ANALYZING MULTICELLULAR INTERACTIONS: A HYBRID COMPUTATIONAL AND BIOLOGICAL PATTERN RECOGNITION APPROACH

A Dissertation
Presented to
The Academic Faculty

by

Douglas Edward White

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in Biomedical Engineering in the
Wallace H. Coulter Department of Biomedical Engineering

Georgia Institute of Technology
May 2015

ANALYZING MULTICELLULAR INTERACTIONS: A HYBRID COMPUTATIONAL AND BIOLOGICAL PATTERN RECOGNITION APPROACH

Approved by:

Dr. Melissa L. Kemp, Advisor
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Todd C. McDevitt, Advisor
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Manu O. Platt
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Eberhard O. Voit
Department of Biomedical Engineering
Georgia Institute of Technology

Dr. Ron Weiss
Department of Biological Engineering
Massachusetts Institute of Technology

Dr. Steve L. Stice
Department of Regenerative Bioscience
University of Georgia

Date Approved: March 06 2015

To my family, friends, cats, and loved ones without whom I would not be writing this document today

ACKNOWLEDGEMENTS

I would like to acknowledge my two thesis advisors for making my trip through graduate school possible. Todd McDevitt and Melissa Kemp provided me with the freedom to pursue my PhD and forge my own field of research. This freedom as well as the mentoring I received from them was invaluable to my progression as a scientist. Todd had a knack for motivating me even when I thought my results were poor, taught me to see the silver lining in every set of data, and to think about existing results in new and creative ways. Melissa provided a sounding board for all of my harebrained ideas and always made time to listen to me no matter what she happened to be doing when I barged into her office.

I would like to thank my committee members, Dr. Stice, Dr. Weiss, Dr. Platt, and Dr. Voit, for providing in depth feedback on my thesis work as it unfolded. Dr. Stice has been an invaluable collaborator, and the opportunity to work with his student Raymond Swetenburg has been truly eye opening and invigorating. Dr. Platt and Dr. Weiss provided in depth questions about the biology regulating stem cell differentiation as well as crucial questions about biological applications of the thesis work. Dr. Voit was a driving force behind the mechanistic modelling side in an effort to understand specific molecular mechanisms at play during differentiation. This project would also not have been possible without the collaboration with Todd Streelman and his students Jonathan Sylvester and Amanda Ballard. Their work on the cichlid fish system provided a fascinating challenge and a very nice compliment to a lot of the *in vitro* work performed in this thesis.

The McDevitt lab members, past and present, provided an amazing scientific atmosphere to learn, grow and collaborate. I feel so privileged to have seen the old guard, Rich, Andreas, Alyssa, Barbara, and Ken as they finished up there PhD's and continually wowed me with the rigor and creativity of their science. Melissa Kinney helped start me out on my PhD career, and has been a great internal McDevitt lab collaborator who is driven and dedicated. There is no doubt she will make an amazing PI someday soon. The same could be said of my classmate Jenna Wilson, who really stepped up to lead the lab after many of the older students defended. My other classmate Anh is one the hardest workers I have ever seen and I have faith she will continue to do amazing things in her life. The younger graduate students, Denise, Marian, Josh, Katy, Liane, Jessie, Alex, Emily, Melissa G, and Olivia are a fun loving group who will continue on the many McDevitt lab traditions. Even though the lab is now separated by 2139 miles, and three time zones, I have faith that the friendship and comradery of the lab will survive.

The Kemp lab members, though there are only a few left, have formed long lasting bonds even after graduation. Catherine, Gaurav and Adam provided amazing mentorship and were a barrel of laughs. Gaurav and Catherine have provided invaluable resources in my job search, and have never hesitated to help me out when I asked for it. Adam performed amazing work while finishing up his PhD, and it was amazing to watch him balance the birth of his son with the challenges of finishing his PhD on time to get back into his MD studies. Ariel has won almost every scholarship on the planet and continues to be a force for good in the lab, always displaying a positive face no matter what life throws her way. She provided a great deal of support as I finished up my PhD, and was always cheering me up when thesis writing seemed impossible. Then there is

Chad, whose pure taste for verse and poetic rhetoric has kept me laughing for many hours. He is an amazing scientist, a driven individual, and I look forward to seeing what he accomplishes throughout his PhD.

My friends made Georgia Tech feel like a home away from home. Though it would be extremely difficult to name all of them here, each and every of them have provided me with support, great times, and fond memories as we all move on to our next stages in life. It's been an amazing ride to spend 5 years watching every one grow and accomplish so many amazing things, both scientifically and personally. Eli and Chris deserve a special shout out for putting up with my messy shenanigans as a roommate for many years and for all the awesome memories we forged.

My mother, father and sister have all been amazing sources of support during my life. Without them, I would never have been motivated to pursue a PhD. My mom constantly reminded me to never let my curiosity die, while my pushed me to be the best that I could be in everything. My younger sister provided an amazing role model and has been a great source of advice as I move forward in my career.

Lastly my wife Megan, made this whole process bearable. She was my rock, confidant and friend throughout the latter half of my PhD, and I cannot imagine accomplishing this task without her love and support. She even agreed to edit my thesis, even when she was busy studying for medical school. I am so happy that she agreed to marry me which, short of the following thesis document, may be one of my greatest achievements in life.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SYMBOLS AND ABBREVIATIONS

2i	Dual inhibition of MAP kinase and GSK3
2D	2 dimensional
3D	3 dimensional
ABM	Agent Based Modeling
β-Catenin	Cadherin Associated Protein Beta
BMP2/4	Bone morphogenetic protein 4
CPM	Cellular Potts method
DCN	Differentiated cluster number
Diff	Differentiated
dlx3b	Distless Homeobox Gene
EB	Embryoid Body
ESC	Embryonic Stem Cell
ERK	Extracellular signal related kinase
FEA	Finite Element Analysis
FGF4	Fibroblast Growth Factor 4
GMM	Gaussian Mixture Models
GPU	Graphical processing unit
GSK3	Glycogen Synthase Kinase 3
JAK	Janus Kinase
Klf4	Kruppel Like Factor 4
LIF	Leukemia Inhibitory Factor
MAP	Mitogen Activated Kinase
MD	Molecular dynamics
NNK	K-nearest neighbor
ODE	Ordinary Differential Equation
PC	Principal Component
PCA	Principal Component Analysis
PDE	Partial Differential Equation
Smad	Contraction of Sma and Mothers Against Decapentaplegic
pSmad	Phosphorylated Smad
SGD	Stochastic Gradient Descent
Sox2	Sry-box Transcription Factor 2
Stat3	Signal Transducer and Activator of Transcription 3
SVM	Support Vector Machine
UCN	Undifferentiated Cluster Number
Wnt	Wingless

SUMMARY

Pluripotent embryonic stem cells (ESCs) can differentiate into all somatic cell types, making them a useful platform for studying a variety of cellular phenomena. Furthermore, ESCs can be induced to form aggregates which recapitulate the dynamics of development and morphogenesis. However, many different factors such as gradients of soluble morphogens, direct cell-to-cell signaling, and cell-matrix interactions have all been implicated in directing ESC differentiation. Though the effects of individual factors have been investigated independently, assaying combinatorial effects has proven inherently difficult due to the spatial and temporal dynamics associated with such cues. Dynamic computational models of ESC differentiation can provide powerful insight into how different cues function in combination, both spatially and temporally. By combining particle based diffusion models, cellular agent based approaches, and physical models of morphogenesis, a multiscale rules based modeling framework was created to investigate complex regulatory cues which govern complex morphogenic behavior in 3D ESC aggregates. The objective of this study was to examine how spatial patterns of differentiation by ESCs arise as a function of the microenvironment. The central hypothesis was that heterogeneity associated with soluble morphogens and cell-cell signaling leads to complex spatial patterns in ESC aggregates.

To test this hypothesis, a computational modeling framework capable of modeling diffusive soluble gradients and cell-cell interactions in multicellular aggregates was developed. Agent based modeling (ABM) captured complex spatio-temporal patterns associated with ESC aggregate differentiation, while a mass spring description accurately captured dynamic cellular behaviors such as movement and division. This served to

approximate the physical properties of both the cells and the aggregate, while providing a means to keep track of local cell-cell interactions in a network structure. Soluble interactions via paracrine/autocrine signaling mechanisms were tested within a robust lattice-based diffusion solver by modulating parameters associated with soluble and local cell-cell interactions.

One major issue with agent based modeling of cellular aggregates is quantifying the spatial pattern of differentiation in a manner that is amenable to experimental validation. To address this shortcoming, a novel spatial pattern recognition system was developed which utilized network theory to extract meaningful spatial metrics from networks in which cells are represented as nodes, and the connections between cells are edges. Images were converted to networks using a combination of image analysis algorithms and custom code to create digital networks. By describing the experimental images and computational results in a common space, quantitative comparisons between spatial images and computational models were performed. This technique uncovered a putative paracrine mechanism which could explain size-dependent differences in differentiation of ESC aggregates.

To further probe the molecular mechanism which could lead to pattern formation in early differentiation of ESC aggregates, a multiscale stochastic ordinary differential equation (ODE) system was constructed. This model included mechanisms for soluble LIF signaling and FGF4 signaling, which are produced by ESCs during the differentiation process. The stochastic ODE based model created in this study captured a wide range of spatial patterns, and showed that FGF4 signaling and inherent stochasticity associated with expression of the pluripotency associated transcription factor Nanog play

a key role in modulating spatial pattern associated with differentiation. This model provides key insight into why cellular aggregates cultured in different conditions may display different spatial differentiation signatures.

This study further demonstrated the modulatory of network based spatial pattern classification and tested a variety of input image types (2D images, 3D confocal stacks, and 2D histological sections) and a wide range of model systems (mesenchymal and neural differentiation in ESC aggregates and gastrulation in cichlid fish). In the context of cichlid fish, this technique was able to segregate fish into different stages of gastrulation based on spatio-temporal differences in protein and RNA expression. Network derived metrics significantly improved classification of mesenchymal phenotypes, and enabled comparisons across different experiments to describe patterns associated with mesenchymal phenotypes in histological samples. Furthermore, this approach predicted novel feedback mechanisms which can help explain the switch from motor neuron production to oligodendrocyte production (the glial switch) in neural differentiation.

This work represents the first attempt to understand emergent dynamic differentiation patterning that result from integration of multiple cues governing ESC aggregate morphogenesis in 3D via computational modeling strategies. Furthermore, this network approach represents a significant and novel advance in the field of pattern recognition and quantitative biology as the first pattern classification platform which utilizes single cell spatial information and modularly compares across multiple systems of interest.

CHAPTER 1 INTRODUCTION

1.1 Motivations for Research

Pluripotent embryonic stem cells (ESCs) have the unique ability to differentiate into cells of the three germ lineages that form all of the tissues and organs of a mature organism. Differentiation of pluripotent ESCs can be induced *in vitro* via a variety of existing approaches to emulate aspects of the developmental program. One of the most widely used techniques relies upon the formation of multicellular aggregates composed of undifferentiated ESCs in suspension culture[1, 2], that spontaneously induce the differentiation of ESCs within the 3D aggregate[3, 4]. Due to the fact that ESC aggregates mimic the physical structure and cellular composition of the morphogenic embryonic microenvironment, they have been used to study aspects of development *in vitro* as well as the formation of primitive tissue complexes[3-5]. Despite the utility of the approach, robust methods to control ESC aggregate differentiation *in vitro* remain limited due to an incomplete understanding of the complex interactions within the 3D multicellular aggregates that mitigate cell fate decisions[6, 7].

