of paraffin sectioned WTC hiPSC organoids at day 10 of differentiation from 2uM, 4uM, and 6uM CHIR concentrations.

4.3.7 Wnt and BMP Signaling Influences Axis Identity and Degree of Extension

As the body axis extends, gradients of key signaling molecules are set up across the embryo. Initially, Wnt signaling is activated throughout the epiblast by extraembryonic BMP and later localized to the posterior embryo where Wnt directly upregulates the expression of TBXT, which in turn upregulates both Wnt and FGF, reinforcing and maintaining their posterior to anterior gradients (Amin et al., 2016). The importance of these signaling relationships for proper axial elongation are highlighted in mouse knockout models. TBXT knockout mice have severely disrupted trunk morphogenesis, failing to generate the notochord and posterior somites (Chesley, 1935; Herrmann et al., 1990; Pennimpede et al., 2012; Wilson and Beddington, 1997; Wilson et al., 1993). When the dorsal inhibitor of BMP, Noggin, is knocked out, murine models have increased posterior BMP signaling, failure of neural tube closure, loss of robust dorsal-ventral patterning, and elongation of the developing tail (McMahon et al., 1998). Deletion of the BMP shuttling molecule, Chordin, results in a shortened body axis, a ventralized phenotype, and underdeveloped anterior spine (Bachiller, 2003). Thus, to interrogate the role of these signaling pathways in a human system, TBXT, Chordin, or Noggin were independently knocked down via an inducible CRISPR interference system in human iPSCs (Larson et al., 2013; Libby et al., 2018; Mandegar et al., 2016). First, RNA guides targeting TBXT were integrated into a Lamin-B GFP-labeled WTC hiPSC line harboring a doxycycline (DOX) inducible CRISPR interference system (CRISPRi)(Libby et al., 2018; Mandegar et al., 2016). Media supplemented with DOX for five days prior to aggregation reduced the number of detected TBXT(+) cells by 90%

(**Figure 4.16 A,B**), and the frequency of TBXT(+)SOX2(+) cells was also significantly

reduced (**Figure 4.16 C**). Interestingly, the TBXT knockdown organoids displayed

multiple extensions emerging at day 5 (**Figure 4.16 D**). These multi-extension

aggregates maintained a reduced level of TBXT expression through day 10 and had

large cavities devoid of cells (**Figure 4.16 E,F**). At day 7, the wildtype control

demonstrated pockets of expression of the axis organizer LEFTY, whereas the TBXT

knockdown aggregates displayed reduced expression of LEFTY that was distributed

throughout the organoids (**Figure 4.16 G**). Both isogenic control and TBXT knockdown

organoids expressed the neuronal marker nestin, the caudal marker CDX2, and

basement membrane marker fibronectin (**Figure 4.16 H**). However, there was a

reduction in caudal HOX genes in the knockdown organoids at day 5 and day 10 of

differentiation, as well as a reduction in CDX2 and FGF8 expression (**Figure 4.16 I,**

Figure 4.17), reminiscent of the absent posterior structures in TBXT knockout mice.

Interestingly, organoids maintained expression of members of the Wnt pathway, such

as AXIN2 and WNT3, suggesting that TBXT is necessary in this system for

unidirectional extension and caudal fate specification, but not for maintaining Wnt

expression and signaling.
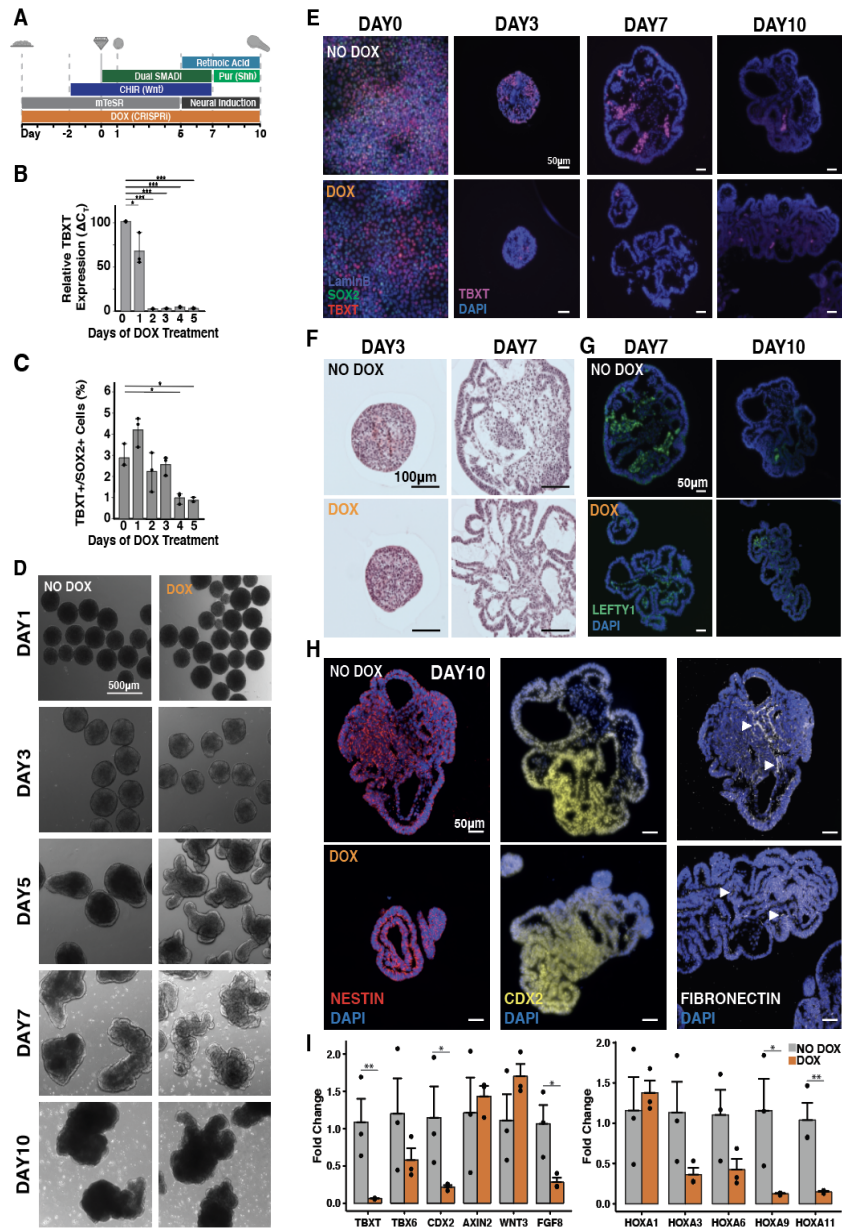
Figure 4.16 TBXT knockdown leads to multiple elongations
A) Schematic of differentiation protocol timeline with knockdown induction. B) Quantification of TBXT knockdown efficiency by reduction in mRNA expression levels measured by qPCR, each collected on day 0 of differentiation. C) Quantification by FLOW of TBXT(+)SOX2(+) cells with DOX treatment over time, each collected on day 0 of differentiation. D) Brightfield images of differentiation time course of knockdown hiPSC lines. E) Immunofluorescence staining of fixed cells (Day 0) or paraffin sections (Days 3 - 10) for TBXT over time in wildtype and KD organoids. F) Histological sections of wildtype and KD organoids at day 3 and 7 of differentiation. G) Immunofluorescence staining of paraffin sections for LEFTY at day 7 and 10 in wildtype and KD organoids. H) Immunofluorescence staining of paraffin sections for NESTIN, CDX2, and Fibronectin at day 10 in wildtype and KD organoids. I) qPCR quantification of HOX genes and genes related to axial pattern specification from day 10 organoids with and without DOX treatment (* signifies p-value < 0.05, ** signifies p-value < 0.005, *** signifies p-value < 0.0005).

149
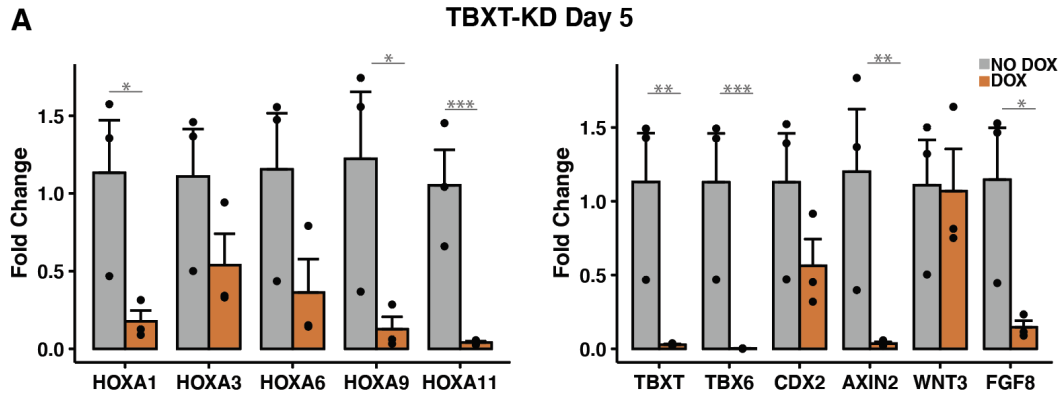
Figure 4.17 qPCR of TBXT KD at day 5 of differentiation
A) Quantification of gene expression by qPCR of HOX genes and genes associated with caudal fates in the TBXT knockdown organoids at day 5 of differentiation (* signifies p-value < 0.05, ** signifies p-value < 0.005, *** signifies p-value < 0.0005).

To interrogate the BMP pathway, Noggin and Chordin were knocked down

separately (**Figure 4.18 A,B**). Both Chordin and Noggin knockdown resulted in

increased extension compared to wildtype organoids (**Figure 4.18 C**). To further

examine extension dynamics, organoids were allowed to continue to grow for an

additional 7 days, to a total of 17 days of differentiation. Noggin knockdown organoids

extended continuously, becoming more polarized through day 13 relative to isogenic

control organoids. The pronounced extensions in Noggin knockdown organoids

appeared analogous to the elongated tails of Noggin knockout mice. In contrast,

Chordin knockdown organoids initially rapidly developed outgrowths without clear

anterior-posterior axis identity, but then halted elongation after day 10 of differentiation,

a morphogenic response reminiscent of the shortened body axis of Chordin knockout

mice (**Figure 4.18 D**).
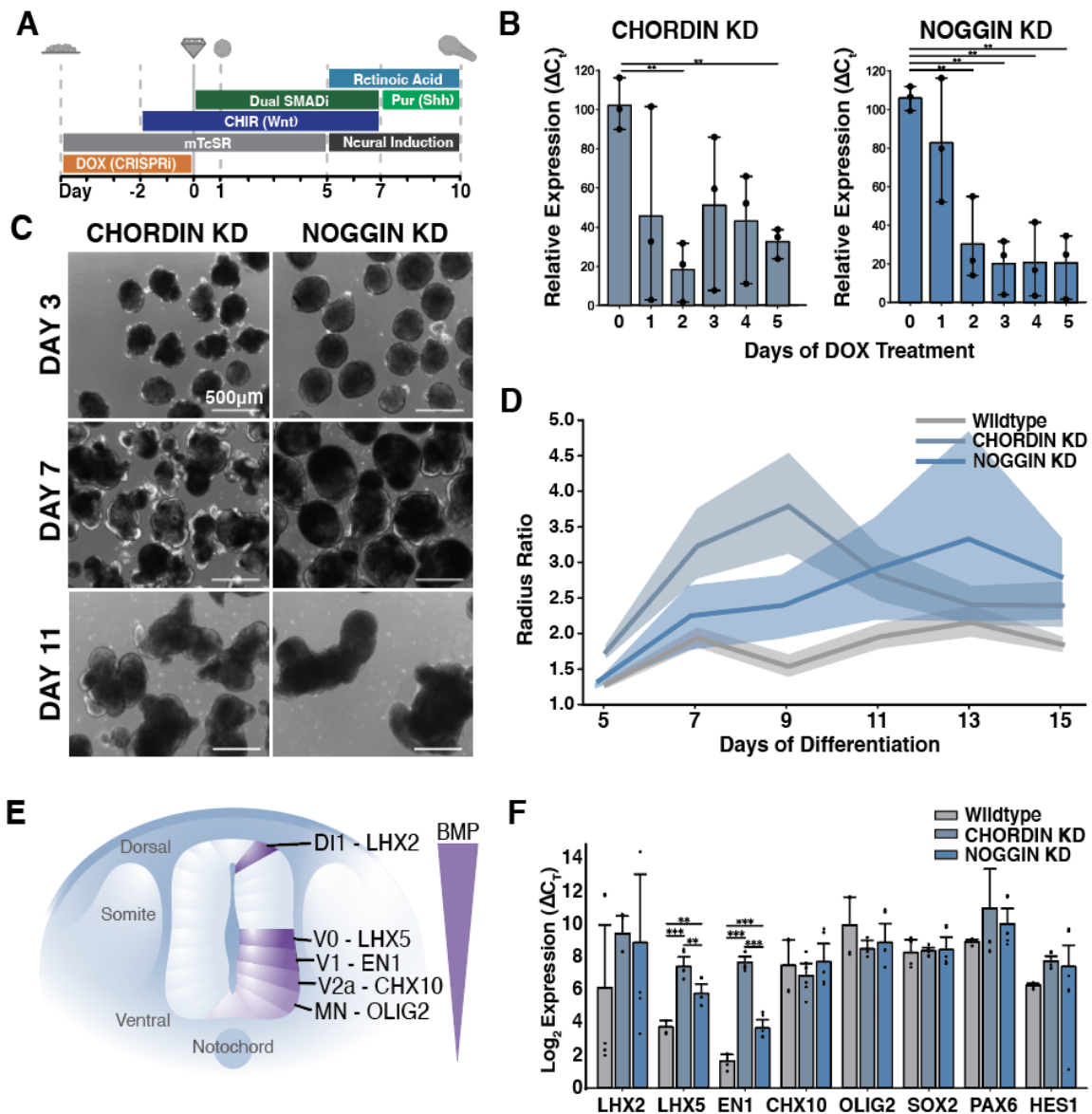
Figure 4.18 Manipulation of endogenous BMP signaling by CRISPR interference
A) Schematic of differentiation protocol timeline with knockdown induction. B) Quantification of Noggin and Chordin knockdown efficiency by reduction in mRNA expression levels measured by qPCR. C) Brightfield images of differentiation time course of knockdown hiPSC lines. D) Quantification of organoid length displayed as aspect ratio of major to minor axis in response to knockdown. E) Schematic depicting BMP signaling and progenitor specification within the developing neural tube. F) Quantification of mRNA expression of progenitor specific transcription factors by qPCR at day 17 of differentiation (n=3). Error bars depict standard deviations. Significance depicted by * indicating $p < 0.05$, ** indicating $p < 0.01$, and *** indicating $p < 0.001$.