Considerable efforts have been made to ascertain the role of individual components of the cellular microenvironment in regulating cell fate decisions. The extent to which cell-cell communication via paracrine[8, 9], autocrine[8, 10-14], or direct contact signaling[12-14] enhances or inhibits differentiation has been investigated in various contexts. Exogenous manipulation has been used to control differentiation by the addition or removal of various soluble factors in a temporally regulated manner in an effort to mimic the sequence of morphogenic cues. Although exogenous factors have proven necessary for the *in vitro* and *in vivo* maintenance or differentiation of ESC

populations, they are not the only factors regulating stem cell behaviors. The biochemical composition of the cellular microenvironment [8, 15] and extracellular matrix (ECM) [16-18] have also been implicated in the regulation of cellular niches. In addition, the mechanics and physical properties of the microenvironment can also impact cell phenotype [19]. Given that cell fate transitions occur in complex environments where biochemical and physical cues coexist, elucidating the role each of these combinatorial factors via experimental studies alone remains a significant challenge. Therefore, new approaches that allow systematic investigation of combinations of parallel factors that regulate stem cell differentiation are needed to more accurately predict cell response to complex micro-environmental cues.

In many instances, computational modeling strategies have been successfully used to identify mechanisms governing integration of complex signals that direct cell fate decisions and correctly predict the resulting phenomena. Depending on the desired resolution of the system, ordinary differential equations (ODEs) can be used to model a variety of stem cells processes; including, but not limited to, cellular signaling events[20-22], protein interaction networks[23], and genetic regulatory networks[24]. Alternatively, to model the structure of cellular aggregates[25, 26], cellular division and tissue formation[27-33], and pattern formation in biological systems[32, 34], agent based modeling has been applied to overlay rules based with physical modeling approaches[35]. Moreover, agent based models have been used to investigate dynamic processes of multicellular systems, such as morphogenesis[28, 36] and formation of physical tissues[37]. Due to their flexibility agent based models provide an attractive, lightweight platform capable of integrating spatial constraints, soluble cues, and cell-cell interactions

in addition to investigating rule sets which could lead to complex spatial patterns. This study uses computational rules based modeling to assess the macroscopic principles regulating stem cell pluripotency while studying the spatial patterns associated with differentiation in complex 3D microenvironments.

Specific Aim 1. Create computational models of 3D ESC aggregate differentiation. The *working hypothesis* is that simple rules governing ESC differentiation would give rise to spatial pattern formation. An agent based computational model (ABM) coupled with a mass spring approach was used to capture ESC aggregate physical properties and growth dynamics. The model predicted that simple cell-cell interactions could produce a wide variety of spatial pattern dynamics. Qualitative validation of the model was performed by comparisons with spatial patterns of the pluripotency marker Oct4.

Specific Aim 2. Develop quantitative descriptions for spatial pattern dynamics occurring in ESC aggregates. The *working hypothesis* was that if cells and their connections were represented as networks, then network theory could be applied to extract quantitative metrics describing spatial information. The rational was to improve computational modeling strategies by directly comparing the networks with experimental data. Spatial network analysis was able to derive quantitative descriptions of spatial patterns associated with loss of the pluripotency transcription factor Oct4; moreover, direct comparisons between model and experiment identified a possible paracrine mechanism capable of explaining size-dependent differentiation kinetics in ESC aggregates. In addition, the modularity and robustness of spatial network analysis was demonstrated by applying it to multiple biological systems (mesenchymal and neural

differentiation in ESC aggregates, as well as gastrulation in cichlid fish), and input modalities (2D images, 3D confocal data and histological sections).

Specific Aim 3. Investigate the role of heterogeneous regulation of the pluripotency associated transcription factor Nanog as a source for spatial pattern formation in ESC aggregates. The *working hypothesis* was that Nanog regulation via inherent stochasticity and FGF4 regulation will lead to spatial pattern formation. A multiscale model incorporating the ESC transcriptional regulatory network and secretion of soluble factors like LIF and FGF4 described a wide range of biological pattern phenomena, and ultimately predicted differentiation patterns associated with Nanog and Oct4 which match those observed in ESC aggregates.

This work is *innovative* because it represents one of the first attempts to computationally dissect how different soluble and local cell-cell cues interact to influence spatial pattern evolution. The spatial network analysis described here is a significant advance in the field of biological spatial pattern recognition as it integrates single cell information into tissue level descriptions of pattern formation. Furthermore, this approach is poised to take advantage of advances in single cell tracking via 3D light sheer microscopy to ultimately describe biological pattern evolution at a cellular level.

CHAPTER 2 BACKGROUND

2.1 Embryonic Stem Cells

2.1.1 Embryonic Stem Cell Culture

Embryonic stem cells (ESCs) have the unique ability to differentiate into all different somatic tissues found in the adult organism, making them important tools for regenerative medicine as well as drug design and discovery. ESCs were first derived in mice[38], followed by lines derived rabbits[39], primates[40] and finally humans[41]. The process by which these lines were derived varied initially, but was standardized to one procedure in which the inner cell mass was extracted from a developing embryo at the blastocyst stage[42]. Based on ESCs differentiation capacity the field sought to induce differentiation towards targeted cell types *in vitro*. Initial differentiation of ESCs centered around cardiac[43-45] and neuronal differentiation[46, 47]; investigation into the mechanisms governing these processes is still an active field of research today[48-54]. Eventually techniques were discovered for differentiating ESCs into other tissues including oocytes[55], liver[56, 57], pancreatic[58-60] often driven by the need for a source of cells for experimental testing and drug screening[61]. One main problem with these cells, however, was that before implantation, the cells had to be differentiated to a high purity, otherwise the formation of teratomas or tumors from undifferentiated ESCs occurred upon implantation [62, 63]. To solve this problem, differentiated cell populations were purified by fluorescent activated cell sorting (FACS), magnetic activated cell sorting (MACS) or other methods to eliminate pluripotent ESCs in the sample [64-66].

In addition to determining the cues that initiate stem cell differentiation, researchers were also interested in what factors were necessary for maintaining ESC pluripotency. These investigations led to the discovery of several key genes involved in regulating stem cell pluripotency. Though the web of interactions has grown quite large, the core regulatory transcription factors have been identified as Oct4, Sox2 and Nanog [67-74]. Particularly Oct4 has been the focus of many investigations as loss of Oct4 is concomitant with the loss of stem cell pluripotency which puts ESCs into a primed differentiation position [75-77]. Nanog has been implicated as the first key differentiation event in this sequence, but is not directly correlated with the loss of pluripotency transition, and appears to undergo stochastic fluctuations the causes of which are largely unknown[78, 79]. Efforts have been made to find molecules or pathways which regulate these core transcription factors. In the case of mouse embryonic stem cells, factors like leukemia inhibitory factor (LIF), which regulates pluripotency via JAK/Stat3 signaling[67], and FGF4 signaling which regulates Nanog via MAK/ERK have established roles in modulating pluripotency[23, 78]. However, others like while Wnt which activate downstream gene expression via β-Catenin/TCF can either induce differentiation [80] or maintain pluripotency [74] depending on their context. Though the mechanisms of action of these specific molecules have been elucidated, the question remains how well ESCs in a 2D culture environment capture ESCs in a developing embryo.

This question led to the investigation of the stem cell niche, which in the case of embryonic stem cells involves an investigation of the embryo. Some of the earliest observed phenomena in development include the development of gradients of morphogens such as Wnt and BMPs and their respective inhibitors; thus differentiation

protocols were derived for stem cells using additions of similar morphogens to the medium. However, there are other factors which can regulate cell fate such as cell matrix interactions, physical properties of the environment, and direct cell-cell interactions. These can be difficult to study fully in a 2D context, particularly if one is interested in how these processes occur during natural development. On method to examine how ESCs function in a 3D cellular environment, is to create aggregates of ESCs via various strategies.

2.1.2 Differentiation of Embryonic Stem Cells as 3D Aggregates

To understand how pluripotent stem cells behave and interact with their in vivo environment, it is desirable to construct a structure which more closely matches the native embryo. To accomplish this objective, aggregates of either human or mouse embryonic stem cells (classically termed embryoid bodies (EBs)) are formed. Various different methods exist to form ESC aggregates, among them hanging drop[81, 82], spontaneous aggregation induced in stirred fluid reactors[83], formation via microbeads[84], and preformation in microwells[85]. While each of these technologies has benefits, recent advancements with preformation of ESC aggregates in microwells using the AggreWell technology allows the formation of large numbers of aggregates of uniform size due to control over initial cell seeding density [85, 86]. In addition to the formation methodologies, methods for culturing aggregates differ as well. Static culture is commonly used as the cells do not experience hydrodynamic effects; however, in bulk culture configurations, ESC aggregates will often agglomerate introducing mass transfer limitations and making it difficult to track how differentiation is progressing within a single aggregate. Hanging drop methods avoid this problem, but are inherently lower

throughput. Alternatively ESC aggregates can be cultured in suspension which prevents them from aggregating, but introduces hydrodynamic forces[83, 86].

Differentiation of ESCs has been extensively studied in ESC aggregate systems and differentiation into all three germ layers: mesoderm, endoderm, and ectoderm has been demonstrated [4]. In some cases the derivation of trophectoderm (trophoblast) has been reported [87]. Several studies have been carried out which look at the spontaneous differentiation of ESC aggregates towards various different phenotypes including osteoblasts[88], primitive endoderm[89-92], and hepatocytes [93, 94]. In the case of most of these finds, the aggregates are placed in a medium which is permissive towards the germ lineage of choice, and then the specialized cell types are isolated[88, 95, 96]. Additional schemas for more homogenous and robust differentiation have been derived for the cardiac[97-99] and neural[100] lineages; however, even in these systems it is difficult to know when to add factors, or to what extent the spatial localization of these factors is important. The mode of delivery in spatial terms can dramatically affect the resulting morphological outcomes [101]. Often such protocols are optimized via brute force, which can be an expensive and time consuming proposition depending on the number of factors, highlighting the need for more robust approaches; particularly the precise spatial and temporal control of morphogens, and also a configuration for rapidly testing the various potential combinations of morphogenic factors.

2.1.3 The Role of Nanog, Oct4, and Sox2 in Maintaining Stem Cell Pluripotency

As mentioned previously the transcription factors Oct4, Sox2 (Sry-box Transcription Factor 2) and Nanog have all been implicated in a web of several transcription factors necessary for maintaining pluripotency[102, 103]. While the roles of Oct4 and Sox2 are

largely known in ESCs[104-106], Nanog has been the subject of a variety of studies investigating its intrinsic variability [68, 102]. In addition to its already complex regulation, Nanog is thought to be bi-allelically expressed in ESCs in LIF conditions[68], but not during culture in the 2i conditions (an alternative method for maintaining pluripotency in which GSK3 and MEK inhibitors are supplied in culture)[102]. Furthermore, there are several theories for how Nanog is regulated during pluripotency, with some indicating a positive feedback [107], and others indicating auto-repression [108]. This inherent difference in opinion is due to the intrinsic stochasticity associated with Nanog expression. In monolayer culture, it is suggested that this heterogeneity is due to local expression of FGF4 which inhibits Nanog expression via the MAP kinase pathway. In cultures where a MAP Kinase inhibitor is used (part of 2i) the heterogeneity is decreased substantially, and the resultant ESC colonies all express Nanog at a high level in a more deterministic manner [102, 109]. This suggests a similar role for FGF4 signaling in ESCs as in the developing embryo where it helps to prime linage development and early differentiation events [110-113].

As mentioned previously Oct4 and Sox2 have been substantially studied in the context of pluripotency. Sox2 and Oct4 have similar architectures, the two transcription factors can form homo/heterodimers and positively feedback on their own transcription. Furthermore, β-Catenin is thought to associate with Oct4 and induce Oct4 expression possibly through Wnt-mediated signaling pathways, suggesting yet another local paracrine mechanism for Oct4 control[105]. However, the pluripotency network does not just consist of Sox2, Oct4 and Nanog. Network analysis and genome wide screens have revealed a whole core host of candidate genes including Rex and Klf4 which all work in

tandem to regulate this state[114]. Klf4 has been shown to transduce LIF mediated Jak/STAT3 signaling by directly influencing Oct4 levels [115, 116]. It is the combination of both local paracrine signaling pathways with the core pluripotency transcription network that creates such dynamic and switch-like responses during the transition of ESCs out of the pluripotent state.

2.2 Computational Modeling

2.2.1 Modeling Cellular Systems via Ordinary Differential Equations

Computational models have been used in the past to elucidate various different levels of cellular interactions. The interesting processes to model depend on the scale; communication between multiple cells requires discrete spatial information about their locations while individual cells operate based on transcriptional networks and signal transduction pathways. Systems of ordinary differential equations (ODEs) are often used to model metabolic networks[117-119], cell signaling networks[120-122], and gene regulatory networks[123, 124] at the single cell or population level. In the particular context of stem cells ODEs have modeled the pluripotency network[125], differentiation[126], and population dynamics[127, 128]. Deterministic ODEs lead to only one solution for a given set of parameters; however stem cell differentiation is driven by complex gene expression networks which hold inherent stochasticity [129]. Thus, stochastic ODEs have also been used to investigate transcriptional dynamics and networks of populations of ESCs [130-132]. A general weakness of ODE approaches is the large number of parameters necessary for the construction of such models and the inherent complexity of biological systems. Furthermore, ODE systems only capture how

the system will change with time, and are not capable of describing specific spatial behaviors. This makes them a great tool for capturing the internal single cell dynamic information, but insufficient at predicting how phenotypic changes modulate intercellular communication in complex three dimensional environments.

To counteract these shortcomings one could link individual cells with external inputs dictated by the environment. Paracrine signaling can link cells over longer distances, but requires a model of diffusion. The classic approach to modeling diffusion utilizes partial differential equations (PDEs). Simple diffusion consumption equations have been used to model nutrients uptake in tumor spheroids, and mass transfer limitations in ESC aggregates [133]. Additionally, consumption and production terms combined into reaction-diffusion equations have been used to model autocrine/paracrine signaling[134, 135] and heterogeneity in intracellular signaling [136]. In the context of diffusion, various different algorithms exist for the solution to complex reaction diffusion systems including: finite element[137], lattice Boltzmann[138], and particle based methods [139, 140]. Finite element analysis is excellent for systems in which the size of the mesh remains small and regularly spaced. However, in the context of large cellular environments, this can be difficult when one must consider the diffusion of species primarily around cells. The lattice Boltzmann implementation can solve this problem utilizing an irregular finite-element scheme for lattice creation [138, 141]. However, it can be difficult again to get a fine enough mesh size as and an exact geometric representation of the cells is necessary to establish the appropriate boundary conditions. Additionally, when the cells are allowed to move this it can be difficult to keep track of the resulting displacement of molecules. Particle based methods can be used to counter

these issues; however, these algorithms are computationally intensive, which can make them difficult to link into higher scale models [139].

2.2.2 The Cell as an Entity - Agent Based Modeling Approaches

As cells can all be considered autonomous entities, another approach to modeling cellular systems is to use agent based modeling (ABM). ABM is a technique used to break a problem down into individual autonomous parts (or agents) which act based on a set of predefined rules and is considered to be an extension of the cellular automata method. This technique is used not just to model cellular systems[142], but has also been used in ecology[143] , economics[144], and psychology[145]. ABM has also been used in the context of cells to model macrophages[146], angiogenesis[147], epithelial interactions[148], and cancer development [149-151]. It is commonly used in combination with multiscale modeling techniques which seek to describe cellular behavior across multiple different scales [148, 149, 152]. In the context of stem cells ABMs have been applied in a scant number of cases, particularly focusing on the derivation of previously unknown gene interaction networks. As mentioned previously, the ability to examine how structural features of the stem cell niche influence the spatial patterns associated with loss of pluripotency is attractive for studying differentiation in 3D aggregate systems. ABM provides a powerful tool which can be leveraged towards understanding the combinations of factors governing stem cell differentiation.

2.2.3 A New Way to View Network Modeling

In the context of ABM, a methodology for conveying information from one cell to another is needed. Network representations are used to represent the flow of information from one node to another. For cells, this technique is typically used to represent proteins

[153-155] and genes [156-158]. Oftentimes, these interactions are not known explicitly, and the goal of the modeling is to determine what the most likely set of interactions between species can explain a set of data [121, 159]. Bayesian network analysis can infer probabilistic connections between nodes in a network [160-162]. This approach has been used to infer the regulatory networks associated with differentiation of stem cells[126], and pluripotency [163]. However, there is another context in which network theory can be useful. When the structure of a network is known, it can be used to convey information between nodes. The nodes can be programmed to accept this information in a certain way, and this potentially could be used to represent integration of cellular communication at individual cells or nodes. Furthermore, previously networks based motif analysis was useful in examining biological interaction networks [125, 164-166]. A natural extension of this approach is to annotate different phenotypes within networks to extract meaningful spatial and temporal information about the network architecture. This represents a novel and insightful application of network theory to biological pattern identification.

2.2.4 Modeling Cellular Interactions in a Relevant 3D Context

Modeling 3D structures and their resulting interactions in complex environments is applicable to multiple different scientific and industrial applications making it an active area of research in computer science. As a starting point, objects can be modeled according to physical laws such as Newton's Laws, or Brownian motion. The difference between how objects are simulated is primarily a function of what information is desired from the simulation, and on what scale the simulations need to be performed. Simulations of large objects can be accomplished via Newtonian mechanics, while simulations of

molecular interactions can be carried out using molecular dynamics (MD)[167]. In the simplest case of rigid bodies a simple physics engine can suffice and many such engines have been created (Open Dynamics Engine, Open Tissue, Jitter) though most are primarily used in game physics (PhysX, Havok, Orge). However, if the goal is to model individual atoms or molecules and their interactions then MD simulations are necessary. While MD is exceedingly accurate it is still computationally intensive, though recent parallelization utilizing GPU architecture has provided some improvement [167-169]. Conversely, the physics models are fast and can model lots of agents interacting in space, which makes them attractive for agent based models. However, cells are deformable objects, and using a purely physics based or MD method to model deformation can be prohibitively computationally expensive.

Consequently, three main methods have emerged for calculating deformation in cellular structures: finite element analysis (FEA), mass spring systems, and cellular autonoma methods. Again, the methodology used here is primarily a function of the desired output of the model. FEA has been used to model physical properties of a single cell[170], tissue[171], or regions of the human body[172, 173]. In all cases the primary goal is to model the deformation and physical stress/strain relationships that these objects observe. Though FEA analysis provides exceedingly accurate results for a single mesh being tested, collisions between multiple meshes or objects dramatically increases the complexity of the model. In contrast cellular autonoma methods operate on predefined lattices and are suited to modeling large numbers of interacting deformable objects. The newest version of this model, called the cellular-Potts method (CPM), relies on the minimization of several discrete energy terms computed on a predefined square lattice.

This technique has been used to successfully model cellular sorting and adhesion, multicellular aggregates, and cancer cell migration [174-177]. One of the main drawbacks of the CPM approach is that the energy terms utilized to drive this procedure are not easy to measure, and can be seen as arbitrarily set to achieve a certain behavior. Mass spring technologies are an amalgamation of CPM and FEA, with defined masses connected by physical springs, but whose overall spring behavior can be governed by energy optimization parameters. However, due to oscillations and instabilities inherent in these systems, much effort has been put into their optimization algorithms to ensure numerical convergence[178] [179]. Classically mass spring systems have been used to model inanimate objects, but has been used to model soft tissue deformations[179]. Additionally, technologies for the parallelization of such systems to decrease computational time via GPU acceleration also exist [180-182].

2.3 Spatial Pattern Recognition

2.3.1 Classical Applications

Pattern recognition is a broad field devoted to the study of the formation and classification of various patterns. In the computational field, patterns are input into various algorithms whose goal is to assign the patterns to various different classes. These algorithms are often accomplished using a machine learning approach, in which a set of defining features capable of distinguishing patterns from one another are slowly developed [183-186]. Pattern matching is used in many different applications; one of the most famous being facial recognition [187, 188]. Such machine learning approaches include, but are not limited to, support vector machines (SVM)[189], decision trees, neural networks, and Bayesian approaches[190]. While these represent the base

algorithms, many other hybrid methods such as stochastic gradient descent (SGD)[191], gradient tree boosting[192], and random forests[193] are commonly used for more complex pattern classification schemes. Most of these algorithms support classification into multiple different pattern classes, but each applies a slightly different approach for multiclass pattern classification. Often multiclass classification is accomplished using either a one vs. one or a one vs. all approach. In one vs. all a single classifier is fit per each pattern class. Thus, to fit a given pattern to a class it is assigned a score from each classifier. This approach is both computationally efficient, and easy to interpret, but does not scale well with large amounts of data or many different pattern classes. In contrast the one vs. one technique constructs a classifier for every pair of classes. Thus a lot more classifiers are constructed (number of classes)*(number of classes -1)/2 making this method generally slower than the one vs. all approach. However, it has advantages when working with large data sets as it uses a smaller subset of the data for training purposes.

2.3.2 Biological Applications

In the context of biology, pattern recognition has been used in a wide variety of systems ranging from histopathology, to whole organism classification. With the advent of higher power microscopy and confocal based approaches, recently there has been a push towards classification and analysis of single cells. One general approach is to use high content imaging or other analogous platforms to image single cells at extremely high resolution, and then extract sets of quantitative metrics describing their shape or distributions of proteins within the cell [194-196]. While these approaches are relegated to the single cell level, many have addressed issues related to single cell identification in tissues via nuclear segmentation algorithms which take advantage of machine learning components [197-208]. Many of these approaches are devoted to the field of

histopathology, where segmentation and automatic identification of cells provides a crucial step towards automated tumor detection methods [197, 202, 203, 207]. These algorithms are now sophisticated enough to segment nuclei in a variety of image types, and many open source platforms have started to incorporate such approaches into their frameworks (IMARIS, Matlab, Cell Profiler, and Illitask). With the advent of such algorithms, another popular technique called image-cytometry has emerged which allows the extraction of protein or RNA species levels on a per cell basis [102, 105]. While this approach has certainly been a useful single cell analysis tool in the context of stem cells, it does not take into account spatial information. Other tissue level approaches have used a variety of machine learning algorithms to classify patterns where the scale has ranged from classification of patterns in local tissues[209-212], to whole embryos/organism classification in the case of *Caenorhabditis elegans* (*C. elegans*) [213-216], and drosophila [200, 217-220]. While classification paradigms exist at both the single cell, and tissue levels, the field has yet to utilize the information inherent in single cell approaches to perform tissue level classifications.