Because of the role of BMP signaling in generating the dorsal-ventral axis of the spinal cord, progenitor domain gene expression was assessed by qPCR. At day 17, expression of LHX2, a marker of dI1 dorsal interneurons, did not significantly change in either knockdown conditions. However, expression of LHX5, a marker of V0 ventral interneuron populations, significantly increased in both Noggin and Chordin knockdown organoids (**Figure 4.18 E,F**). Although markers and regulators of neural maturation, Sox2, Pax6, and Hes1, were not significantly upregulated in the knockdowns, LHX5 and EN1, transcription factors that mark the central V0 and V1 domains, were upregulated. Interestingly, genes associated with the more extreme poles of the dorsal-ventral axis (LHX2, CHX10, and OLIG2) did not show significant differences, suggesting a potential change in threshold of BMP signaling that affects more central regions within the dorsal-ventral axis. Altogether, these results suggest that BMP signaling impacts both the anterior-posterior and dorsal-ventral progenitor emergence of the extending human organoids in a manner that recapitulates cellular patterning of murine neural tube development*.

4.4 Discussion

Here, hiPSC-derived organoids were demonstrated to model axial extension via the specification of progenitors that organize axial elongation and neural tube formation. This model enables robust axial extension as a function of both culture density and Wnt signaling. Within extending organoids, a persistent TBXT(+)SOX2(+) NMP population was observed, reminiscent of populations present in the caudal lateral epiblast and in the tail bud during axial elongation. Extending organoids displayed regionalized co-

linear patterning and timed expression of HOX genes. Single-cell RNA sequencing

confirmed the presence of caudal transcriptomic profiles and populations of neuronal

progenitors and paraxial mesoderm. Finally, axial extension of organoids was shown to

be influenced by Wnt and BMP signaling, such that perturbation of endogenous TBXT,

Chordin, or Noggin influences relative expression of neuronal markers along the

anterior-posterior and dorsal-ventral axis, as has been observed in parallel mouse

models (Bachiller, 2003; Chesley, 1935; McMahon et al., 1998; Pennimpede et al.,

2012; Wilson and Beddington, 1997; Wilson et al., 1993). Overall, the organoids

described herein provide reliable and faithful models of the gene regulatory networks,

multicellular organization, and structural outcomes that contribute to the development of

the primitive central nervous system in humans.

Stem cell-derived organoid models that reflect multiple organ systems (Lancaster

et al., 2013; Martins et al., 2020; Renner et al., 2017; Sato et al., 2009) and stages of

development (Beccari et al., 2018; van den Brink et al., 2020; Harrison et al., 2017;

Moris et al., 2020; Veenvliet et al., 2020) can now be used to investigate aspects of

human embryogenesis that were previously elusive due to poor tissue quality, technical

difficulties isolating and imaging embryonic tissues, and ethical concerns regarding

tissue procurement. Organoid platforms are amenable to varying culture conditions,

genetic modification, or small molecule drug screens, enabling the systematic

manipulation of experimental conditions at higher scale and speed than possible in most

model organisms. This work demonstrates that as multiple examples of axial elongation

are developed(Beccari et al., 2018; van den Brink et al., 2020; Marikawa et al., 2020;

153

Moris et al., 2020; Trivedi et al., 2019; Veenvliet et al., 2020), refining the protocols of organoid culture and differentiation can lead to more complex models than previously possible, and pushes the limits of organoid utility beyond cell signaling toward coordinated multicellular organization and morphogenesis en route to structurally and functionally mature tissues.

4.5 Bibliography

Amin, S., Neijts, R., Simmini, S., van Rooijen, C., Tan, S.C., Kester, L., van Oudenaarden, A., Creyghton, M.P., and Deschamps, J. (2016). Cdx and T Brachyury Co-activate Growth Signaling in the Embryonic Axial Progenitor Niche. Cell Rep. *17*, 3165–3177.

Bachiller, D. (2003). The role of chordin/Bmp signals in mammalian pharyngeal development and DiGeorge syndrome. Development *130*, 3567–3578.

Beccari, L., Moris, N., Girgin, M., Turner, D.A., Baillie-Johnson, P., Cossy, A.-C., Lutolf, M.P., Duboule, D., and Arias, A.M. (2018). Multi-axial self-organization properties of mouse embryonic stem cells into gastruloids. Nature.

Beck, F., Erler, T., Russell, A., and James, R. (1995). Expression of Cdx-2 in the mouse embryo and placenta: Possible role in patterning of the extra-embryonic membranes. Dev. Dyn. *204*, 219–227.

Beddington, R.S.P., and Robertson, E.J. (1999). Axis Development and Early Asymmetry in Mammals. Cell *96*, 195–209.

Bel-Vialar, S., Itasaki, N., and Krumlauf, R. (2002). Initiating Hox gene expression: in the early chick neural tube differential sensitivity to FGF and RA signaling subdivides the HoxB genes in two distinct groups. Development *129*, 5103–5115.

van den Brink, S.C., Alemany, A., van Batenburg, V., Moris, N., Blotenburg, M., Vivié, J., Baillie-Johnson, P., Nichols, J., Sonnen, K.F., Martinez Arias, A., et al. (2020). Single-cell and spatial transcriptomics reveal somitogenesis in gastruloids. Nature.

Butler, A., Hoffman, P., Smibert, P., Papalexi, E., and Satija, R. (2018). Integrating single-cell transcriptomic data across different conditions, technologies, and species. Nat. Biotechnol. *36*, 411–420.

Butts, J.C., McCreedy, D.A., Martinez-Vargas, J.A., Mendoza-Camacho, F.N., Hookway, T.A., Gifford, C.A., Taneja, P., Noble-Haeusslein, L., and McDevitt, T.C. (2017). Differentiation of V2a interneurons from human pluripotent stem cells. Proc. Natl. Acad. Sci. *114*, 4969–4974.

Cambray, N., and Wilson, V. (2007). Two distinct sources for a population of maturing axial progenitors. Development *134*, 2829–2840.

Carpenter, E.M. (2002). Hox Genes and Spinal Cord Development. Dev. Neurosci. *24*, 24–34.

Chesley, P. (1935). Development of the short-tailed mutant in the house mouse. J. Exp. Zool. *70*, 429–459.

del Corral, R.D., and Storey, K.G. (2004). Opposing FGF and retinoid pathways: a signalling switch that controls differentiation and patterning onset in the extending vertebrate body axis. BioEssays *26*, 857–869.

Ericson, J., Briscoe, J., Rashbass, P., and Heyningen, V.V. Graded Sonic Hedgehog Signaling and the Specification of Cell Fate in the Ventral Neural Tube. 16.

Gouti, M., Tsakiridis, A., Wymeersch, F.J., Huang, Y., Kleinjung, J., Wilson, V., and Briscoe, J. (2014). In Vitro Generation of Neuromesodermal Progenitors Reveals Distinct Roles for Wnt Signalling in the Specification of Spinal Cord and Paraxial Mesoderm Identity. PLoS Biol. *12*, e1001937.

Grady, L. (2006). Random walks for image segmentation. Pattern Anal. Mach. Intell. IEEE Trans. On *28*, 1768–1783.

Halır, R., and Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG, (Citeseer), pp. 125–132.

Harrison, S.E., Sozen, B., Christodoulou, N., Kyprianou, C., and Zernicka-Goetz, M. (2017). Assembly of embryonic and extraembryonic stem cells to mimic embryogenesis in vitro. Science *356*, eaal1810.

Henrique, D., Abranches, E., Verrier, L., and Storey, K.G. (2015). Neuromesodermal progenitors and the making of the spinal cord. Development *142*, 2864–2875.

Herrmann, B.G., Labeit, S., Poustka, A., King, T.R., and Lehrach, H. (1990). Cloning of the T gene required in mesoderm formation in the mouse. Nature *343*, 617–622.

Hookway, T.A., Butts, J.C., Lee, E., Tang, H., and McDevitt, T.C. (2016). Aggregate formation and suspension culture of human pluripotent stem cells and differentiated progeny. Methods *101*, 11–20.

Kim, D., Langmead, B., and Salzberg, S.L. (2015). HISAT: a fast spliced aligner with low memory requirements. Nat. Methods *12,* 357–360.

Lancaster, M.A., and Knoblich, J.A. (2014). Organogenesis in a dish: Modeling development and disease using organoid technologies. Science *345*, 1247125–1247125.

Lancaster, M.A., Renner, M., Martin, C.-A., Wenzel, D., Bicknell, L.S., Hurles, M.E., Homfray, T., Penninger, J.M., Jackson, A.P., and Knoblich, J.A. (2013). Cerebral organoids model human brain development and microcephaly. Nature *501*, 373–379.

Larson, M.H., Gilbert, L.A., Wang, X., Lim, W.A., Weissman, J.S., and Qi, L.S. (2013). CRISPR interference (CRISPRi) for sequence-specific control of gene expression. Nat. Protoc. *8*, 2180–2196.

Law, C.W., Chen, Y., Shi, W., and Smyth, G.K. (2014). voom: precision weights unlock linear model analysis tools for RNA-seq read counts. Genome Biol. *15*, R29.

Liao, Y., Smyth, G.K., and Shi, W. (2014). featureCounts: an efficient general purpose program for assigning sequence reads to genomic features. Bioinformatics *30*, 923–930.

Libby, A.R., Joy, D.A., So, P.-L., Mandegar, M.A., Muncie, J.M., Mendoza-Camacho, F.N., Weaver, V.M., Conklin, B.R., and McDevitt, T.C. (2018). Spatiotemporal mosaic self-patterning of pluripotent stem cells using CRISPR interference. ELife *7*.

Liem, K.F., Tremml, G., Roelink, H., and Jessell, T.M. (1995). Dorsal differentiation of neural plate cells induced by BMP-mediated signals from epidermal ectoderm. Cell *82*, 969–979.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. Nat. Methods *3*, 637–646.

Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. Cell Stem Cell *18*, 541–553.

Marikawa, Y., Chen, H.-R., Menor, M., Deng, Y., and Alarcon, V.B. (2020). Exposure-based assessment of chemical teratogenicity using morphogenetic aggregates of human embryonic stem cells. Reprod. Toxicol. *91*, 74–91.

Martins, J.-M.F., Fischer, C., Urzi, A., Vidal, R., Kunz, S., Ruffault, P.-L., Kabuss, L., Hube, I., Gazzerro, E., Birchmeier, C., et al. (2020). Self-organizing 3D human trunk neuromuscular organoids. Cell Stem Cell *26*, 172–186.

McGinnis, C.S., Patterson, D.M., Winkler, J., Hein, M.Y., Srivastava, V., Conrad, D.N., Murrow, L.M., Weissman, J.S., Werb, Z., Chow, E.D., et al. (2018). MULTI-seq: Scalable sample multiplexing for single-cell RNA sequencing using lipid-tagged indices (Genomics).

McMahon, J.A., Takada, S., Zimmerman, L.B., Fan, C.-M., Harland, R.M., and McMahon, A.P. (1998). Noggin-mediated antagonism of BMP signaling is required for growth and patterning of the neural tube and somite. Genes Dev. *12*, 1438–1452.

Miyaoka, Y., Chan, A.H., Judge, L.M., Yoo, J., Huang, M., Nguyen, T.D., Lizarraga, P.P., So, P.-L., and Conklin, B.R. (2014). Isolation of single-base genome-edited human iPS cells without antibiotic selection. Nat. Methods *11*, 291–293.

Moris, N., Anlas, K., van den Brink, S.C., Alemany, A., Schröder, J., Ghimire, S., Balayo, T., van Oudenaarden, A., and Martinez Arias, A. (2020). An in vitro model of early anteroposterior organization during human development. Nature *582*, 410–415.

Oginuma, M., Harima, Y., Tarazona, O.A., Diaz-Cuadros, M., Michaut, A., Ishitani, T., Xiong, F., and Pourquié, O. (2020). Intracellular pH controls WNT downstream of glycolysis in amniote embryos. Nature *584*, 98–101.

Ortmann, D., and Vallier, L. (2017). Variability of human pluripotent stem cell lines. Curr. Opin. Genet. Dev. *46*, 179–185.

Pennimpede, T., Proske, J., König, A., Vidigal, J.A., Morkel, M., Bramsen, J.B., Herrmann, B.G., and Wittler, L. (2012). In vivo knockdown of Brachyury results in skeletal defects and urorectal malformations resembling caudal regression syndrome. Dev. Biol. *372*, 55–67.

Renner, M., Lancaster, M.A., Bian, S., Choi, H., Ku, T., Peer, A., Chung, K., and Knoblich, J.A. (2017). Self-organized developmental patterning and differentiation in cerebral organoids. EMBO J. *36*, 1316–1329.

Sato, T., Vries, R.G., Snippert, H.J., van de Wetering, M., Barker, N., Stange, D.E., van Es, J.H., Abo, A., Kujala, P., Peters, P.J., et al. (2009). Single Lgr5 stem cells build crypt-villus structures in vitro without a mesenchymal niche. Nature *459*, 262–265.

Schiffmann, Y. (2007). Self-organization in and on biological spheres. Prog. Biophys. Mol. Biol. *95*, 50–59.

Steventon, B., Duarte, F., Lagadec, R., Mazan, S., Nicolas, J.-F., and Hirsinger, E. (2016). Species-specific contribution of volumetric growth and tissue convergence to posterior body elongation in vertebrates. Development *143*, 1732–1741.

Tanabe, Y., and Jessell, T.M. (1996). Diversity and Pattern in the Developing Spinal Cord. Science *274*, 1115–1123.

Trivedi, V., Fulton, T., Attardi, A., Anlas, K., Dingare, C., Martinez-Arias, A., and Steventon, B. (2019). Self-organised symmetry breaking in zebrafish reveals feedback from morphogenesis to pattern formation (Developmental Biology).

Turner, D.A., Rué, P., Mackenzie, J.P., Davies, E., and Martinez Arias, A. (2014). Brachyury cooperates with Wnt/β-catenin signalling to elicit primitive-streak-like behaviour in differentiating mouse embryonic stem cells. BMC Biol. *12*, 63.

Veenvliet, J.V., Bolondi, A., Kretzmer, H., Haut, L., Scholze-Wittler, M., Schifferl, D., Koch, F., Pustet, M., Heimann, S., Buschow, R., et al. (2020). Mouse embryonic stem cells self-organize into trunk-like structures with neural tube and somites. BioRxiv.

van der Walt, S., Schönberger, J.L., Nunez-Iglesias, J., Boulogne, F., Warner, J.D., Yager, N., Gouillart, E., and Yu, T. (2014). Scikit-image: image processing in Python. PeerJ *2*, e453.

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

Wilson, V., and Beddington, R. (1997). Expression of T Protein in the Primitive Streak Is Necessary and Sufficient for Posterior Mesoderm Movement and Somite Differentiation. Dev. Biol. *192*, 45–58.

Wilson, V., Rashbass, P., and Beddington, R. (1993). Chimeric analysis of T (Brachyury) gene function. Development *117*, 1321–1331.

Wilson, V., Olivera-Martinez, I., and Storey, K.G. (2009). Stem cells, signals and vertebrate body axis extension. Development *136*, 2133–2133.

Wymeersch, F.J., Huang, Y., Blin, G., Cambray, N., Wilkie, R., Wong, F.C., and Wilson, V. (2016). Position-dependent plasticity of distinct progenitor types in the primitive streak. Elife *5*.

Yamaguchi, T.P. (2001). Heads or tails: Wnts and anterior–posterior patterning. Curr. Biol. *11*, R713–R724.

Yu, G., Wang, L.-G., Han, Y., and He, Q.-Y. (2012). clusterProfiler: an R Package for Comparing Biological Themes Among Gene Clusters. OMICS J. Integr. Biol. *16*, 284–287.

Zhang, Z., Yu, D., Seo, M., Hersh, C.P., Weiss, S.T., and Qiu, W. (2019). Novel Data

Transformations for RNA-seq Differential Expression Analysis. Sci. Rep. *9*.

Zheng, Y., Xue, X., Resto-Irizarry, A.M., Li, Z., Shao, Y., Zheng, Y., Zhao, G., and Fu, J. (2019).

Dorsal-ventral patterned neural cyst from human pluripotent stem cells in a neurogenic niche.

Sci. Adv. *5*, eaax5933.

5 Directed cell migration drives spontaneous symmetry breaking during simulated gastrulation

5.1 Introduction

Human embryos begin as an aggregate of totipotent stem cells that transition through a series of lineage restriction and symmetry breaking events to self-organize into the tissues of the body. While mechanisms of pre-patterning cell fates are known for specific model organisms such as *Xenopus* (Keller and Danilchik, 1988; Moon et al., 1993) and *Drosophila (Wieschaus, 2016)*, the asynchronicity of mammalian embryogenesis and the complexity of studying embryos post-implantation has limited the ability to construct mechanistic models of embryo organization in mammals (Płusa and Piliszek, 2020). Recent advances in human organoid systems have allowed for *in vitro* interrogation of morphogenic processes reminiscent of those found during early development, including models of central nervous system development (Velasco et al., 2020), gastrulation and axial extension (van den Brink et al., 2014; Turner et al., 2017), and multi-system co-emergence (Silva et al., 2020; Takasato et al., 2015). However, unlike the highly coordinated, robust sequence of tissue forming events found during embryogenesis, organoid structures form in a highly variable, disordered arrangement (Lou and Leung, 2018; Qian et al., 2019), suggesting the need for more mechanistic understanding of the processes underlaying multicellular structure assembly.

Computational models of multicellular behavior have long been used to understand developmental pattern formation in contexts including coat patterning (Marcon and Sharpe, 2012; Turing, 1952), limb formation (Scoones and Hiscock, 2020), and

somitogenesis (Pantoja-Hernández et al., 2020; Tomka et al., 2018). Broadly, models fall into three classes: continuum models, lattice-based models, and lattice free approaches, with agent based models (ABMs) representing a large portion of the latter two groups (Glen et al., 2019; González-Valverde and García-Aznar, 2018; Voss-Böhme, 2012). Although lattice-based models have had enormous success predicting pattern formation in various 2D contexts, they require hardware acceleration to scale to 3D and need parametric tuning to replicate *in vitro* measurements (González-Valverde et al., 2016; Libby et al., 2019; Madhikar et al., 2018). In contrast, particle-based ABMs are amenable to simultaneous simulation of tens of thousands of cells on CPUs, are inherently n-dimensional, and incorporate direct biological measurements relatively easily (González-Valverde and García-Aznar, 2018). Despite extreme simplification of the complex processes occurring within a cell, the scalability of ABMs makes them a tractable system to model the adhesion and collective migration processes underlaying organoid formation.

In this chapter, I present a particle-based ABM simulating the dynamics underlaying symmetry breaking in a gastruloid model of neural tube formation. Initial physical parameters were fit based on the detailed measurements of individual and collective cell behavior derived in chapters 2 and 3. The resulting parameters were then refined using measurements of the single cell structure of the elongating neural organoids presented in chapter 4. Finally, a dynamic, full-scale simulation of particle migration in organoids was performed, demonstrating how the interactions between differential cell ratios and chemotaxis enable a randomly mixed, spherical cell distribution to break symmetry and

form axially polarized populations. By exploring the dynamic range of cell ratios, cell-type specific chemoattraction, and surface seeking behavior, this simulation provides a testable, mechanistic model for organoid polarization. Further, by providing a high-level description of the individual multicellular elements involved in organoid polarization, this model suggests potential targets for modulating the organizing processes underlaying gastrulation and neural tube formation *in vitro.*

5.2 Methods

5.2.1 Generation of organoids

Elongating and non-elongating organoids were generated as described in Chapter 4. Briefly, hiPSCs derived from the WTC11 line (Coriell Cat. #GM25256) were cultured in feeder-free conditions on growth-factor reduced Matrigel (BD Biosciences) coated plates and fed daily with mTeSR1 medium (STEMCELL Technologies)(Ludwig et al., 2006). Cells were passaged by dissociating with Accutase (STEMCELL Technologies) the re-seeded in mTeSR1 supplemented with the small molecule Rho-associated coiled-coil kinase (ROCK) inhibitor Y-276932 (10 µM; Selleckchem) (Park et al., 2015) at a seeding density of 12,000 cells per cm^2. To induce differentiation, hiPSCs were seeded at 125,000 cells per cm^2 in mTeSR1 medium supplemented with Y-276932 and small molecule GSK inhibitor CHIR99021 (2µM, 4µM, or 6µM; Selleckchem). Two days later, cells were singularized with Accutase (STEMCELL Technologies), counted using a Countess II FL (Life Technologies), and seeded at 810,000 cells per well into 800µm X 800µm PDMS microwell inserts in a 24 well plate (~270 wells/insert; 3,000 cells per microwell) (Hookway et al., 2016) in mTeSR1

(STEMCELL Technologies) supplemented with Y-276932 (10 µM), CHIR99021 (2µM, 4µM, or 6µM), ALK5 small molecule inhibitor SB431542 (10µM, Selleckchem), and small molecule BMP inhibitor LDN193189 (0.2µM, Selleckchem). After ~18 hours, condensed organoids were transferred to rotary culture at a density of approximately 270 organoids per well, or allowed to reattach to 96-well plates for imaging as described below.

5.2.2 Attached organoid imaging and colony tracking

Organoids were allowed to reattach to optically clear 96-well plates (Corning) coated with growth factor reduced Matrigel (BD Biosciences) for 48 hours and fed with mTeSR1 medium (STEMCELL Technologies) supplemented with CHIR99021 (2µM, 4µM, or 6µM), SB431542 (10µM), and LDN193189 (0.2µM). After 48 hours, fresh media was exchanged, then colonies were imaged for 24 hours with images taken every 5 minutes on an incubated spinning disk confocal Observer Z1 (Zeiss) using a motorized filter wheel (Yokogawa) and imaged using a Prime 95B (Photometrics) digital CMOS camera at 10x magnification (NA 0.45, 0.91 μm x 0.91 μm per pixel), as described in Chapter 3. Colonies were segmented using the dense cell tracking pipeline described in Chapter 3, yielding a data set of total cell behavior in adherent colonies corresponding to days 3-4 of the elongating organoid differentiation protocol from Chapter 4.

5.2.3 Colony morphology characterization

The perimeter of each colony was calculated by finding the largest connected component of each cell mesh for each frame, then calculating the boundary of that connected component. An elliptical model was fit to this contour by least squares (Halır

and Flusser, 1998) to calculate eccentricity and the ratio of semi-major to semi-minor axes. Area and circularity were also calculated, giving a second measure of colony shape. Rosette structures were manually segmented by selecting an ellipse around the visible density anomaly in the final frame of each colony image. The cell tracks enclosed this elliptical contour were then projected backwards through time to estimate rosette size and shape over time. Cell radius was estimated as half the average distance between detected cell centers of mass in non-rosette regions, which was consistent between CHIR treated and non-treated colonies.

5.2.4 Cell migration and growth rate estimation

Cell velocity and active/passive percentages was calculated by first interpolating individual tracks by a factor of 4, then smoothing with a rolling average of 5 samples, then calculating velocity magnitude and direction as described in Chapter 3. Smoothed tracks were divided into regions of time where velocity was greater than at least 0.1 µm/min and where a cell changed direction less than 3 degrees/minute, which were called active migration, and other times, where cells were considered quiescent. The proportion of the two times were calculated, and used as transition probabilities in a persistent random motion model (Campos et al., 2010).

Cell growth was estimated using the total number of cell tracks in each frame. A growth rate was then estimated by normalizing the total number of cells by the initial cell number in each colony, yielding an approximately exponential empirical distribution. An exponential model was fit to this curve using least squares, giving the mean cell

doubling time according to. This was scaled by the observation time to calculate the probability of cell division in a single simulation step.

5.2.5 Rotary culture and phase imaging

Organoids were cultured in 6-well plates on a rotary orbital shaker as described in Chapter 4. Media was exchanged every other day for up to 10 days. At day 5, organoids were transferred to Neural Induction Media (DMEM F:12 (Corning), N2 supplement (Life Technologies), L-Glutamine (VWR), 2µg/ml heparin (Sigma Aldrich), non-essential amino acids (Mediatech INC), penicillin-streptomycin (VWR), supplemented with fresh 0.4µg/ml ascorbic acid (Sigma Aldrich) and 10ng/ml brain derived neurotrophin factor (BDNF, R&D Systems)) supplemented with CHIR99021 (2µM, 4µM, or 6µM; Selleckchem), SB431542 (10µM, Selleckchem), and LDN193189 (0.2µM, Selleckchem). From day 7 onwards, organoids were fed with Neural Induction Media supplemented with retinoic acid (10nM, Sigma Aldrich), purmorphamine (300nM, EMD Millipore). Organoids were imaged daily at 4x magnification in phase using an EVOS M5000 imaging system (ThermoFisher Scientific), and organoid images segmented as described in Chapter 4, giving a set of longitudinal organoid contours over the course of 10 days of differentiation.

At days 3 and 5, organoids from 6-well plates were individually transferred using wide bore pipette tips into the center 60 wells of an uncoated ultra-low attachment 96-well plate (Corning), seeding exactly one organoid per well, and imaged using an inverted Axio Observer Z1 (Zeiss) microscope with incubation as described in Chapter 4. Wells were imaged at 5x magnification in TL Brightfield every 20 minutes for 48

hours, and the resulting time lapse images segmented, giving a denser set of contours covering the 48-hour period of organoid polarization and initial axis extension. Gross motion of organoid structures was calculated using optical flow (Farnebäck, 2003), giving an estimate for multicellular migration velocity magnitude, direction, and duration. Organoid volume was estimated by calculating the area of the exterior contour of each organoid segmentation, calculating the effective radius of a circle with the same area, then scaling that effective radius to a sphere, i.e:

$$V_{eff} \cong \frac{4\pi}{3} \, r_{eff}^{\,3} = \frac{4\pi}{3} \left(\frac{A}{\pi}\right)^{3/2}$$

corresponding to a uniform circular inflation of the contour in all directions. Cell number within the estimated volume was given by assuming optimal sphere packing of spherical cells with the cell radius estimated as half the cell-cell distance from 2D cell tracking, i.e.

$$n_{cell,eff} \cong 0.74 * \frac{V_{eff}}{V_{cell}} = 0.74 * \left(\frac{r_{eff}}{r_{cell}}\right)^3$$

5.2.6 Whole Mount Light Sheet Imaging and Segmentation

Organoids were fixed with 4% paraformaldehyde (VWR) for 40 minutes, washed three times with PBS, permeabilized with 1.5% Triton X-100 (Sigma-Aldrich) for 1 hour, blocked in 5% normal donkey serum (Jackson Immunoresearch) for 1 hour, and probed with primary and secondary antibodies overnight. Nuclei were stained with a 1:10000 dilution of Hoechst 33342 (Thermo Fisher) included with secondary antibodies. Samples were then embedded in 1.5% low melt agarose (BioReagent) and drawn up into ~1mm imaging capillaries and subsequently imaged on the Zeiss Z.1 Light sheet Microscope equipped with two PCO. edge SCMOS cameras at 5X (NA 1.34, aqueous objective).

Multiple views of each volume were fused in Zen Black 2014 (Zeiss) using the Multi-view fusion module with rotation and translation alignment on intensity of the Hoechst channel. Cell puncta were segmented in Imaris v9.5.1 (Oxford Instruments) using the Spot detection tool on each of the channels corresponding to Hoescht, TBXT, and SOX2 expression, with a minimum spot diameter of 5 µm, a maximum spot diameter of 20 µm, and an intensity threshold corresponding to at least 20% above the estimate of the local fluorescent background for that channel.