CHAPTER 3 COMPUTATIONAL MODELS OF 3D MULTICELLULAR STEM CELL AGGREGATES[1]

3.1 Introduction

Embryonic stem cells (ESCs) can differentiate into cells of the three germ lineages: mesoderm, endoderm and ectoderm. Multicellular aggregates composed of undifferentiated ESCs in suspension culture [221, 222], provide an enticing platform to study complex morphogenetic events [223-227]. Despite the recent success of this approach in generating complex tissues, robust methods to control aggregate differentiation *in vitro* remain limited due to an incomplete understanding of the complex interactions within the 3D multicellular aggregates that mitigate cell fate decisions [228, 229]. Therefore, although many studies have examined the effects of individual signaling pathways or molecules on ESC differentiation[14, 16, 19, 58, 70, 74, 75, 83, 86], new approaches that allow systematic investigation of combinations of parallel factors that regulate stem cell differentiation are needed to more accurately predict cell responses to complex micro-environmental cues. Computational modeling strategies provide one method to systematically investigate cues governing spatial patterns of differentiation in ESC aggregates.

---

[1] Modified from:

White DE, Kinney MA, McDevitt TC, Kemp ML. *Spatial Pattern dynamics of 3D Stem Cell Loss of Pluripotency via Rules Based Computational Modeling*. PLoS Comput Biol. 9.3(2013)

Classically, computational modeling has been used in a variety of contexts to identify specific mechanisms leading to observed biological phenomena [21, 120, 131, 159, 230, 231]. While systems of ordinary differential equations have been used to model protein and gene interaction networks for single cells [20-22, 24] extending these methodologies to groups of cells in a tissue is problematic for a variety of reasons. ODEs are highly parameterized systems and while this level of description works well in the context of isolated cells, it is computationally prohibitive to screen all parameter sets to find conditions which modulate spatial pattern formation on a tissue level. Furthermore, in the context of tissues, it is important to consider spatial heterogeneity of soluble cue secretion [133, 232], which adds additional computational complexity in the form of partial differential equation solvers. Keeping track of all these parameters and the added computational intensity makes rapid screening of parameter sets modulating spatial pattern formation extremely difficult.

Alternatively agent based modeling (ABM) provides a flexible, modular framework to study complex regulatory cues. Simple rules can be defined to allow cells to transition between different states. These functions can be defined in any manner, and previous work has defined such transitions in the form of probability density functions (PDFs) [231]. PDFs defined in this manner have fewer parameters and can be easily modified to govern transition as a function of any input, which allows for a rapid screening of different rule types leading to spatial pattern differentiation. However, there is a trade-off with ABM models of this type; while they allow rapid sampling of parameter space, it is difficult to extract information about specific molecular mechanisms which give rise to the high level rules.

Another challenge with modeling multicellular aggregates is that the physical location of each cell as well as cell-cell interactions must be monitored and maintained through time. The aggregate undergoes substantial cell division, and in some cases migration, which can make modeling physical interactions a computationally intensive task [15, 33, 37, 150, 176]. ABM models provide an answer to these problems and a variety of approaches ranging from the cellular Potts method [174-177], to mass spring tissue approximations [172, 179] have been used to model biological systems [35]. While approaches such as the cellular Potts method allow for energy based approximations of physical cell parameters, these models do not scale well to large cell numbers. Various advances in mass spring systems [179, 180] coupled with the ease of generating spatial data structures over large point clouds (such as KD-Trees) make this a viable option for simulating aggregates with many cells. This study showed that a mass spring approach captured relevant physical parameters of ESC aggregates during division and differentiation; also demonstrated the utility of computational rules based modeling to predict emergent spatial patterns associated with the pluripotent transcription factor in Oct4 and investigate macroscopic principles that can play important roles in determining cell fate transitions.

3.2 Materials and Methods

3.2.1 Cell Culture

A murine embryonic stem cell line (D3) transfected with an Oct4-GFP construct was used (phOCT3-EGFP1; provided by Wei Cui, Ph.D., Imperial College, London, UK). For this particular experiment these cells were used after several passages and splits, and thus

did not show robust GFP expression under confocal or flow cytometry; thus immunostaining was necessary to visualize Oct4 expression. These cells were cultured in monolayer on 100 mm tissue culture plates coated with 0.67% gelatin in Dulbecco's modified Eagle's medium (DMEM) supplemented with 15% fetal bovine serum(FBS) (Hyclone, Logan, UT), 2mM L-glutamine (Mediatech), 100 U/ml penicillin, 100 ug/ml streptomyocin, and 0.25 ug/ml amphotericin (Mediatech), 1x MEM nonessential amino acis solcuiotn (Mediatech), 0.1 mM 2-mercaptoethanol (FisherChecmical, Fairlawn, NJ), and 10^3 U/ml leukemia inhibitory factor (LIF) (Chemicon Internation, Temecula, CA). Cells were passaged every 2-3 days prior to reaching 70% confluence.

3.2.2 Immunostaining and Confocal Microscopy

ESC aggregates were collected for staining and fixed in 10% formalin for 45 minutes. Aggregates were permeabilized for 30 minutes in 1.0% TritonX-100, re-fixed in formalin for 15 minutes, and blocked in blocking buffer (2% bovine serum albumin, 0.1% Tween-20 in PBS) for 3 hours. Samples were stained with a goat Oct4-antibody (Santa Cruz) overnight at 4 °C. After three washes in blocking buffer, aggregates were subsequently stained with a secondary donkey anti-goat Alexa Fluor 488 conjugated antibody (1:200 Santa Cruz) for 4 hours at 4 °C. Staining with Alexa Flour 546 Phalloidin (1:20 Molecular Probes) and Hoescht (1:100) was performed concurrently for 25 minutes at 4 °C. Samples were washed and resuspended in blocking buffer, and imaged using a Zeiss LSM 510 Confocal Microscope using Ar, He, Ne and Chameleon lasers. A single image was taken at the top of the aggregate and at a depth of 25 μm into the aggregate. For each time-point, 25 images were obtained.

3.2.3 ESC Aggregate Formation and Culture

Undifferentiated embryonic stem cells were dissociated from monolayer culture using 0.05% trypsin-EDTA solution (Mediatech) to obtain a single cell suspension and added to AggreWells™ (Stem Cell Technologies) [85] six well plate inserts to form six thousand aggregates of either 250 (1.5 million cells/ml) or 1000 (6 million cells/ml) cell per ESC aggregate. Aggregates were allowed to form in the wells for 20 hours, at which point they were removed and transferred into rotary culture at 60 RPM on a 100 mm plate at a density of 2000 aggregates in 10 mls of medium [233]. Aggregates were re-fed every 2 days, and 75% of the spent medium was replaced with fresh medium at each exchange. Aggregates were cultured in this manner for up to 7 days of differentiation.

3.2.4 Pattern Analysis

Spatial patterns of Oct4 expression were classified into six different categories, random, inside-out, outside-in, connected, differentiated, or undifferentiated. For an image to be classified as undifferentiated, 90% or more of the cells in the image had to positively express Oct4. Conversely, for an image to be classified as differentiated no more than 10% of the cells could stain positive for Oct4. If the number of positive Oct4 cells fell in between these two levels, the pattern of differentiation was classified as either random, inside-out, outside-in, or connected. Inside-out patterns were characterized by differentiation in the middle of the aggregate and undifferentiated cells on the outside. Conversely, outside-in patterns exhibited differentiation on the outside and undifferentiated cells in the middle. Connected patterns were defined as multiple distinct connected regions of cells of the same state. Random patterns displayed no identifiable pattern. For each time point, 25 confocal images of ESC aggregates were analyzed.

Pattern matching was performed on the output from the model as well as the experimental confocal images. Two blinded observers were used to classify the experimental confocal images. In the case of the *in silico* results, metrics such as the total number of differentiated cells, and average number of distinct cell clusters were used to aid in classifying patterns, with a total of 73 1000-cell aggregates and 66 250-cell aggregates analyzed. A cluster was defined as two or more cells of the same type and clusters were identified throughout the entire 3D aggregate.

3.2.5 Analysis of ESC Aggregate Size

ESC aggregates were harvested at various time points and fixed for 45 minutes in 10% formalin. ESC aggregates were imaged using bright field microscopy via a 4x objective on an EVOS microscope. Three representative images were taken for each sample. Images were analyzed by using threshold, watershed, and image particle detection operations in ImageJ. Aggregate radius was derived by computing the cross sectional area, approximating the EB as a circle, and calculating the radius accordingly. The circularity of the aggregates was calculated by fitting an ellipse to their area, and taking the ratio of the minor to the major axes.

3.2.6 Rules Based Modeling

Rules based modeling was carried out using probabilities to govern state changes. Linear, hyperbolic and Hill-type probabilities have been previously examined in the context of stem cell differentiation and thus similar Hill-type probability laws were used in this work [231]. All of these rules were designed to be functions only of the number of nearest neighbors to reduce complexity. For the “random” rule, a basal probability associated with the state change was set to 1% as this led to complete differentiation of the aggregate. In the case of the “positive feedback” rule, the differentiation probability

was influenced according to (3-1 3-1.

$$P(x) = \frac{\beta}{\varepsilon},\ \varepsilon = 12 \quad (3\text{-}1)$$

where β represents the number of nodes connected in the differentiated state, normalized by the total possible number of neighboring nodes ε, which for a face-centered cubic or hexagonal close-packed spherical packing arrangement is 12. In the case of the "competing feedback" rule the probability was determined according to equation 3-2.

$$P(x) = \frac{1}{1+e^{\gamma-\beta}} \quad (3\text{-}2)$$

where β represents the number of neighboring nodes in the differentiated state, and γ represents the number of neighboring nodes in the undifferentiated state. This function produces a similar sigmoidal shape as the Hill function, but does not require the inclusion of the additional Hill coefficient.

3.2.7 Determining Growth Rate

After aggregate sizes were determined, the number of cells in a spheroid was approximated by first determining the volume of the spheroid based upon known aggregate radii. Next an average cell volume was calculated using the average cell diameter of 6.6 +/- .3287 μm. By assuming a maximal close packed configuration for spheroids (.7408), the volume of the aggregate was adjusted to contain the cells. Cell numbers were calculated by dividing the adjusted aggregate volume by the volume of a

single spheroid. To calculate the cellular growth rate equation 3-3 was applied between discrete time points.

$$c(t) = c_0 e^{kt} \quad (3\text{-}3)$$

This method produced growth rates over which closely matched the proliferation of Oct4+ pluripotent mESCs in 2D. By applying this same process for later day ESC aggregates, a division time for an Oct4- cell was also calculated. The growth rates could then be fit to equation 3-4 determine doubling times that were used in the model for Oct4+ and Oct4- negative cells.

$$t_2 = \frac{\ln(2)}{\ln(1+r)} \quad (3\text{-}4)$$

3.2.8 Dynamic Modeling

Modeling of dynamic cell movement was accomplished using custom C# code with the aid of XNA package for vector math and 3D visualization. Cells were modeled as rigid spheres connected by springs to depict cell-cell physical connections. A complete collision detection algorithm was used to resolve all possible collisions at each time step of the simulation. Simulations were run for a period of 144 hours (6 days), until 99 percent of the cells had changed state or until 40,000 total cells existed in the model as computational time increased dramatically with increased cell number above this limit. Cells were allowed to change fate instantaneously. The kinetics of the simulations were fit to model growth curves, thus the probabilities were given different weights to assure pattern formation was observed. In the case of the random and positive feedback rules, no weights were applied to the rules as the random parameter induced complete differentiation over the 144 hour simulation period. However, in the case of the

competing feedback rule a 0.01 weight was applied to adjust differentiation kinetics to match those observed experimentally. 10 simulations were run for each different rule set and simulations were run in parallel using an Intel Core i7 X980 3.33 GHz CPU with 12.0 GB of RAM.

3.2.9 Software Tools

Rules based modeling was achieved using a Python language with the following freely available software packages: pyode, numpy, matplotlib, python imaging library (PIL) and vpython. Physical aggregation simulations of structure were performed using PyODE as the underlying physics engine. Results were plotted via the aid of numpy and matplotlib. The 2D aggregate slices were visualized using PIL. The 3D aggregate was visualized using vpython.