5.2.7 Cell-cell spacing and 3D organoid shape and position

Detected nuclei were merged by clustering all spot detections that co-occurred within a 5 µm, then classifying detections as either containing a TBXT+ detection, or otherwise being a TBXT- cell. SOX2 detection in stained samples was ubiquitous (>99% positive clusters), so no attempt was made to classify cells by SOX2 status. Cell neighborhood size was assessed by calculating the distance to the nearest 20 neighbors and detecting the smallest number of neighbors where the mean distance to neighbors rapidly increases, giving a mean neighborhood size of approximately 12 cells. The convex hull of the organoid was calculated (Barber et al., 1996), and the organoid convex volume, ellipsoid parameters, and the distance of each cell to the convex centroid were measured to give a normalized radial cell position within the organoid volume. Cells were clustered using k-means clustering with k evaluated between 1 and 20 clusters by the silhouette coefficient using the implementations in scikit-learn v0.23 (Pedregosa et al., 2011). Cluster centroids were estimated and measures of both

normalized radial cluster position and distance between clusters was calculated to estimate the extent of cluster polarization and cluster convergence.

5.2.8 Initialization of Randomly Packed Spherical Organoids

Initial cell position within an organoid was generated by simulating a close random packing of hard spheres with a uniform radius equal to the average cell radius, enclosed within a spherical volume defined by the mean radius of the segmented organoids. Cell positions were drawn uniformly at random and placed following the u-statistic for spheres:

$$r \sim \mathcal{U}(0,1)^{1/3}; \cos(\theta) \sim \mathcal{U}(-1,1); \; \varphi \sim \mathcal{U}(0,2\pi)$$

$$\vec{x}_{cell} = \begin{pmatrix} r * sin(\theta) * \cos(\varphi) \\ r * \sin(\theta) * \sin(\varphi) \\ r * \cos(\theta) \end{pmatrix}$$

A running index of previously placed cells was maintained, and any newly sampled cells which collided with a previous cell were redrawn until they no longer collided, or ten samples failed in a row, whichever occurred first (Visscher and Bolsterli, 1972). Previous mixing studies, as presented both in Chapter 2 and in (Turaga et al., 2020) showed that aggregates of two cell types are initially uniformly mixed, so cell populations were assigned to either "red" cells representing TBXT- cell types, or "gold" cells, representing the TBXT+, polarizing fraction of cells uniformly and at random:

$$P(gold) = 1 - P(red) = \mathcal{U}\left(0, \frac{n_{gold}}{n_{red} + n_{gold}}\right)$$

Initializing cells uniformly throughout the volume agreed best with empirical measurements, but other distributions including inside-out, planar polarized, or ring shaped were also easily generated using this sampling method.

5.2.9 Simulation of cell-cell adhesion

Cell-cell adhesion was simulated with a modified Lennard-Jones potential (González-Valverde and García-Aznar, 2018; González-Valverde et al., 2016):

$$\vec{F}_{adh} = w_{adh} \sum_{j=0}^{neighbors(i)} \left[\frac{r_{adh}^4}{(\vec{r}_i - \vec{r}_j)^5} - \frac{r_{adh}^2}{(\vec{r}_i - \vec{r}_j)^3} \right]$$

where r_{adh} was the empirically derived mean cell-cell spacing (7.5 μm), and w_{adh} was a tunable parameter to control how strongly cells experienced adhesion forces (set to 1.0 for simulations described below). This function generated a profile where cells very close to each other were strongly repulsed, while cells within several multiples of r_{adh} were mildly attracted to one-another. Cells far away felt an attractive force that diminished with the 3rd power of the distance, enabling force calculations to be performed on a sparse graph of cells within a distance of $7 * r_{adh}$ (where $\left| \vec{F}_{adh} \right| <$ 0.01 μm/min^2), speeding up the simulation as cell numbers increased from O(n^2) to O(n log(n)). Cell type specific values for r_{adh} were possible in this framework, but empirical results showed no difference in cell packing between the two cell types, so a single value was used for red-red, red-gold, and gold-gold interactions.

5.2.10 Simulation of stochastic cell migration

To simulate a stochastic active passive model of cell migration, cells were partitioned into a two-state migration model using the empirically derived mean durations for active migration and quiescence. Cells in the active state selected a velocity direction uniformly at random, and a velocity magnitude from a normal distribution with mean 0.33 μm/min (+/- 0.05 μm/min), which produced similar results to

the empirical heavy tailed distribution but had fewer parameters to optimize when matching empirical and simulation behavior. Cells in the active state transitioned to quiescence with a probability of $P(Q_t|A_{t-1}) = \frac{\Delta t}{t_{active}}$ where Δt was the simulation time step and t_{active} = 15 minutes, the mean active migration time from the data set. Cells in the quiescent state were defined to have an intrinsic velocity magnitude of 0 µm/min, so only moved in response to external forces. Quiescent cells transitioned to an active state with an analogous probability $P(A_t|Q_{t-1}) = \frac{\Delta t}{t_{quiescent}}$ where Δt was the simulation time step and $t_{quiescent}$ = 40 minutes, again derived from the empirical mean time. At colony initialization, cells were assigned active or passive at random with probability of initial state given by the expected ratio of ~3:8 active to quiescent cells. Intrinsic velocity was not evaluated during force integration, but was instead evaluated during velocity integration to update cell positions only, and so represented a stochastic impulse input to the simulation, rather than a true force. These results gave the coupled system of differential equations:

$$\frac{d\vec{V}_{adh}}{dt} = \vec{F}_{adh}$$

$$\frac{d\vec{x}}{dt} = \vec{V}_{adh} + \vec{V}_{act}$$

Which were evaluated numerically using adaptive step-size Euler integration that capped the maximum force evaluated at any step to 20 µm/min² and the maximum velocity evaluated at any step to 2.0 µm/min. Aggregates were stably integrated forward in time in time steps of Δt = 12 seconds for several days in all conditions described below.

172

5.2.11 Simulation of cell growth

Cell growth was incorporated into the simulation by giving each cell a probability to divide at each time step given by $P(divide) = \Delta t / t_{division}$ where Δt was the time step and $t_{division}$ was the empirical cell division time of 3,258 minutes/division (~2.2 days). Upon division, a new cell was generated, duplicating the cell type, but initializing the cell migratory state independently. To place the new cell, a direction was drawn randomly around the dividing cell, and then the current and new cells were placed at opposite ends of this axis, spaced $\frac{r_{adh}}{2}$ apart, which was empirically determined as a ratio where sudden appearance of a cell did not cause a large spike in the force balance of cells. Cell type specific division rates were not supported by empirical data, and divisions that induced a transition in cell state were not explored, so only one division rate was used. Using these simulation parameters, organoids grew exponentially over several days of simulation time while maintaining approximately equal ratios of cell types, cell migration states, and cell-cell spacing.

5.2.12 Simulation of chemoattractive migration

To incorporate cell-cell chemoattraction, the "gold" cells were modified to alter their active migration velocity direction each time step. For each "gold" cell, the centroid of the nearest 5 "gold" cells was calculated, then the direction of any actively migrating cells modified with the re-weighting

$$\hat{V}_{act} := [w_{chemo} * \hat{V}_{chemo} + (1 - w_{chemo}) * \hat{V}_{act})$$

where $\hat{V}_{chemo}$ was the direction of the centroid, and $\hat{V}_{act}$ was the current active migration direction for each active cell. w_{chemo} was the strength of attraction, where 0 represented

no chemoattraction and 1 represented only chemoattraction and was tuned as described in the experiments below.

5.2.13 Simulation of surface-seeking migration

The surface of each simulated organoid was estimated by calculating the convex hull of each organoid, then calculating the nearest vertex of the hull for each "gold" cell. Analogous to the chemoattractive model, the only active migration velocity direction was modified for each cell at each time step, according to

$$\hat{V}_{act} := [w_{surf} * \hat{V}_{surf} + (1 - w_{surf}) * \hat{V}_{act})$$

Where $\hat{V}_{surf}$ was the direction of the nearest surface vertex and $\hat{V}_{act}$ was the current active migration direction for each active cell. w_{surf} was the strength of the surfacing pseudo-force, where 0 represented no surfacing impulse, and 1 represented directed migration towards the surface, and was tuned in the experiments below. Where chemoattraction and surfacing were both simulated simultaneously, the two weights were set to obey the inequality

$$w_{chemo} + w_{surf} \leq 1.0$$

with parameter settings sampled along the resulting simplex.

5.2.14 Statistical methods and sampling

Empirical measurements were made based on at least 3 biological replicates and across multiple batches where available. Each setting of aggregate parameters was simulated at least 3 times with distinct random seeds. Distributions were compared using the Mann–Whitney U test. Two tailed t-tests were used to assess p-values in all

174

pairwise statistics presented, with significance reported as *p<0.05, **p<0.01, and ***p<0.001.

5.3 Results

5.3.1 Organoid parameters derived from differentiating colonies

Although there are well known behavioral and structural differences between adherent colonies and organoids, local cell neighborhoods in plated organoids maintain a quasi-3D nature for some time, so the colony morphology of adhered elongating and non-elongating organoids was characterized to compare with suspension culture (**Figure 5.1 A**). CHIR-pretreated (elongating) and untreated (non-elongating) organoids were allowed to reattach to plates for 48 hours, then analyzed using the cell tracking pipeline from Chapter 3 to generate detailed measures of cell-cell adhesion, migration rate, and proliferation between days 3 and 4 of differentiation. Gross colony morphology was largely round for the entire 24-hour period of observation, with both colony size and number of detected cells increasing following an exponential growth curve with a doubling time of 21.2 hours (**Figure 5.1 B-D**). CHIR pre-treated organoids were overall larger, and grew at faster rates during the 24-hour period, but no differences were found between regions in either CHIR pre-treated or untreated colonies. Cells were packed in a hexagonal arrangement with an average of 5.9 neighbors. Cell-cell spacing grew slightly over 24 hours in both protocols, from 13.5 μm to 14.0 μm on average. CHIR pre-treated colonies increased in spreading at the colony edge vs untreated, consistent with CHIR-induced density changes shown in Chapters 2 and 3 (**Figure 5.1 E-G**).
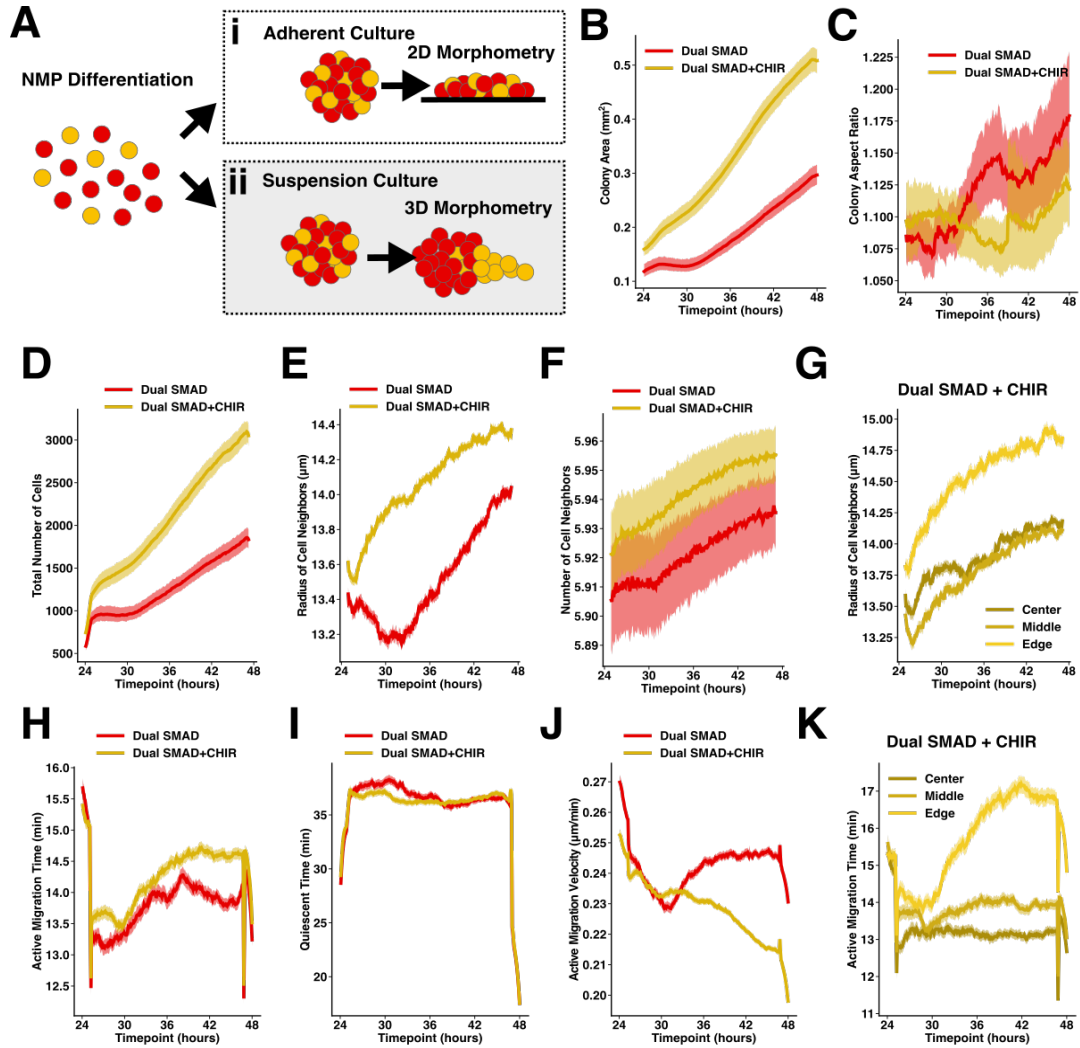
Figure 5.1 2D morphometry of re-attached organoids between day 3 and 4
A) Schematic of approach to compare 2D and 3D organoid structure. B) Growth of adherent colony area over 24 hours for CHIR pre-treated (yellow) and untreated (red) organoids. C) Change in adherent colony ratio of major to minor axis over 24 hours. D) Growth in cell number. E) Average radius of nearest neighbors to each cell over time. F) Average number of nearest neighbors over time. G) Region stratified cell-cell spacing for CHIR-pretreated colonies for cells in the center, middle, and edge of a colony. H) Active and I) passive migration periods for CHIR pretreated colonies. J) Velocity of actively migrating cells in CHIR-pretreated colonies. K) Migration persistence of actively migrating cells in CHIR-pretreated colonies.