3.2.10 Statistical Analysis

All experiments were performed in triplicate and data was presented as the mean +/- the standard deviation. Significance was determined using a student's t-test with a significance level of $p < 0.05$.

3.3 Results

3.3.1 Modeling Embryoid Body Structure

The overarching goal in constructing a model description of ESC aggregates was to accurately recapitulate the overall multicellular structure based on the physical properties of individual mouse embryonic stem cells (mESCs). Prior models of multicellular structures have described the individual cellular agents as incompressible objects consisting of ellipsoids [234]. In this study a physics-based modeling approach was

implemented in which cells were modeled as incompressible rigid spheres as this was a powerful and portable method for representing complex aggregate shapes [235, 236]. To determine if modeling mESCs as spheres was appropriate, the effective surface area (Figure 3-1 A) and radii (Figure 3-1 B) of individual ESCs were experimentally determined via Coulter counter analysis. The average surface area to volume ratio of the mESC line was 3.26 +/- 0.15, which is only ~8% higher than the theoretical value of 3.00 for a spheroid. Due to the increased computational costs associated with ellipsoid collision detection algorithms and the relatively low error in the surface area-to-volume ratio (< 10%), each cellular agent was represented as a sphere. The distribution of cell radii from the Coulter counter measurements (Figure 3-1 A, B) were used to create the population of spheres for each agent in the aggregate simulations. These cell agents were randomly seeded into a box, which served as an initial boundary for the simulation, and then forced to aggregate using a gravitational point source into the subsequent multicellular spheroidal structure. This is analogous to centrifugation into a pyramidal well were centrifugation forces cells to initiate the aggregation process.

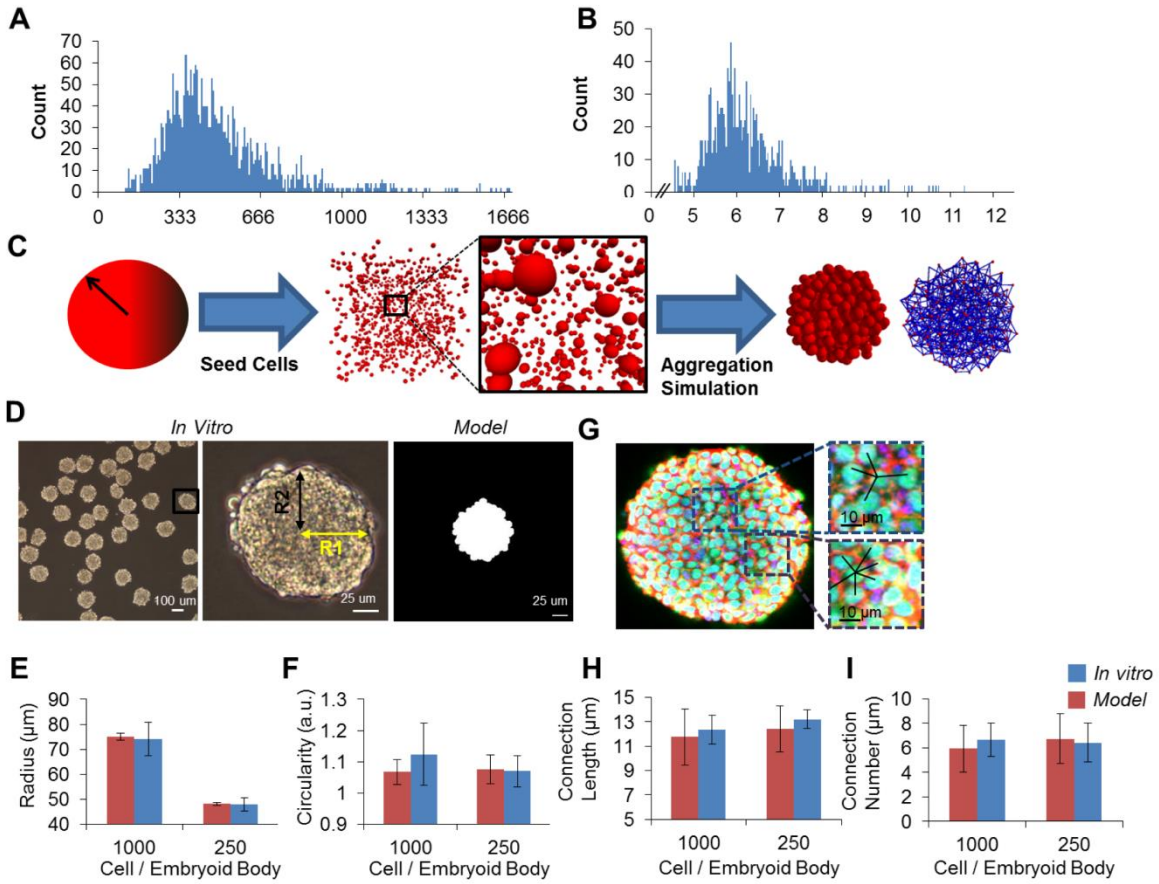

Figure 3-1 : Aggregate Modeling Methodology. Dissociated mouse embryonic stem cells (mESCs) were analyzed via Coulter counter for surface area (A) and radius (B). *In silico* aggregates were generated in a physical modeling process in which cells were generated and then forcibly aggregated (C). ESC multicellular aggregates were formed via ultra-high throughput methods for two initial cell numbers: 250 and 1000 - a representative image for 1000 cell/aggregate seeding density(D). The black box in first column is digitally enlarged in the second column. Circularity was calculated by fitting the aggregate to an ellipse and taking the ratio of the two radii labeled R1 and R2 respectively. Aggregates were analyzed for two macro scale aggregate properties: radius (E) and circularity (F). Confocal images were used to analyze local aggregate properties – a representative 1000 cell/aggregate seeding density (G). Aggregate local properties were assessed by two metrics: connection length (H) and number of connections (I).

The structures of *in silico* and *in vitro* aggregates were assessed for aggregates of 250 and 1000 cells seeding densities using four parameters: radius, circularity, connection count, and connection lengths. Size and circularity were used to assess the entire aggregate structure and were experimentally determined through the analysis of phase contrast

images, while similar measurements were obtained using projections of the *in silico* aggregates onto a 2-dimensional plan (Figure3-1 D). The results indicated that the model appropriately captured the macroscopic features of the relative aggregates since there were no statistical differences between the model and experimental metrics (Figure 3-1 F). The connection count and connection length parameters were calculated from the spatial distribution of individual cells comprising the aggregates and served as quantifiable metrics for assessing local micro-scale organization within aggregates. These parameters were assessed by individual cell labeling performed in confocal microscopy images and via computational algorithms for the *in silico* aggregates (Figure 3-1 G). As an example, the blue box in Figure 3-1 G highlights a cell with an average connection length of 14.87 +/- 2.07 microns and connection number of 4. I*n silico* aggregates were "virtually sectioned" (at a 10 μm thickness) to perform similar analysis on a 2D projection, and neither the average circularity nor the connection lengths differed statistically from the experimentally derived aggregate values (Figure 3-1 H,I). Overall, quantitative comparison of four different physical parameters indicated that the physical model accurately captured the structure of individual aggregates on both the aggregate and cellular scales, providing an accurate structural framework for subsequent analysis of spatial patterning.

3.3.2 Spatial Patterns Occur During Differentiation

Throughout the subsequent discussion of the results, pluripotent cells that exhibit loss of Oct4 expression are simply referred to as "differentiated", acknowledging the caveat that Oct4- cells are not terminally differentiated. As Oct4 is concomitant with loss of pluripotency, it was used to monitor the pluripotent state of the cells [125, 237-239]. Loss

of Oct4 has commonly been modeled as a bi-stable transition, which causes an all or none response [114, 125, 130]. The temporal patterns of loss of pluripotency were evaluated in 250- and 1000-cell ESC aggregates via confocal microscopy to examine Oct4 expression. Starting from a homogeneous population of undifferentiated cells, spatial heterogeneity (as defined by loss of Oct4 expression) was observed over the differentiation time course. In order to capture the diversity of spatial pattern heterogeneity, a classification system was developed. Based on preliminary results from both the experimentally derived aggregates and the model, six different categories of patterns were proposed: Oct4+, inside-out, outside-in, connected, random, and Oct4- (Figure 3-2). These patterns were loosely grouped into three larger categories: Oct4+, transitioning, and Oct4-. The outside-in, inside-out, connected, and random patterns were all considered transition patterns as they captured intermediate stages of the differentiation process. Inside-out patterns were characterized by differentiation in the middle of the aggregates and undifferentiated cells on the outside; conversely outside-in patterns exhibit differentiation on the outside and undifferentiated cells in the middle. Connected patterns were defined as multiple distinct connected regions of cells of the same state, whereas random patterns were classified as no identifiable pattern based on a lack of connectivity.

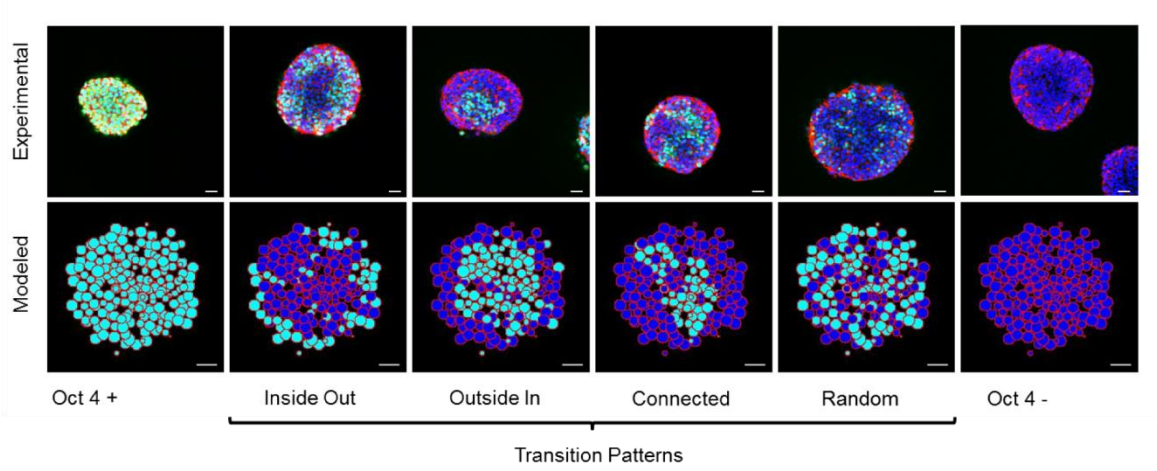

Figure 3-2: Classification of spatial differentiation patterns within ESC aggregates. Six classifications are used to described Oct4 expression patterns: undifferentiated, inside-out, outside-in, connected, random, and differentiated. Confocal images shown on top are stained with DAPI (blue), phalloidin (red) and Oct4 (green) (with a scale bar of 25 μm). The model generates patterns similar to confocal images with Oct4- cells (dark blue) and Oct4+ cells (cyan). Scale bar represents 25 μm.

In the smaller 250-cell ESC aggregates, Oct4 expression persisted for up to six days (Figure 3-3 A). Rapid loss of Oct4 was observed between days 3 and 5 (Figure 3-3 C) and the patterns associated with differentiation were classified entirely as "connected" (Figure 3-3 D). In 1000-cell aggregates, differentiation patterns were assessed over a 7-day period (Figure 3-3 B). Differentiation was observed to occur at a later time than the smaller 250-cell aggregates, with transition patterns occurring from days 4 to 7 (Figure 3-3 E). The spatial patterns in the 1000-cell aggregates associated with differentiation were more varied than the 250-cell aggregates but also were primarily classified as "connected" (Figure 3-3 F). At each time point, pattern classification for each aggregate size was performed to generate temporal differentiation profiles for each time point (Figure 3-3 D,F). The trajectories of differentiation were calculated by assessing how the number of differentiated, undifferentiated, and transitioning patterns changed over time.