For the vast majority of colonies, cell migration followed the typical stochastic active passive migration cycle found in wild type colonies with similar periods of 14 minutes active followed by 40 minutes passive migration time on average. In CHIR pre-treated colonies, the active migration period of cells on the periphery increased from 14

to 17 minutes per cycle, while the velocity of migration fell from 0.25 µm/min to 0.22

um/min, consistent with similar trends found in BMP4 differentiation, but distinct from

CHIR differentiation. Active migration time and directionality increased on the periphery

of CHIR pre-treated colonies, suggesting that the periphery of colonies was

recapitulating the migratory phenotype observed during optoWnt differentiation as

described in Chapter 2 (**Figure 5.1 H-K**).

5.3.2 Local cell neighborhood structures replicated in 3D organoids

One of the current challenges of developing accurate organoid models of *in vivo*

tissues is the disordered formation of mesoscale and macroscale structures, where

individual tissue components become juxtaposed in artificial ways not observed in living

organisms. Previous studies of force aggregation of mixed populations of cells have

demonstrated that microwell aggregation thoroughly randomizes the spatial composition

of the aggregate (Turaga et al., 2020). Hence, multicellular structures must form post-

aggregation, with multiple interacting mechanisms such as differentiation, clonal

expansion, cell sorting, or cell migration responsible for creating higher order

organization. This spontaneous formation of ordered structures makes organoid

systems an ideal platform to study symmetry breaking from a disordered, spherically

symmetric distribution of single cells.
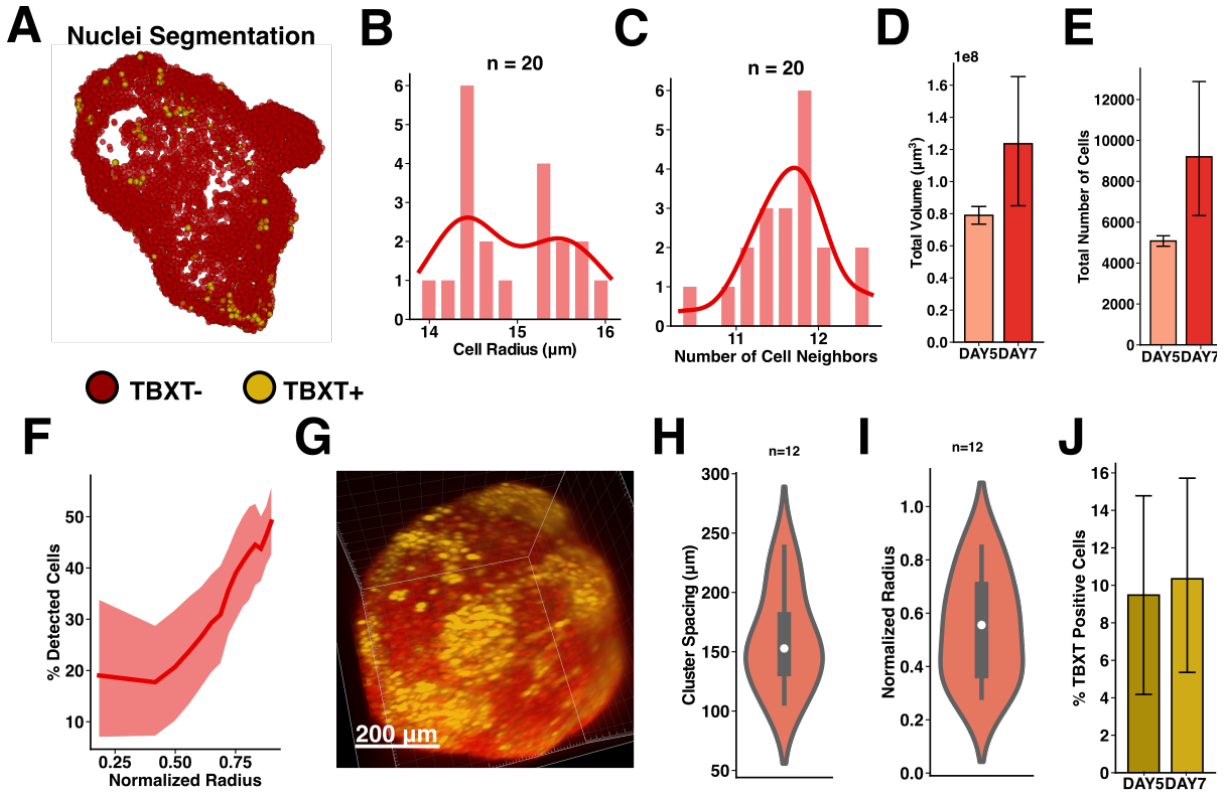
Figure 5.2 3D morphometry of organoids during polarization and elongation
A) Representative segmentation of 3D organoid in light sheet. B) Average cell-cell spacing in 3D CHIR-pretreated organoids. C) Average neighborhood size in 3D CHIR-pretreated organoids. D) Organoid volume at days 5 and 7. E) Estimated total number of cells in organoids at days 5 and 7. F) Percent of detected cells at normalized depth compared to the expected number in an organoid of uniform cell density (0.0 is the center of the organoid and 1.0 is the surface). G) Day 5 organoid with TBXT+ clusters (gold) surrounded by TBXT- cells (red). H) Cluster to cluster spacing, and I) normalized cluster position for aggregations. J) % TBXT+ cells detected in organoids on day 5 and day 7.

To understand the packing of cells in 3D organoids, we quantified cell-cell spacing in day 5 and day 7 light sheet volumes of CHIR treated elongating organoids (n=20 organoids, **Figure 5.2 A**). Cell-cell spacing was comparable to that found in 2D, with mean spacing of 14.9 µm (+/- 0.63 µm), although the distribution was somewhat bimodal with peaks around 14 µm and 16 µm. Cell neighborhood size was found to be 11.6 cells on average, consistent with a random close packing of nearly equal sized spheres (**Figure 5.2 B-C**). Overall organoid volume was estimated by RANSAC estimation of the convex hull of a sample of detected cells, giving measures of the size

178

of the organoid, total number of cells in the organoid, and the radial distribution of detected cells (**Figure 5.2 D-F**). Regional variations in cell density were low at all locations of the organoid, but scattering artifacts deep in the organoid prevented robust detection of cell position deeper than ~100 μm, leading to a drop from approximately 50% cell detection at the surface, to only 20% cell detection near the core. Overall, the consistency of cell spacing with both a random sphere packing model and with the density measurements found in 2D suggest that cell-cell adhesion is similar in the 2D colonies and 3D aggregates.

At day 5, TBXT+ cells formed ~11-12 aggregations throughout the organoid, consisting of 147 cells per cluster, with an average center-to-center spacing of 160 μm (**Figure 5.2 G-H**). The aggregations appeared throughout the organoid (**Figure 5.2 I**), and were marked by streams of TBXT+ cells connecting each cluster, suggesting a mechanism of polarization by chemoattraction of TBXT+ cells to the nearest cluster. This hypothesis was supported by the observation of several day 5 organoids undergoing elongation with a clear polarized TBXT+ pocket in the organoid extension. Approximately 10-12% of all cells were TBXT+ at day 5 and day 7 post-elongation (**Figure 5.2 J**), consistent with flow cytometry results at day 0 and suggesting that the proportion of TBXT+ cells is relatively constant over time.

Although light sheet was important to understand 3D aggregate structure, the limited number of samples, and limited segmentation volume prevented a dense estimation of organoid growth over the entire 10-day differentiation protocol. A set of phase images were collected daily at 24-hour intervals over the entire protocol and then

segmented, producing a large dataset of organoid contours (**Figure 5.3 A**). To estimate

volume, the area of each contour was converted to 3D by inflating each contour using a

cylindrical distance transform, producing a dense time course of organoid volume over

time (**Figure 5.3 B**). Since cell density was found to be relatively constant between days

and between 2D and 3D, cell counts were estimated by multiplying organoid volume by

cell density and then assuming near-perfect sphere packing to estimate the unoccupied

volume. Organoid growth for the first 7 days was exponential, with a doubling time of ~2

days, approximately half the rate of growth derived from 2D (**Figure 5.3 C**). Initial

aggregate size at day 1 was estimated at a median of 1,325 cells (IQR: 1,073 to 1,609),

suggesting that approximately 50% of cells loaded at day 0 were incorporated into the

final aggregate. Median cells per aggregate at day 3 were 1,907 cells (IQR: 1,534 to

2,821) with 3,723 cells (IQR: 3,152 to 5,454) at day 4, consistent with 2D cell tracking

(**Figure 5.3 C**). 4,648 cells (IQR: 3,787 to 5,987) were estimated at day 5 and 8,313

cells (IQR: 6,223 to 10,504) at day 7, consistent with 3D measurements from light

sheet. The distribution of estimated cell number across organoids was heavy tailed, with

an approximately gamma shape (**Figure 5.3 D**) Organoid growth after approximately

day 7 became sigmoidal, suggesting a change in growth pattern which likely

corresponded to the media switch at day 5 and subsequent changes in the nutrient

balance.
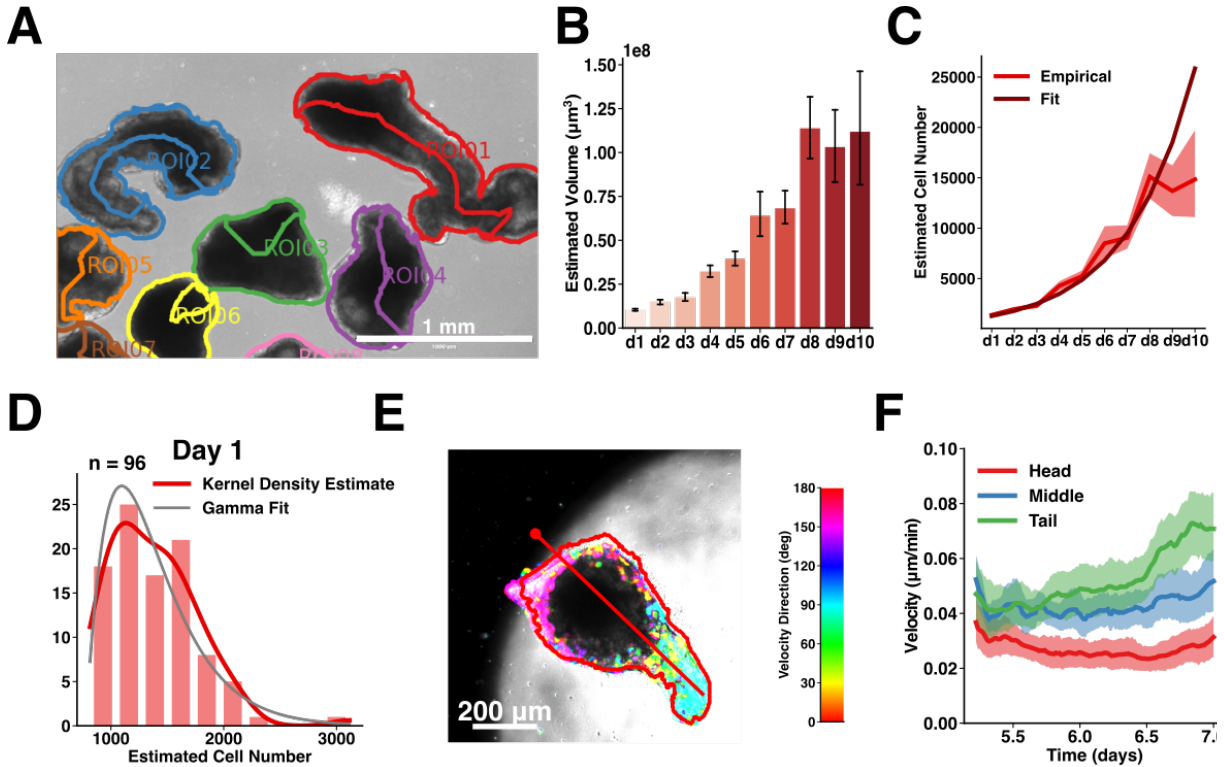
Figure 5.3 Quasi-3D analysis of longitudinal images of organoids
A) Example segmentation of elongating organoids segmented in bright field. B) Estimated organoid volume from day 1 through day 10. C) Estimated cell number per organoid from day 1 through day 10 (light red) and exponential fit of cell growth from day 1 through day 7 (dark red). D) Distribution of aggregate sizes at day 1 (red), with a gamma distribution fit (grey) E) Optical flow analysis of an elongating organoid during extension showing the elliptical fit head-tail axis (red line, dot at head), the organoid segmentation, and the direction of elongation for all points moving at least 0.01 μm/min. F) Velocity magnitude of the head, middle, and tail regions of elongating organoids over 48 hours.

To further complement this dataset, 48-hour time lapse videos of elongating organoids were taken for several different CHIR dosages between days 5 and 7 of differentiation. Although videos were too low resolution for single cell tracking, contours were extracted around each organoid which matched growth curves from the longitudinal imaging. During elongation, motion of particles was detectible in phase at the periphery of elongating colonies in both the head and tail regions. The average speed of these particles was estimated as 0.05 μm/min, much slower than the average velocity of individual cells, suggesting that optical flow was measuring an ensemble

average of many cells, moving with random velocities that tended to cancel out in regions where the aggregates remained round. However, during elongation, the extending regions almost doubled in average velocity to 0.1 µm/min, and the movement of large portions of the tail became aligned in the same direction, suggesting that the elongation phenomenon was at least partially a result of aligned cell migration (**Figure 5.3 E, F**).

Overall these results suggest that dense measures derived from 2D culture replicate key single cell and cell neighborhood behaviors in 3D elongating organoids. Cell-cell adhesion was found to be similar between 2D and 3D bulk populations. TBXT+ cells were retained through at least day 7 of differentiation and maintained at a steady ~10% of overall cell population. Organoid growth was slower than in 2D, but obeyed a similar exponential distribution through at least day 7. Finally, while elongation velocities were much slower for whole aggregates, tail extension occurred at speeds consistent with partially correlated directed migration of cells. These results suggest that dense parameter estimates derived from 2D, with appropriate scaling, can also apply to cell behavior in 3D during at least early stages of stem cell differentiation.

5.3.3 Migration and differential cell adhesion induce simulated symmetry breaking

Although the behavior was not identical, the similarities between 2D and 3D suggested that average cell parameters were well conserved during the critical period between force aggregation at day 0, and polarization and elongation between days 3 and 5. Unfortunately, the relatively large size of elongating organoids, the high opacity

of interior organoid structures, and the lack of a live TBXT reporter precluded direct observation of TBXT+ cell coalescence in 3D. Further, elongating organoids were observed to have differential behavior when embedded in matrix (Matrigel) or neutral gels (alginate or agarose), including loss of 3D organization (i.e. plate down) and failure to elongate to the same extent as unembedded controls. Further, the low throughput of light sheet imaging precluded a large scale, multi-factorial exploration of necessary and sufficient parameters to produce organoid polarization.

Given the consistent estimates of cell adhesion and migration parameters, an agent-based model (ABM) of cell behavior was developed to explore the robustness and stability of organoid polarization *in silico.* Based on the cell behavior data derived from 2D and confirmed in 3D, agents were modeled as the interaction between 4 core forces: cell-cell adhesion modeled as a Leonard-Jones potential, cell migration modeled using a stochastic active/passive migration model, cell growth modeled using a simple stochastic exponential growth rate, and a lumped damping term representing internal viscosity, migration contact inhibition, and cell-ECM interactions. Organoids were modeled as a close random packing of spheres starting from 1,500 cells at day 1 and simulated forward in time for 144 hours until day 7 of differentiation. Simulated and experimental organoids were compared several time points to assess how organoid morphology compared between *in silico* and *in vitro* models of elongation.
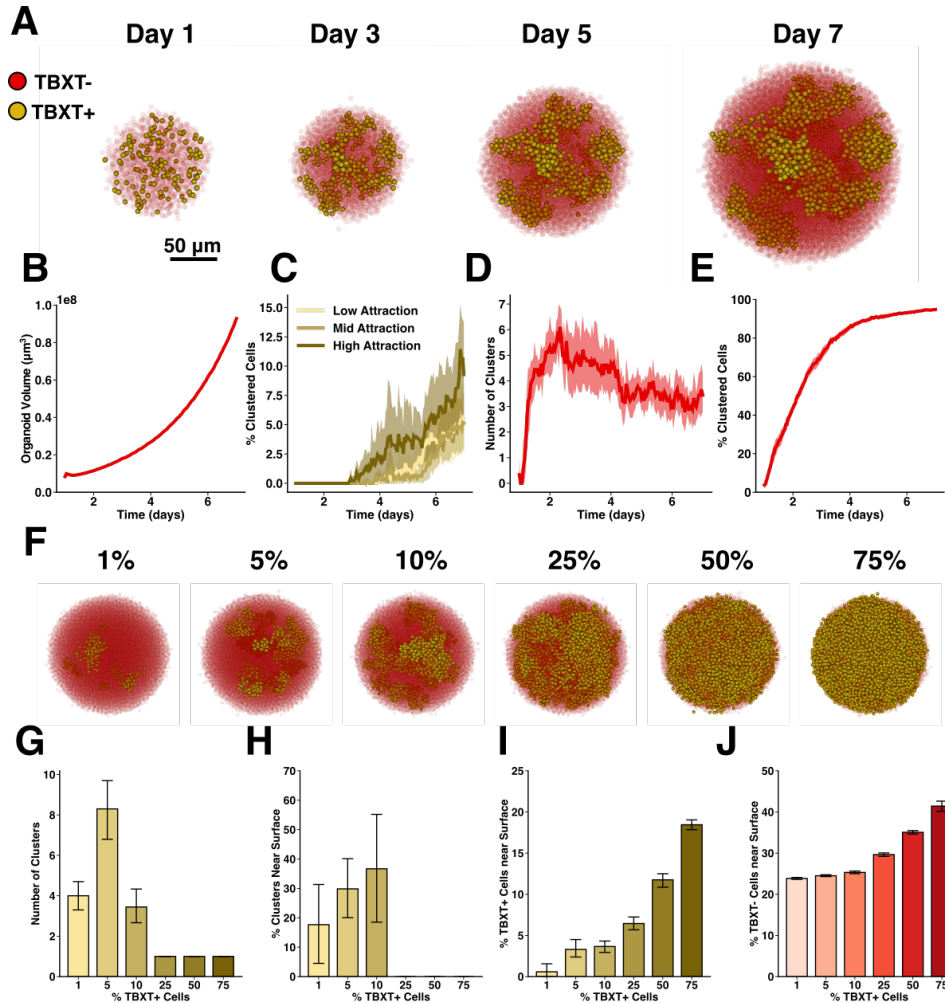
Figure 5.4 Simulations of chemoattractive symmetry breaking
A) Time series of a single simulated aggregate over 7 days. B) Organoid volume growth over 7 days.
C) Mixing in 10% TBXT+ aggregates with low, mid, or high levels of mutual chemoattraction. D)
Number of clusters in 10% TBXT+ high chemoattraction organoids. E) Percentage of TBXT+ cells in a
cluster. F) Clustering behavior of organoids with between 1% and 75% TBXT+ cells with high
chemoattraction. G) Number of clusters at day 7 for organoids with different percentages of TBXT+
cells. H) Percentage of TBXT+ cells in the core of organoids. I) Percentage of TBXT+ cells near the
surface of organoids.

Randomly mixed models were simulated consisting of 10% TBXT+ cells (gold) and

90% TBXT- cells (red) (**Figure 5.4 A**). Cells were simulated with uniform cell behaviors

derived from the bulk cells in the 2D DS+CHIR differentiation, resulting in organoid

growth curves similar to *in vitro* results, and maintaining uniform mixing between TBXT-

and TBXT+ cells over 7 days (**Figure 5.4 B**). A chemoattractive term was added to the

simulation (**Figure 5.4 C**)., where TBXT+ cells underwent biased migration towards the center of mass of the nearest other TBXT+ cells, resulting in clear phase separation at day 3, and formation of ~4 tightly separated clusters at 7 days, with almost all TBXT+ cells contained in a cluster (**Figure 5.4 A, D, E**). The percentage of TBXT+ cells was varied from 1% to 75%, resulting in divergent behaviors (**Figure 5.4 F, G**). At between 1% and 10% TBXT+ cells, 4-8 polarized clusters robustly self-assembled after 5 days, with clusters appearing near the surface of the organoid (20%-40% of the time, **Figure 5.4 H**). Once 25% or more cells were TBXT+ positive, the chemoattractive force still robustly induced clustering, but resulted in formation of a large, centrally located TBXT+ cluster rather than polarized, off-axis clusters, displacing an increasing number of TBXT- cells to the surface (**Figure 5.4 I, J**). The divergence between polarized clusters and more centrally localized inside-outside clusters was reminiscent of the effect of increasing CHIR dosage, where organoids with a low to mid level of CHIR treatment robustly elongated, where organoids with higher CHIR doses formed a central TBXT+ population and did not undergo axial extension.

5.3.4 Surface seeking behavior increases robustness of polarization

Analysis of average cell position within 3D organoids suggested a higher proportion of TBXT+ cells were detected at or near the organoid surface than would be expected, even accounting for losses due to light sheet scattering. Several migratory cell types have been previously shown to undergo Warburg-like metabolic reprogramming during transition to a migratory phenotype (Bhattacharya et al., 2020; Miyazawa and Aulehla, 2018), suggesting that oxygen tension might provide a

"surfacing" signal. Other plausible mechanisms include migration along gradients of chemoattractive morphogens such as FGF8 secreted by a surface population (Sheeba and Logan, 2017), mechanical exclusion from the organoid interior (Winklbauer and Parent, 2017), or destabilization of the TBXT gene expression network for cells too far from the aggregate surface. To test the hypothesis that TBXT "surfacing" contributes to organoid polarization, a second bias term was added to the directional migration calculation, where TBXT+ preferentially migrated towards the surface of the organoid in addition to migration towards the nearest TBXT+ cluster, with the relative strength of the two attraction terms controlled by a weight.
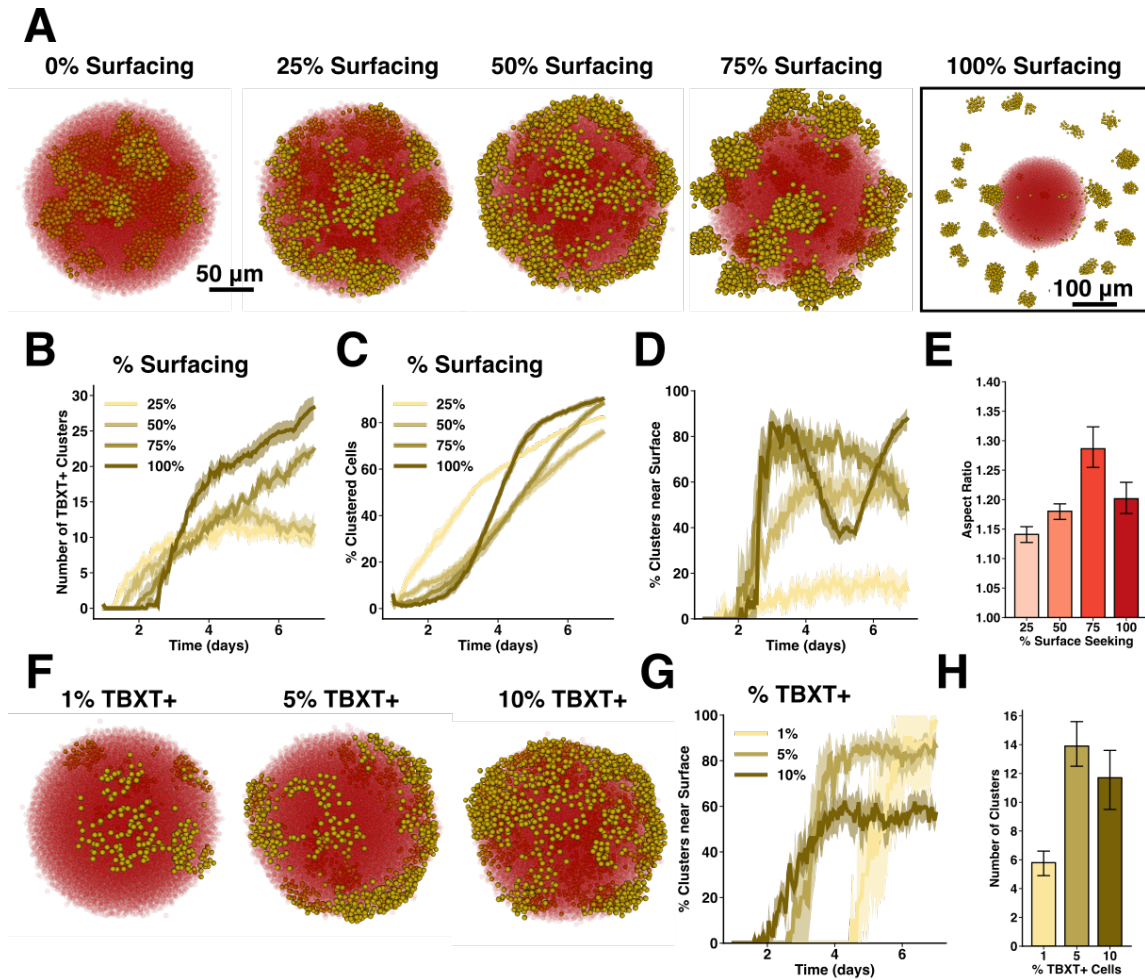
Figure 5.5 Simulation of surface seeking behavior to model organoid polarization
A) Day 7 organoids with 10% TBXT+ cells with different intensity of surface seeking. B) Number of TBXT+ clusters over time for 10% TBXT+ organoids. C) Percent of TBXT+ cells in a cluster over time for 10% TBXT+ organoids. D) Percentage of clusters near the aggregate surface for 10% TBXT+ organoids E) Overall aspect ratio of 10% TBXT+ organoids. F) Morphology of TBXT+ clusters at 1%, 5% and 10% TBXT+ cells with 50% surfacing. G) Percent of clusters near surface at different TBXT+ % with 50% surfacing. H) Number of TBXT clusters with 50% surfacing.

Imposing at least 50% bias towards surfacing resulted in strong agglomeration of TBXT+ clusters at the surface of the aggregate, although >75% bias resulted in cluster fragmentation and 100% bias resulted in phase separation from the aggregate, indicating that surfacing alone was not sufficient to explain aggregate polarization (**Figure 5.5 A-C**). Coalescence of clusters was strongest at 50% surfacing to 50% chemoattraction, with both >50% of clusters at the surface of the organoid and clear

extension of clusters out of the organoid altering the overall aspect ratio (**Figure 5.5 D-E**). More TBXT+ cells reached the surface at lower TBXT+ percentages, but additional clusters formed, demonstrating the tension between surfacing and chemoattraction in forming polarized surface clusters (**Figure 5.5 F-H**). These results suggest that a combination of cell clustering by mutual chemoattraction and polarization through migration towards the organoid surface are sufficient to induce the symmetry breaking and phase separation events underlaying axial polarization in gastruloids.

5.4 Discussion

The complex, interwoven extra-cellular, inter- and intra-cellular processes that enable tissue patterning during morphogenesis are likely not amenable to a tractable mathematical treatment (Glen et al., 2019), and as such can only be approximated by *in silico* methods. However, phenomenological modeling remains a powerful tool for elucidating the minimal set of dynamic components necessary to produce an outcome, and predicting how those components may interact in a biological system. In this chapter, I presented an empirically derived model of organoid symmetry breaking and polarization capable of recapitulating important aspects of axis formation. I showed that cell distributions within microwell aggregates are well explained by random mixing models, and that cell migration and cell-cell adhesion act to maintain this randomly mixed state in the absence of additional forces. Particle-based ABMs were employed to efficiently simulate *in vitro* behavior of cells, and the resulting *in silico* organoids faithfully recapitulated many aspects of *in vitro* structure, shape, and organization. By generating quantitative predictions of organoid formation dynamics, my model offers

testable hypotheses to explain self-organization events, implicating chemotactic and

metabolic signals as likely key components to induce polarization in elongating

organoids.