Although the types of patterns associated with differentiation only changed slightly with aggregate size (Figure 3-3 D,F), the kinetics of the process appeared to change appreciably. The 250-cell aggregates began differentiating at ~ day 3 and finished within one day, whereas the 1000-cell aggregates started a similar process later at ~ day 4, and took 3 days to nearly exhaust Oct4 expression (Figure 3-3 C,E).

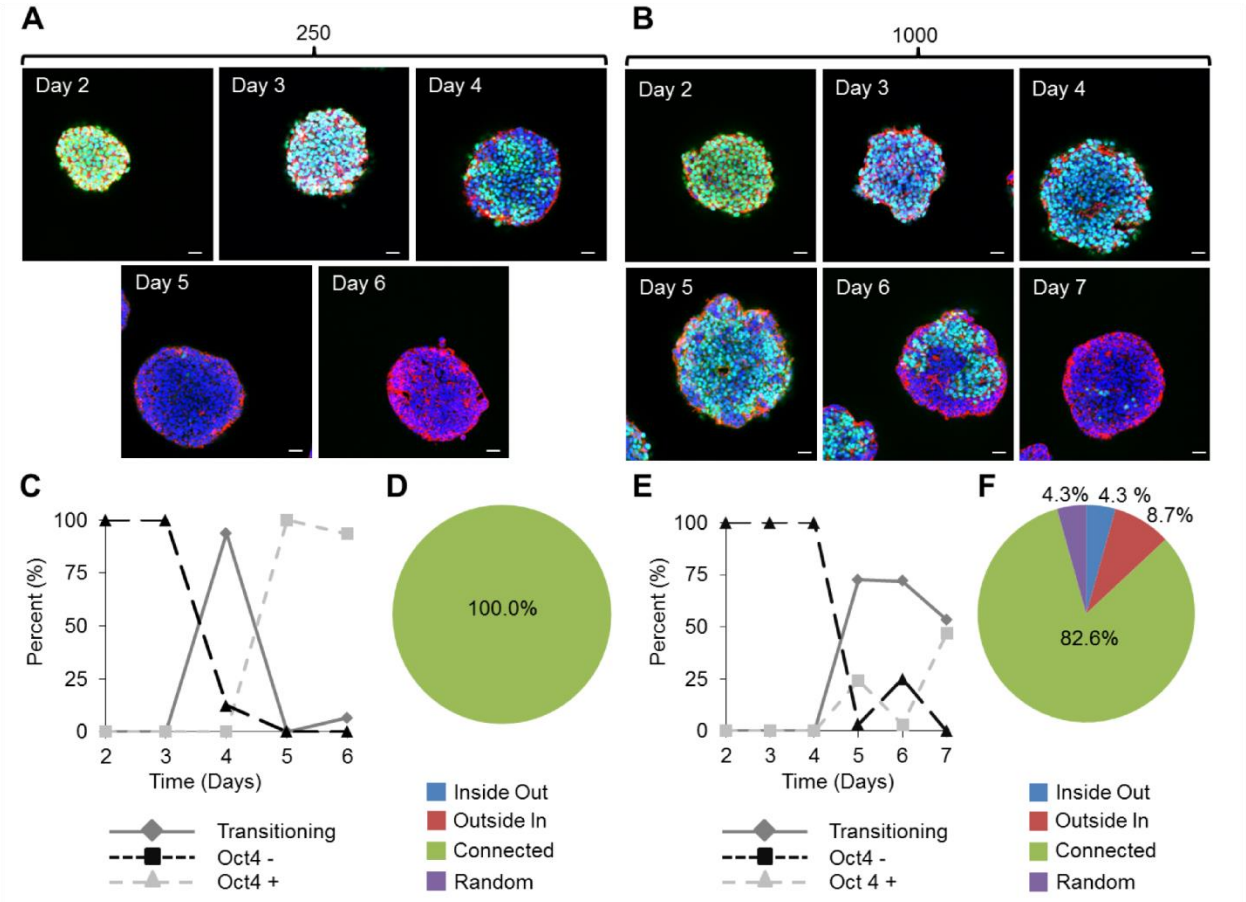

Figure 3-3: Differentiation kinetics differ by aggregate size. Confocal images of aggregates stained with DAPI (blue), phalloidin (red), and Oct4 (green) shown at a depth of 25 μm for ESC aggregates of (A) 250 and (B) 1000 cells. (C, E) Temporal dynamics of observed patterns for 250 cell aggregates (C) and 1000 cell aggregates (E). (D, F) The overall distribution of observed patterns for 250 cell aggregates over 6 days in culture (D) and 1000 cell ESC aggregates over 7 days in culture (F). Scale bars on all images are 25 μm.

3.3.3 Rules Based Modeling of the Cellular Microenvironment

After validating the generation of an appropriate 3D geometry for ESC aggregates, rules based modeling was performed by creating network structures, in which "nodes" represented individual cells, and "connections" represented physical interactions between adjacent cells; nodes were allowed to convey information with the macrostructure along the defined connections. The goal was to determine if simple rules accurately produced the distribution of spatial patterns observed experimentally. During these initial simulations the macro-structures were assumed to be static (i.e. no proliferation, migration or apoptosis). Cells could exist in either of two states: undifferentiated (Oct4 +) or differentiated (Oct4 -). The transition between these two states was chosen as binary based on previous modeling work [125] and occurred based on different rule formulations: "random", "positive feedback" or "competing feedback" (Figure 3-4). The random rule configuration represented a stochastic, basal differentiation probability (Figure 3-4 A). The positive feedback rule was based on a paradigm in which differentiated cells biased neighboring cells to differentiate (Figure 3-4 B) and was inspired by differentiation induced via direct cell-cell interactions [240, 241]. Finally, the competing feedback rule depicts a situation where differentiated cells promoted subsequent differentiation of neighboring cells while undifferentiated cells inhibited this transition (Figure 3-4 C). Positive feedback in this rule was based upon the known role of soluble factors to maintain pluripotency [242], while negative feedback comes from the differentiation induced via the cell-cell interactions discussed above [240, 241].

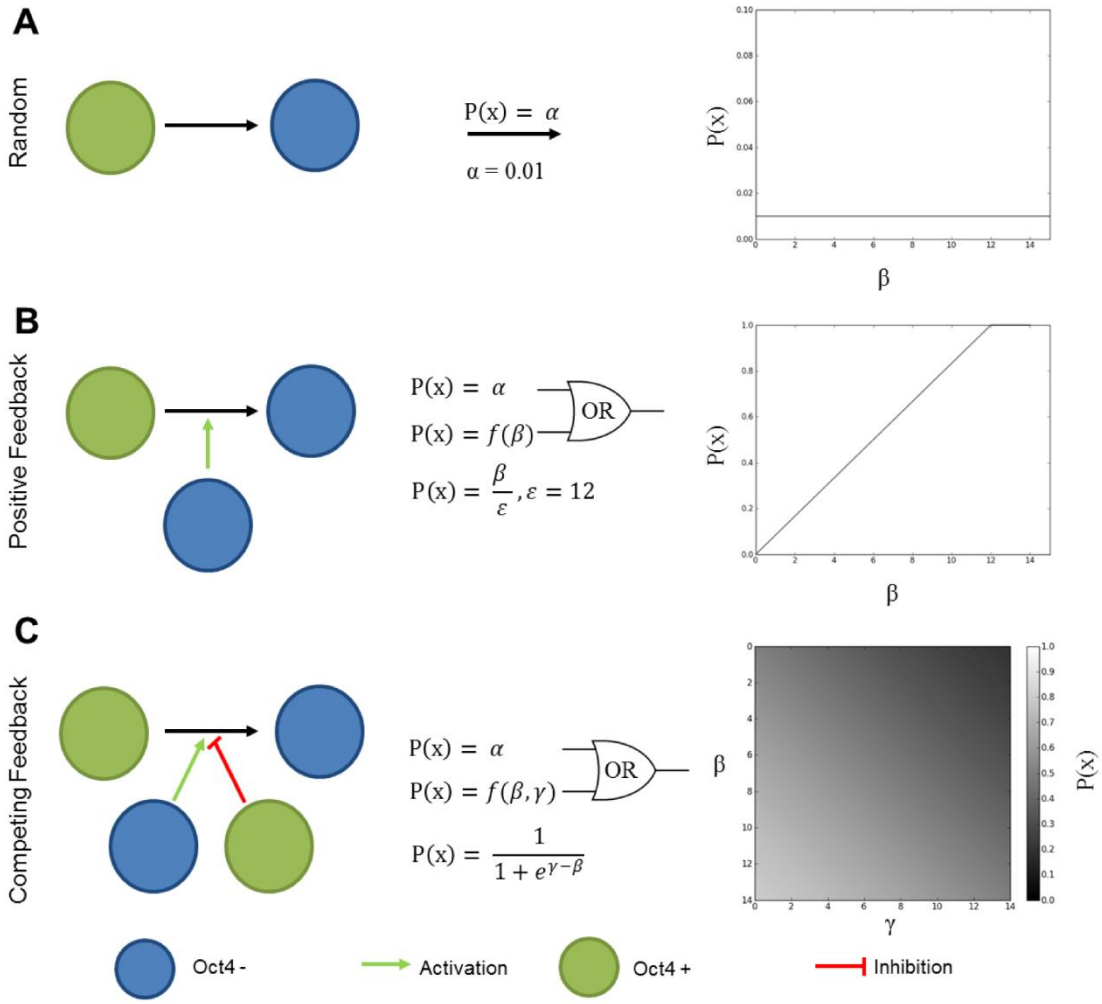

Figure 3-4: Overview of Rule Configurations. Three different rule configurations governing transitions from the undifferentiated to differentiated state are shown: random (A), positive feedback (B), and competing feedback (C). The random rule is governed by a constant probability of differentiation denoted by α (A). The positive feedback rule takes into account Oct4- cells denoted by β, and allows them to positively influence the differentiation probability (B). The competing feedback rule take into account β and also the number of Oct4+ cells denoted by γ (C). In the case of the positive feedback and competing feedback rules, the probability are combined using an or-gate logical operator. Representation of the probaility density functions P(x) are shown for each rule.

3.3.4 Quantitative *in Silico* Pattern Analysis

To glean insight into the evolution of simulated spatial patterns, two quantitative metrics were used, undifferentiated cluster number (UCN) and differentiated cluster number

(DCN), to assess pattern formation and simulation trajectories against a normalized time (τ) axis. Analyzing the cell phenotype transitions by the UCN and DCN metrics revealed distinct paths of pattern formation for each of the different rules. From such curves, critical points ($\tau = 0.2$, $\tau = 0.4$, and $\tau = 0.6$) representing rapid changes or important regions across all rules were chosen and representative aggregate slices were displayed (Figure 3-5 B,E,H). Analysis of the trajectories themselves revealed insight about the types of clusters being formed in the "connected" patterns. In the "positive feedback" scenario, the loss of Oct4 expression was characterized by a high number of differentiated or undifferentiated clusters, suggesting localized intercellular neighbor influences regulating phenotype transition (Figure 3-5 D). The peak in UCN at $\tau = 0.6$ was characterized by a large number of isolated pockets of undifferentiated cells. In contrast, the "competing feedback" rule peaked through a high number of differentiated clusters, but never amassed a high undifferentiated cluster number (Figure 3-5 G) which matched the larger isolated and persistent clusters of Oct4 positive cells experimentally observed in both the 250- and 1000-cell aggregates. Taken together, these data indicate that the "competing feedback" rule matched the patterns observed biologically with the highest fidelity for the different size aggregates examined.