Particle based models of organoid can be extended to model morphogenic events in

other organoid developmental contexts. Within my model, polarization occurs at the

scale of the entire organoid, but in other organoid models such as neural tube organoids

(Bérenger-Currias et al., 2020) or multilineage cardiac gut organoids (Silva et al., 2020),

the signaling gradients that enable structure formation likely have spatially restricted

ranges and are impacted by compartment boundaries between tissue types. This cell

and organoid tracking system enables quantification of compartmentalized cell

behavior, as well as behavior transitions during differentiation, enabling compilation of a

catalog of morphometric and kinetic cell phenotypes for use in prospective modeling, as

has been previously proposed (Moraru et al., 2008). Further, through quantification of

the impacts of inducible genetic perturbations, morphogen response dynamics, and

optogenetic stimulation, the combination of empirical cell tacking and ABM enables

prospective engineering of synthetic constructs to direct the evolution and patterning of

organoid systems. The combination of phenomenological organoid particle simulations

with machine learning optimization techniques (Briers et al., 2016; Libby et al., 2019)

represents a first step towards the rational design of organoid differentiation, and

ultimately to controlled engineering of *in vitro* tissue.

5.5 Bibliography

Barber, C.B., Dobkin, D.P., and Huhdanpaa, H. (1996). The quickhull algorithm for convex hulls. ACM Trans. Math. Softw. TOMS *22*, 469–483.

Bérenger-Currias, N.M.L.P., Mircea, M., Adegeest, E., van den Berg, P.R., Feliksik, M., Hochane, M., Idema, T., Tans, S.J., and Semrau, S. (2020). Early neurulation recapitulated in assemblies of embryonic and extraembryonic cells (Developmental Biology).

Bhattacharya, D., Azambuja, A.P., and Simoes-Costa, M. (2020). Metabolic Reprogramming Promotes Neural Crest Migration via Yap/Tead Signaling. Dev. Cell *53*, 199-211.e6.

Briers, D., Haghighi, I., White, D., Kemp, M.L., and Belta, C. (2016). Pattern synthesis in a 3D agent-based model of stem cell differentiation. In 2016 IEEE 55th Conference on Decision and Control (CDC), (Las Vegas, NV, USA: IEEE), pp. 4202–4207.

van den Brink, S.C., Baillie-Johnson, P., Balayo, T., Hadjantonakis, A.-K., Nowotschin, S., Turner, D.A., and Martinez Arias, A. (2014). Symmetry breaking, germ layer specification and axial organisation in aggregates of mouse embryonic stem cells. Development *141*, 4231–4242.

Campos, D., Méndez, V., and Llopis, I. (2010). Persistent random motion: Uncovering cell migration dynamics. J. Theor. Biol. *267*, 526–534.

Farnebäck, G. (2003). Two-frame motion estimation based on polynomial expansion. In Scandinavian Conference on Image Analysis, (Springer), pp. 363–370.

Glen, C.M., Kemp, M.L., and Voit, E.O. (2019). Agent-based modeling of morphogenetic systems: Advantages and challenges. PLOS Comput. Biol. *15*, e1006577.

González-Valverde, I., and García-Aznar, J.M. (2018). Mechanical modeling of collective cell migration: An agent-based and continuum material approach. Comput. Methods Appl. Mech. Eng. *337*, 246–262.

González-Valverde, I., Semino, C., and García-Aznar, J.M. (2016). Phenomenological modelling and simulation of cell clusters in 3D cultures. Comput. Biol. Med. *77*, 249–260.

Halır, R., and Flusser, J. (1998). Numerically stable direct least squares fitting of ellipses. In Proc. 6th International Conference in Central Europe on Computer Graphics and Visualization. WSCG, (Citeseer), pp. 125–132.

Hookway, T.A., Butts, J.C., Lee, E., Tang, H., and McDevitt, T.C. (2016). Aggregate formation and suspension culture of human pluripotent stem cells and differentiated progeny. Methods *101*, 11–20.

Keller, R., and Danilchik, M. (1988). Regional expression, pattern and timing of convergence and extension during gastrulation of Xenopus laevis. Development *103*, 193–209.

Libby, A.R.G., Briers, D., Haghighi, I., Joy, D.A., Conklin, B.R., Belta, C., and McDevitt, T.C. (2019). Automated Design of Pluripotent Stem Cell Self-Organization. Cell Syst. *9*, 483-495.e10.

Lou, Y.-R., and Leung, A.W. (2018). Next generation organoids for biomedical research and applications. Biotechnol. Adv. *36*, 132–149.

Ludwig, T.E., Bergendahl, V., Levenstein, M.E., Yu, J., Probasco, M.D., and Thomson, J.A. (2006). Feeder-independent culture of human embryonic stem cells. Nat. Methods *3*, 637–646.

Madhikar, P., Åström, J., Westerholm, J., and Karttunen, M. (2018). CellSim3D: GPU accelerated software for simulations of cellular growth and division in three dimensions. Comput. Phys. Commun. *232*, 206–213.

Marcon, L., and Sharpe, J. (2012). Turing patterns in development: what about the horse part? Curr. Opin. Genet. Dev. *22*, 578–584.

Miyazawa, H., and Aulehla, A. (2018). Revisiting the role of metabolism during development. Development *145*, dev131110.

Moon, R.T., Campbell, R.M., Christian, J.L., McGrew, L.L., Shih, J., and Fraser, S. (1993). Xwnt-5A: a maternal Wnt that affects morphogenetic movements after overexpression in embryos of Xenopus laevis. Development *119*, 97–111.

Moraru, I.I., Morgan, F., Li, Y., Loew, L.M., Schaff, J.C., Lakshminarayana, A., Slepchenko, B.M., Gao, F., and Blinov, M.L. (2008). Virtual Cell modelling and simulation software environment. IET Syst. Biol. *2*, 352–362.

Pantoja-Hernández, J., Breña-Medina, V.F., and Santillán, M. (2020). Reaction-diffusion model for the arrest of oscillations in the somitogenesis segmentation clock. ArXiv200902364 Nlin Q-Bio.

Park, S., Kim, D., Jung, Y.-G., and Roh, S. (2015). Thiazovivin, a Rho kinase inhibitor, improves stemness maintenance of embryo-derived stem-like cells under chemically defined culture conditions in cattle. Anim. Reprod. Sci. *161*, 47–57.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., et al. (2011). Scikit-learn: machine learning in Python. J. Mach. Learn. Res. *12*, 2825–2830.

Płusa, B., and Piliszek, A. (2020). Common principles of early mammalian embryo self-organisation. Development *147*, dev183079.

Qian, X., Song, H., and Ming, G. (2019). Brain organoids: advances, applications and challenges. Development *146*, dev166074.

Scoones, J.C., and Hiscock, T.W. (2020). A dot-stripe Turing model of joint patterning in the tetrapod limb. Development *147*, dev183699.

Sheeba, C.J., and Logan, M.P.O. (2017). The Roles of T-Box Genes in Vertebrate Limb Development. In Current Topics in Developmental Biology, (Elsevier), pp. 355–381.

Silva, A.C., Matthys, O.B., Joy, D.A., Kauss, M.A., Natarajan, V., Lai, M.H., Turaga, D., Alexanian, M., Bruneau, B.G., and McDevitt, T.C. (2020). Developmental co-emergence of cardiac and gut tissues modeled by human iPSC-derived organoids (Developmental Biology).

Takasato, M., Er, P.X., Chiu, H.S., Maier, B., Baillie, G.J., Ferguson, C., Parton, R.G., Wolvetang, E.J., Roost, M.S., Chuva de Sousa Lopes, S.M., et al. (2015). Kidney organoids from human iPS cells contain multiple lineages and model human nephrogenesis. Nature *526*, 564–568.

Tomka, T., Iber, D., and Boareto, M. (2018). Travelling waves in somitogenesis: Collective cellular properties emerge from time-delayed juxtacrine oscillation coupling. Prog. Biophys. Mol. Biol. *137*, 76–87.

Turaga, D., Matthys, O.B., Hookway, T.A., Joy, D.A., Calvert, M., and McDevitt, T. (2020). Single cell determination of cardiac microtissue structure and function using light sheet microscopy. Tissue Eng. Part C Methods.

Turing, A.M. (1952). The chemical basis of morphogenesis. Philos. Trans. R. Soc. Lond. B Biol. Sci. *237*, 37–72.

Turner, D.A., Girgin, M., Alonso-Crisostomo, L., Trivedi, V., Baillie-Johnson, P., Glodowski, C.R., Hayward, P.C., Collignon, J., Gustavsen, C., Serup, P., et al. (2017). Anteroposterior polarity and elongation in the absence of extra-embryonic tissues and of spatially localised signalling in gastruloids: mammalian embryonic organoids. Development *144*, 3894–3906.

Velasco, S., Paulsen, B., and Arlotta, P. (2020). 3D Brain Organoids: Studying Brain Development and Disease Outside the Embryo. Annu. Rev. Neurosci. *43*, 375–389.

Visscher, W.M., and Bolsterli, M. (1972). Random Packing of Equal and Unequal Spheres in Two and Three Dimensions. Nature *239*, 504–507.

Voss-Böhme, A. (2012). Multi-Scale Modeling in Morphogenesis: A Critical Analysis of the Cellular Potts Model. PLoS ONE *7*, e42852.

Wieschaus, E. (2016). Positional Information and Cell Fate Determination in the Early Drosophila Embryo. In Current Topics in Developmental Biology, (Elsevier), pp. 567–579.

Winklbauer, R., and Parent, S.E. (2017). Forces driving cell sorting in the amphibian embryo. Mech. Dev. *144*, 81–91.

6 Discussion and Future Considerations

6.1 Summary

The results presented in this dissertation represent a set of models and analysis pipelines to interrogate the influence of cell migration on pattern formation during early development. In Chapter 2, single cell behavior in mosaic hiPSC colonies and confluent sheets was explored. Cells were demonstrated to mix freely during pluripotency, but gained divergent dynamic behaviors upon differentiation which resulted in phase separation reminiscent of symmetry breaking events *in vivo*. Chapter 3 elaborated on this model, generating a whole colony cell tracking pipeline based on a novel ensemble of neural nets. By segmenting each colony in both space and time, pluripotent stem cell behavior was shown to have a bimodal structure, with central and edge cells exhibiting divergent phenotypes despite remaining universally pluripotent. Interventions to modulate behavioral heterogeneity through colony size, media, and substrate were explored, demonstrating that subtle alterations in pluripotent culture protocols had marked impact on colony structure. During differentiation, cell behavior dynamics were shown to be predictive of both differentiation protocol and cell lineage commitment, with divergent cell behaviors converging on ultimately similar germ layer patterning events. In Chapter 4, a 3D organoid model of symmetry breaking and axial elongation was developed, enabling quantification of the kinetics of a convergent extension-like polarization mechanism, and implicating TBXT+ cells as a key driver of organoid elongation. Finally, in Chapter 5, I applied insights from 2D cell tracking and 3D structural analysis of elongating organoids to propose an *in silico* model of symmetry

breaking and organoid polarization, demonstrating that an empirically derived combination of chemoattractive and surface-seeking migration is sufficient to produce organizer-like structures from an initially randomly mixed 3D population. Together, these results define a critical role for individual and collective cell migration on the formation of 2D and 3D patterns during hiPSC differentiation, and implicate coordinated cell migration as a key organizing principle for inducing macroscale tissue structure during organism development. However, this treatment of cell migration in early development is by no means comprehensive, and so raises additional questions that can be explored on future studies.

6.2 Invasive migration to induce multicellular patterning

Advances in inducible control of gene expression, through systems such as CRISPR interference and activation (Gilbert et al., 2014; Mandegar et al., 2016), optogenetic stimulation (Repina et al., 2019; Sako et al., 2016), and synthetic direct reprogramming (Farber and Qian, 2020; Fernandopulle et al., 2018) provide new tools to induce spatiotemporal patterns of activation and inhibition of gene regulatory networks distinct from classic pleiotropic stimulation from soluble morphogens. However, it is unclear to what extent direct and indirect actuation of gene expression recapitulates the dynamic processes necessary to induce self-assembly of heterogeneous cell constructs. In particular, mistimed expression of multipotent transcription factors (Guye et al., 2016), or incomplete epigenetic reprogramming (Habibi and Stunnenberg, 2017) can lead to abiological intermixing of cell populations and incomplete association of key cell types when forming multilineage tissues. While the biological patterns of cell migration and

tissue compartment formation are well studied during embryogenesis, it is unknown to what extent organoid differentiation recapitulates the formation and chemotaxis of migratory cell subtype pools, and which *in vitro* culture conditions are simultaneously permissive of tissue boundary formation, while still enabling multilineage association to occur. By combining dense cell tracking, mosaic differentiation systems built on CRISPR interference (Libby et al., 2019) or optogenetic control (Repina et al., 2019), and directed differentiation towards innervating (Fernandopulle et al., 2018) or vascularizing (Ng et al., 2016) cell types, it should be possible to systemically screen for conditions which actively favor the invasive migration of neuro- or vascular-progenitors into otherwise homogenous bulk differentiations (e.g. to skeletal muscle or foregut), producing organoids with tissue-like nervous or circulatory systems. Similarly, recapitulating the migration of neural crest cells (Martik and Bronner, 2017) in a mosaic *ex vivo* model of craniofacial development would enable untangling of the complex gene regulatory networks that underlay the formation of the face and skull, and help understand how migratory cells resolve positional cues within complex 3D environments. By better understanding how to optimize for and control cell migration in *in vitro* morphogenesis, we would gain the ability to engineer multilineage tissues that recapitulate the neurulation, vascularization, and germ lineage associations found *in vivo*.