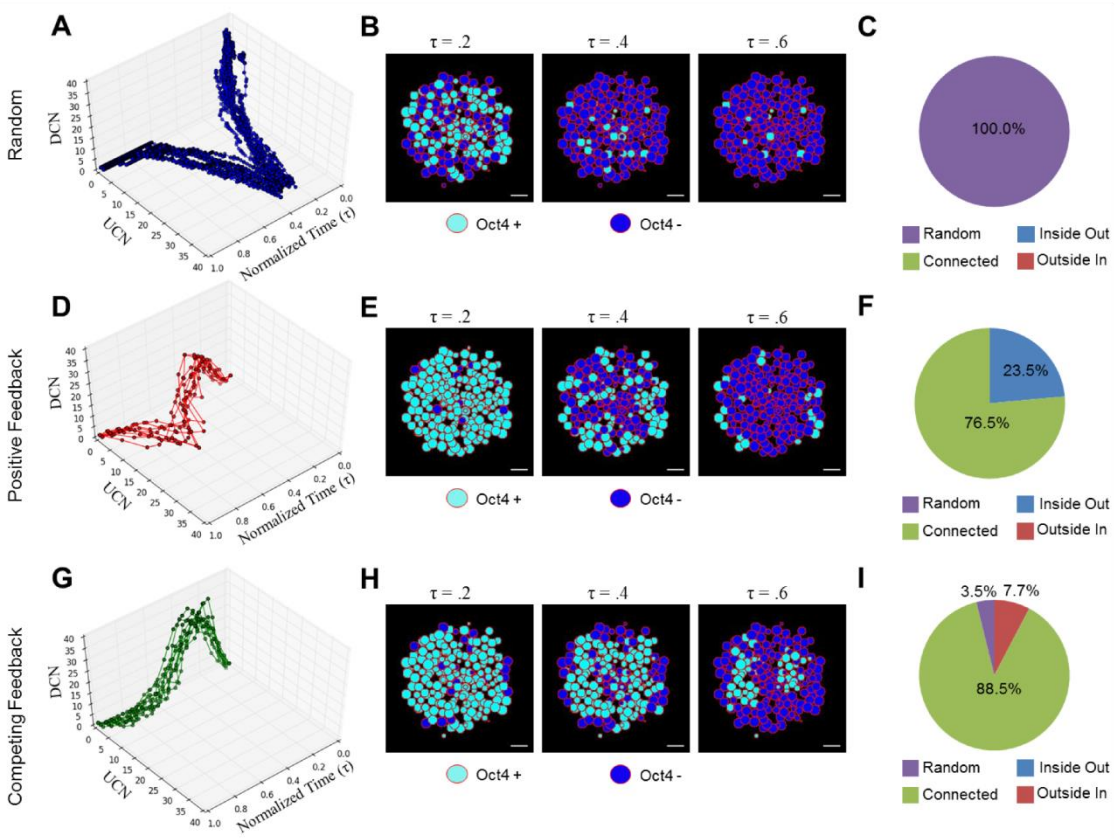

Figure 3-5: Spatial pattern trajectories of differentiation for 1000 cell aggregates. (A, D, G) Pattern trajectories are shown for all three rules for 1000-cell aggregates plotted against a normalized time axis (τ) where the time step was divided by the total number of time steps required for the simulation to complete. (B, E, H) Representative "virtual sections" of aggregates over the course of a simulation. Cyan represents Oct4+ cells while blue signifies Oct4- cells. (C, F, I) Differences in the kinetics of modeled differentiation for all three rules. All scale bars are 25 μm.

Furthermore, the trajectory analysis provided novel information about the evolution of certain patterns. For example, random differentiation was characterized by a high initial spike in the DCN as this signified the emergence of several small clusters of differentiated cells (Figure 3-5 A), and the duration of this spike represents how long the random patterns persisted throughout the duration of the model. If the UCN remained fairly low, the pattern transitioned into a connected phenotype, again evidenced by the low number of undifferentiated clusters of cells (Figure 3-5 B). When the UCN remains

at 1, this signified either an inside-out, or outside-in pattern. However, if the UCN transitioned towards a high value, this signified that the differentiation was still largely governed by random patterns (Figure 3-5 A). Overall, these measures provide quantitative metrics for assessing the types of patterns formed, and the evolution of these patterns over time. One limitation of the model, however, was that the kinetics of pattern formation could not explain the differences in kinetics experimentally observed between different aggregate sizes (Figure 3-3 D,F). This suggested that although a static size aggregate modeling approach was sufficient for predicting the prevalence of different spatial pattern classifications, it did not fully capture the kinetics of experimental Oct4 loss. In order to further investigate the kinetics of the pattern transitions, the modeling framework was modified to include cell division and aggregate growth.

3.3.5 Dynamic Rules Based Modeling

Dynamic processes, such as cell division and growth of the ESC aggregates, were implemented to test the hypothesis that cell division and growth influence spatial patterns of Oct4 expression loss. To investigate the effect of cell division on this loss of pluripotency transition, a revised model which simulated growing multicellular structures was created. With this approach, cells were modeled as rigid spheres while cell-cell connections were represented by springs that maintained the overall macro-structure of the aggregate (Figure 3-7 A). The size of differentiating ESCs did not change appreciably with time, which reduced complexity from the model description (Figure 3-7 B). The first step was to determine an estimated cellular division rate for Oct4+ and Oct- cells. This was accomplished using experimental growth data approximated from the size of the aggregates (Figure 3-7 C), which yielded a division rate of ~ 21 hours for Oct4+ cells and ~ 51 hours for Oct4- cells which is consistent with the literature [243]. Using these division rates, the rules derived in the former static model were applied to the new

dynamic model. As an internal control, division simulations with no rules were run to monitor any bias introduced by the model (Figure 3-6) and found that the cells grew in an exponential manner (Figure 3-6 A,F), while the density of the aggregates remained constant (Figure 3-6 C). The circularity of the network also remained constant with time (Figure 3-6 D). Connection length remained constant as a function of the aggregate radius, whereas the connection number decreased on the outer layer of the aggregate, as was expected (Figure 3-6 E). Taken together these results suggest that structurally no bias was introduced into the model by introducing cell division.

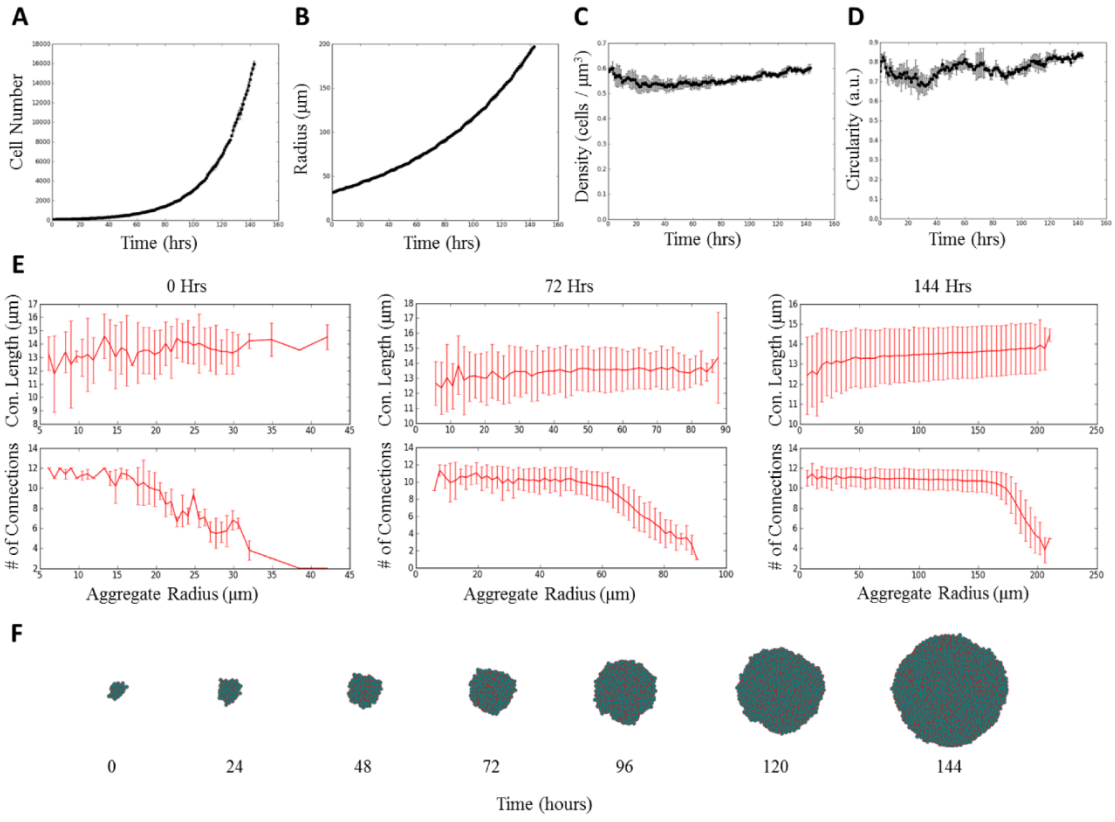

Figure 3-6: Dynamic modeling does not alter network structure. Control metrics for the dynamic cell simulations were plotted as functions of time: cell number, (A), radius (B), density (C), and circularity (D). Internal network parameters for the average connection length and average number of connections as a function of aggregate radius (E) behaved as expected. A representative trace for the growth of an aggregate starting at 50 cells shows a visual representation of the growing aggregate structrues (F).