6.3 Organoid Patterning through Coupled Differentiation and Axial Elongation

Convergent extension plays a critical role in patterning tissues during embryogenesis, most noticeably during axial elongation (Perez-Vale and Peifer, 2020), and limb formation (Hopyan, 2017), but also less obviously during internal structure development as in kidney duct elaboration (Lienkamp et al., 2012) or vascular network growth (Kirkegaard et al., 2019). The process of cell extrusion creates a polarized axis, enabling the localization of signaling centers while simultaneously enabling temporal patterning of cell types through a balance of pro-plasticity signals emanating from the bulk and pro-differentiation signals secreted by the axial pole, or vice versa. *In vivo* studies have long demonstrated the instructive properties of transplanted organizers in patterning either total axes in non-mammalian vertebrates (Catala et al., 1995; Ding et al., 2017; Schneider and Mercola, 1999) or partial axes in mammals (Beddington and Robertson, 1999; Tam and Behringer, 1997), which requires the recruitment and coordination of multiple tissue types to form structures such as ectopic heads, spines, or tails. Analogously for *in vitro* organoids, a self-assembling organizer system, such as the one presented in Chapters 4 and 5, could be introduced into non-neural organoid differentiations with semi-permissive conditions at an early stage, and then stimulated at the desired point during differentiation to induce axis formation. By driving a mechanical sorting process as well as polarized signaling reminiscent of the cues provided during segmentation, this organizer-like structure could be used to generate patterned HOX expression in non-neural contexts, enabling study of how tissue positional identity is

established, as well as acting as a source of stratified cell types. Organizer-based macroscale patterning could further be used to study the *ex vivo* mechanics of folding or looping processes, such as those that occur during heart and gut development (Nowotschin and Hadjantonakis, 2020; Silva et al., 2020; Zamir et al., 2003). By providing whole organoid mechanical and morphogenic patterning cues, *in vitro* organizers would represent a step towards true "micro-organs", enabling the creation of globally stratified, multi-lineage constructs with both microscale, mesoscale, and macroscale organization.

6.4 Prospective Simulation for Directed Organoid Design

Although stunning progress has been made in organoid differentiation over the last 10 years, the culture methods and differentiation protocols employed are still empirically derived, leading to anachronistic differentiation of cell lineages and abiological association of multicellular constructs relative to *in vivo* development (Lou and Leung, 2018; Qian et al., 2019; Takebe and Wells, 2019). Heterogeneous cell mixing has been employed to generate organizer populations that lead to spontaneous multicellular patterning (Bérenger-Currias et al., 2020; Rivron and Rivron, 2018), to create poised domains biased towards cell fates (Libby et al., 2018, 2019), or to control cell fate by modulating the cell-cell interactions that propagate differentiation signals throughout the colony (Glen et al., 2018), but recapitulating complex *in vivo* developmental processes requires combinatorial application of multiple of these engineering principles. While the resulting experimental design matrix is too large for naive *in vitro* optimization, *in silico* approaches have been successfully employed to explore the behavioral design space of

multicellular interactions (Appleton et al., 2019; Briers et al., 2016; Libby et al., 2019; Mehdipour et al., 2018), yielding empirically testable predictions of the genetic and behavioral perturbations required to induce cell patterning events.

Unfortunately, measures of cell behavior dynamics and gene regulatory network kinetics have historically been performed in a bespoke, experiment specific manner, although attempts exist to provide more systematic comparisons (Diego et al., 2017; Jiménez et al., 2017; Maiuri et al., 2012). Analogous to chemical design space (Vanhaelen et al., 2020), a catalog of cell lineage, behavior, and relevant gene regulatory networks can be assembled for early hiPSC differentiation protocols. Machine learning models capable exploring large numbers of branching experimental paradigms quickly (Libby et al., 2019; Silver et al., 2016) and optimizing for cooperative and emergent behaviors (Baker et al., 2020; Berner et al., 2019) can then be employed to target either designed multi-cellular behaviors (Briers et al., 2016; Libby et al., 2019) or to evolve novel interactions through autocurricula learning (Baker et al., 2020; Sukhbaatar et al., 2017). Through rapid, *in silico* optimization, the available classes of multicellular assembly processes could be rapidly expanded, and their robustness optimized prior to *in vitro* work, leading to an increase in the complexity and biological fidelity of organoid differentiation. Further, systemization of cell behaviors and lineage transitions may reveal novel general design principles that can be employed to better direct all classes of hiPSC differentiation, providing another step on the road towards rational organoid design.

6.5 Conclusions

The results presented in this dissertation contribute significantly to the fields of developmental biology and bioengineering, creating a set of biological models and analysis pipelines that enable quantification of dynamic pattern formation events during morphogenesis. The development of cell, colony, and organoid tracking pipelines enables multi-scale characterization of the 4D process of lineage specification and multilineage differentiation during gastrulation and axis elongation. Through study of heterotypic cell behavior and mosaic differentiation, this dissertation provides tools to measure the distribution of cell variation and approaches to measure population coalescence and cell-cell interactions in space and time. Finally, through phenomenological modeling of organoid growth and polarization, this dissertation demonstrates a proof of concept mechanism for stochastic self-assembly and polarization in an *in vitro* organizer. These contributions provide a foundation for future improvements in multi-lineage tissue assembly, rational design of stem cell differentiation, and *in silico* models of developmental biology.

6.6 Bibliography

Appleton, E., Mehdipour, N., Daifuku, T., Briers, D., Haghighi, I., Moret, M., Chao, G., Wannier, T., Chiappino-Pepe, A., Belta, C., et al. (2019). Genetic Design Automation for Autonomous Formation of Multicellular Shapes from a Single Cell Progenitor. BioRxiv.

Baker, B., Kanitscheider, I., Markov, T., Wu, Y., Powell, G., McGrew, B., and Mordatch, I. (2020). Emergent Tool Use From Multi-Agent Autocurricula. ArXiv190907528 Cs Stat.

Beddington, R.S.P., and Robertson, E.J. (1999). Axis Development and Early Asymmetry in Mammals. Cell *96*, 195–209.

Bérenger-Currias, N.M.L.P., Mircea, M., Adegeest, E., van den Berg, P.R., Feliksik, M., Hochane, M., Idema, T., Tans, S.J., and Semrau, S. (2020). Early neurulation recapitulated in assemblies of embryonic and extraembryonic cells (Developmental Biology).

Berner, C., Brockman, G., Chan, B., Cheung, V., Dębiak, P., Dennison, C., Farhi, D., Fischer, Q., Hashme, S., Hesse, C., et al. (2019). Dota 2 with Large Scale Deep Reinforcement Learning. ArXiv191206680 Cs Stat.

Briers, D., Haghighi, I., White, D., Kemp, M.L., and Belta, C. (2016). Pattern synthesis in a 3D agent-based model of stem cell differentiation. In 2016 IEEE 55th Conference on Decision and Control (CDC), (Las Vegas, NV, USA: IEEE), pp. 4202–4207.

Catala, M., Teillet, M.-A., and Le Douarin, N.M. (1995). Organization and development of the tail bud analyzed with the quail-chick chimaera system. Mech. Dev. *51*, 51–65.

Diego, X., Marcon, L., Müller, P., and Sharpe, J. (2017). Key features of Turing systems are determined purely by network topology. ArXiv Prepr. ArXiv170809645.

Ding, Y., Ploper, D., Sosa, E.A., Colozza, G., Moriyama, Y., Benitez, M.D.J., Zhang, K., Merkurjev, D., and De Robertis, E.M. (2017). Spemann organizer transcriptome induction by early beta-catenin, Wnt, Nodal, and Siamois signals in *Xenopus laevis*. Proc. Natl. Acad. Sci. *114*, E3081–E3090.

Farber, G., and Qian, L. (2020). Reprogramming of Non-myocytes into Cardiomyocyte-like Cells: Challenges and Opportunities. Curr. Cardiol. Rep. *22*.

Fernandopulle, M.S., Prestil, R., Grunseich, C., Wang, C., Gan, L., and Ward, M.E. (2018). Transcription Factor-Mediated Differentiation of Human iPSCs into Neurons: Rapid differentiation of iPSCs into neurons. Curr. Protoc. Cell Biol. *79*, e51.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Glen, C.M., McDevitt, T.C., and Kemp, M.L. (2018). Dynamic intercellular transport modulates the spatial patterning of differentiation during early neural commitment. Nat. Commun. *9*.

Guye, P., Ebrahimkhani, M.R., Kipniss, N., Velazquez, J.J., Schoenfeld, E., Kiani, S., Griffith, L.G., and Weiss, R. (2016). Genetically engineering self-organization of human pluripotent stem cells into a liver bud-like tissue using Gata6. Nat. Commun. *7*, 10243.

Habibi, E., and Stunnenberg, H.G. (2017). Transcriptional and epigenetic control in mouse pluripotency: lessons from in vivo and in vitro studies. Curr. Opin. Genet. Dev. *46*, 114–122.

Hopyan, S. (2017). Biophysical regulation of early limb bud morphogenesis. Dev. Biol. *429*, 429–433.

Jiménez, A., Cotterell, J., Munteanu, A., and Sharpe, J. (2017). A spectrum of modularity in multi-functional gene circuits. Mol. Syst. Biol. *13*, 925.

Kirkegaard, J.B., Nielsen, B.F., Trusina, A., and Sneppen, K. (2019). Self-assembly, buckling and density-invariant growth of three-dimensional vascular networks. J. R. Soc. Interface *16*, 20190517.

Libby, A.R., Joy, D.A., So, P.-L., Mandegar, M.A., Muncie, J.M., Mendoza-Camacho, F.N., Weaver, V.M., Conklin, B.R., and McDevitt, T.C. (2018). Spatiotemporal mosaic self-patterning of pluripotent stem cells using CRISPR interference. ELife *7*.

Libby, A.R.G., Briers, D., Haghighi, I., Joy, D.A., Conklin, B.R., Belta, C., and McDevitt, T.C. (2019). Automated Design of Pluripotent Stem Cell Self-Organization. Cell Syst. *9*, 483-495.e10.

Lienkamp, S.S., Liu, K., Karner, C.M., Carroll, T.J., Ronneberger, O., Wallingford, J.B., and Walz, G. (2012). Vertebrate kidney tubules elongate using a planar cell polarity–dependent, rosette-based mechanism of convergent extension. Nat. Genet. *44*, 1382–1387.

Lou, Y.-R., and Leung, A.W. (2018). Next generation organoids for biomedical research and applications. Biotechnol. Adv. *36*, 132–149.

Maiuri, P., Terriac, E., Paul-Gilloteaux, P., Vignaud, T., McNally, K., Onuffer, J., Thorn, K., Nguyen, P.A., Georgoulia, N., Soong, D., et al. (2012). The first World Cell Race. Curr. Biol. *22*, R673–R675.

Mandegar, M.A., Huebsch, N., Frolov, E.B., Shin, E., Truong, A., Olvera, M.P., Chan, A.H., Miyaoka, Y., Holmes, K., Spencer, C.I., et al. (2016). CRISPR interference efficiently induces specific and reversible gene silencing in human iPSCs. Cell Stem Cell *18*, 541–553.

Martik, M.L., and Bronner, M.E. (2017). Regulatory Logic Underlying Diversification of the Neural Crest. Trends Genet. *33*, 715–727.

Mehdipour, N., Briers, D., Haghighi, I., Glen, C.M., Kemp, M.L., and Belta, C. (2018). Spatial-Temporal pattern Synthesis in a Network of Locally Interacting Cells. In 2018 IEEE Conference on Decision and Control (CDC), (Miami Beach, FL: IEEE), pp. 3516–3521.

Ng, E.S., Azzola, L., Bruveris, F.F., Calvanese, V., Phipson, B., Vlahos, K., Hirst, C., Jokubaitis, V.J., Yu, Q.C., Maksimovic, J., et al. (2016). Differentiation of human embryonic stem cells to HOXA+ hemogenic vasculature that resembles the aorta-gonad-mesonephros. Nat. Biotechnol. *34*, 1168–1179.

Nowotschin, S., and Hadjantonakis, A.-K. (2020). Guts and gastrulation: Emergence and convergence of endoderm in the mouse embryo. In Current Topics in Developmental Biology, (Elsevier), pp. 429–454.

Perez-Vale, K.Z., and Peifer, M. (2020). Orchestrating morphogenesis: building the body plan by cell shape changes and movements. Development *147*, dev191049.

Qian, X., Song, H., and Ming, G. (2019). Brain organoids: advances, applications and challenges. Development *146*, dev166074.

Repina, N.A., Bao, X., Zimmermann, J.A., Joy, D.A., Kane, R.S., and Schaffer, D.V. (2019). Optogenetic control of Wnt signaling for modeling early embryogenic patterning with human pluripotent stem cells (Bioengineering).

Rivron, N., and Rivron, N. (2018). Formation of blastoids from mouse embryonic and trophoblast stem cells. Protoc. Exch.

Sako, K., Pradhan, S.J., Barone, V., Inglés-Prieto, Á., Müller, P., Ruprecht, V., Čapek, D., Galande, S., Janovjak, H., and Heisenberg, C.-P. (2016). Optogenetic Control of Nodal Signaling Reveals a Temporal Pattern of Nodal Signaling Regulating Cell Fate Specification during Gastrulation. Cell Rep. *16*, 866–877.

Schneider, V.A., and Mercola, M. (1999). Spatially distinct head and heart inducers within the Xenopus organizer region. Curr. Biol. *9*, 800–809.

Silva, A.C., Matthys, O.B., Joy, D.A., Kauss, M.A., Natarajan, V., Lai, M.H., Turaga, D., Alexanian, M., Bruneau, B.G., and McDevitt, T.C. (2020). Developmental co-emergence of cardiac and gut tissues modeled by human iPSC-derived organoids (Developmental Biology).

Silver, D., Huang, A., Maddison, C.J., Guez, A., Sifre, L., Van Den Driessche, G., Schrittwieser, J., Antonoglou, I., Panneershelvam, V., Lanctot, M., et al. (2016). Mastering the game of Go with deep neural networks and tree search. Nature *529*, 484–489.

Sukhbaatar, S., Kostrikov, I., Szlam, A., and Fergus, R. (2017). Intrinsic Motivation and Automatic Curricula via Asymmetric Self-Play. ArXiv Prepr. ArXiv170305407.

Takebe, T., and Wells, J.M. (2019). Organoids by design. Science *364*, 956–959.

Tam, P.P.L., and Behringer, R.R. (1997). Mouse gastrulation: the formation of a mammalian body plan. Mech. Dev. *68*, 3–25.

Vanhaelen, Q., Lin, Y.-C., and Zhavoronkov, A. (2020). The Advent of Generative Chemistry. ACS Med. Chem. Lett. *11*, 1496–1505.

Zamir, E.A., Srinivasan, V., Perucchio, R., and Taber, L.A. (2003). Mechanical Asymmetry in the Embryonic Chick Heart During Looping. Ann. Biomed. Eng. *31*, 1327–1336.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

David Joy

8170EA2048D646D...

Author Signature

12/7/2020

Date