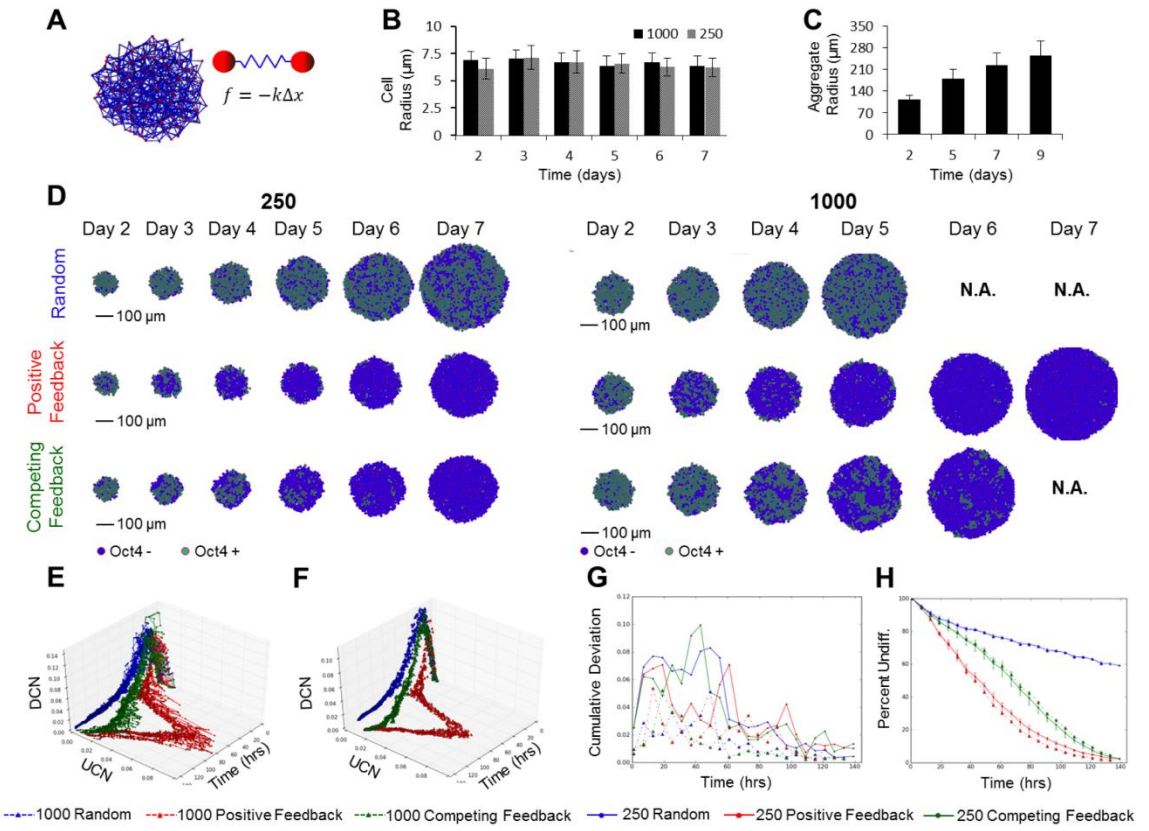

Figure 3-7: The effects of cell division on spatial pattern trajectories. (A) Mass spring schema for modeling cellular division and adhesion. (B) Coulter coulter data showing the radius of stem cells from embryoid bodies and how they change over the 6 day differentiation time course. (C) Estimated sizes of embryoid bodies used to estimate growth rates. (D) Representative virtual slices of simulated 250 and 1000 cell dividing embryoid bodies for the three different rules. Oct4+ cells are shown in teal and Oct4- cells in blue. UCN, DCN time plot trajecory for the 250 cell ESC aggregates (E) and 1000 cell ESC aggregates (F). (G) Summation of the variance in the UCN, DCN trajectories. (H) The percent of undifferetiated cells as a function of time.

Next the spatial patterns formed under the various rule configurations during enlargement of the aggregates over a 5-7 day period were examined (the pattern trajectories now normalized by the cell number to account for cell growth). Trajectories simulated over the 6 day culture period indicated consistent pattern distributions and their evolution over time (Figure 3-7 E,F). While few differences were observed in the trajectories as a function of aggregate size, the cumulative variance in the UCN and DCN metrics was

greater for the 250 cell aggregates (Figure 3-7 G), which suggested a more heterogeneous population of differentiated cells. At later time points (> Day 5) the positive feedback rule generated clusters of undifferentiated cells predominantly on the outer edges of the ESC aggregate. In contrast, the competing feedback rule produced larger clusters of undifferentiated cells localized towards the center of the aggregate. The trajectories of these growth simulations matched the general shape of the trajectories for the static simulations (Figure 3-5), with a notable exception for the random rule. Analysis of the percentage of undifferentiated trajectories revealed that over half the cells remained undifferentiated at the seven day time point, which may explain the absence of the DCN to UCN transition observed in the previous static model simulations.

The motivation behind modeling a dynamic aggregate structure was to more closely recapitulate the emergent morphogenic processes occurring over the transitional 5-7 day period and to investigate whether the inclusion of cell division and aggregate enlargement influenced emergent spatial patterns (Figure 3-7 H). In the case of the random and competing feedback rules, no observable difference in the percentage of differentiated cells appeared, however the positive feedback rule resulted in a lower percentage of undifferentiated in the 1000 cell/aggregate density versus 250 cell aggregates, which was likely due to the total number of cells present. Taken together these results suggested that differential cell division did not significantly influence the formation or evolution of patterns over time and the pattern formation process was dominated by the regulatory mechanisms encoded in the probabilistic rules.

3.4 Discussion

Differentiation is a complex biological process involving the coordinated regulation of multiple genes by intrinsic and extrinsic factors. Rather than attempt to model the intricate network of genetic circuitry, signaling mechanisms, and environmental cues, loss of pluripotency was approached from a simplified perspective designed to elucidate the most basic principles dictating pattern formation in a spherical multicellular system. A dynamic modeling framework capable of recapitulating the physical properties of ESC aggregates for multiple aggregate sizes was developed. This framework allowed simulation of how different probabilistic rules were manifested in the emergence of spatial patterns and examination of the evolution of these patterns over time. Through comparison of the simulated pattern trajectories with the pattern classification developed for the experimental data, it was determined that all possible pattern classes could be explained by the rules employed.

Differentiation is classically thought of as a binary transition from the undifferentiated pluripotent stem cell state to differentiated phenotypes [5,31,46]. Models of stem cell differentiation often consider these events on a population basis [244] or at an intracellular signaling level [245]. This study shows that the transition from Oct4+ to Oct4- states produces dynamic, spatially heterogeneous patterns in a continuous manner. These results indicated that modeling embryonic stem cell (ESC) fate decisions as a stochastic, binary process is sufficient to predict the dominant emergent spatial patterns of loss of pluripotency. Additionally, 250- and 1000-cell ESC aggregates underwent loss of pluripotency through the same intermediate patterns, which suggested that the macro processes governing this early stem cell transition (while occurring at different rates) were largely independent of size.

Our modeling approach created *in silico* aggregates with similar properties to *in vitro* ESC aggregates. The spring based constraint representing cellular adhesions was able to accurately capture the evolution of the aggregate architecture. Additionally, the high-level rules described here were able to reproduce the emergence of a variety of spatial patterns, all of were observed experimentally in ESC aggregates. Both modeled and experimental aggregates demonstrated enrichment in connected patterns of cells. Quantification of model simulation outputs by differentiated and undifferentiated cluster number (DCN and UCN) allowed visual representation of the time evolution of connected patterns in 3D multicellular aggregate systems. The use of the UCN and DCN metrics provided information about not only the types of connected patterns formed, but also the transitions between the different types of spatial patterns. A combination of analyses using these metrics and manual pattern identification indicated that the "competing feedback" rule scheme (that accounted for opposing influences of neighboring pluripotent and differentiated cells) produced a distribution of spatial patterns that closely resembled the experimentally observed spatial patterns for both 250- and 1000-cell aggregates. This was observed in both the static and dynamic division models, and could be hypothetically represented biologically as a combination of cell-cell signaling and the complex interplay of local soluble factors and other chemical gradients that influence pluripotent cell fate decisions [241, 246-248]. However, this is only one possible explanation for a coarse-grained description, and a variety of other signaling pathways and molecules are likely involved in regulating this transition. Interestingly, none of the rules generated spatial patterns if cells were not allowed to also spontaneously differentiate at a low rate random rate ($\alpha = 0.0075$). Though, for the

positive feedback rule, this follows directly from the construction of the probability equations, the formation of strongly connected patterns was also not observed in the competing feedback rule without the inclusion of this low stochastic rate (data not shown) suggesting random component is important to capture the experimentally observed transition patterns.

The static structure model predicted small differences in the Oct4 expression kinetics of aggregates with different cell seeding numbers should exist; however, the slight changes observed did not capture the full extent of the variation present in the experimental results. The inclusion of cell division (with a faster division rate for pluripotent cells) was also not able to explain the difference in the kinetics of this process. These results indicated that additional factors in the changing culture environment may modulate the kinetics of Oct4 loss in a size-dependent manner; hydrodynamic effects [83], diffusion limitations and/or local chemical gradients may need to be taken into account for changing aggregates sizes [249, 250] in order to reproduce the experimentally observed differences in differentiation kinetics. It is also possible that the rules chosen did not describe the system with sufficient resolution. Furthermore, the model was constructed such that all cells have the same strength when affecting other cells, which assumes that cells convey the same amount of information regardless of the amount of cell connections or the amount of shared cell area, an assumption that may need to be refined as more detailed information about the nature and directionality of intracellular communication is included. Future developments will account for cell migration, local versus distal cell-cell communication and diffusion within the aggregate to investigate how these traits affect the physical microenvironment.

The top-down modeling approach described in this study provided new insight into the spatial pattern development associated with differentiation of ESCs in 3D aggregate structures. Surprisingly, without explicitly modeling diffusion gradients or specific signal transduction mechanisms, features of temporal and spatial regulation were elucidated. The proposed modeling technique demonstrated validity for examining spatial pattern formation during differentiation. The UCN and DCN metrics provided a preliminary way to examine spatial pattern formation; however, short of counting clusters by hand, no direct comparisons between computational models and experimental data could be achieved, highlighting the need for a common pattern classification system.

CHAPTER 4 QUANTITATIVE PATTERN ANALYSIS

4.1 Introduction

In order to accurately compare computational modeling and experimental data sets, a common description for both systems was needed. Typically comparisons between spatial patterns are by deriving sets of metrics which can be used to distinguish between different classes of patterns. These metrics are then used as inputs for machine learning algorithms which seek to define sets of features which can classify or distinguish different pattern types. Pattern classification schemes can be broken down into two broad classes: supervised vs. unsupervised. In the case of unsupervised learning, the computer attempts to extract meaningful metrics (such as sets of pixels from an image) and then assigns them to different classes. In the case of supervised learning the user provides a set of ground truths which define values for a given pixel of location in space. Thus supervised pattern classification can be reformulated into an optimization problem where a ground truth set of data is compared to the classification algorithm outputs until the differences are minimized. Each of these algorithms have different costs, while unsupervised algorithms are unbiased they require a lot of data to ultimately extract meaningful information. However, supervised pattern schemas can be trained with significantly less data, but are inherently more biased as the user specifies what classes of pattern exist in the system. In many cases, such as biological images, it can be difficult to obtain a large enough data set to train unsupervised algorithms. To this end, *in silico* images are often created (rearrangements and transformation of existing biological images) to provide enough data for classifier optimization. Thus, though unsupervised approaches often produce models with more sensitivity due to their unbiased nature, there

exists a trade-off between the amount of data available and the type of pattern classification schema which can be used.

In the context of biological spatial pattern classification numerous features are often extracted from images of large scale tissues, or of images single cells. Classification algorithms for pattern recognition in Drosophila and *C. elegans* have gained widespread popularity, and are robust for each system of interest [251-257]. Classification approaches analyzing how single cells change size and shape for studying changes in cellular morphology have been applied as well [194, 196]. Often similar metrics can be extracted from images at either scale such as shape, entropy, complexity, and texture metrics; however interpreting the physical meaning of metrics like complexity and entropy can be difficult [194, 196]. Furthermore, though some attempts have been made to bridge these scales [258], as of yet, no one has utilized single cell information to inform spatial pattern classification. This would be ideal for classifying complex aggregates of cells, as information about how many cells display a given phenotype would be available, as well as information about their localization and spatial distributions. One way to address these challenges is to view cellular aggregates as large networks of cells, and utilize network theory to extract quantitative metrics from such networks.

In Chapter 3, only two metrics (the undifferentiated (Oct4+) cluster number (UCN) and differentiated (Oct4-) cluster number(DCN)) were used to describe spatial patterns. Though the UCN and DCN metrics provided some discrimination between different pattern types, follow up studies indicated that they did not provide sufficient resolution to distinguish between all experimentally observed pattern types. Therefore to further discriminate between different spatial patterns, a more complete set of metrics

was needed. These studies addressed this issue by deriving a set of novel network based metrics for classification of various spatial and temporal patterns. Network based analysis was used because it inherently maintains single cell spatial information while also linking neighboring cells together. This allows traversal of the networks to extract phenotypic subnetworks or clusters which were further analyzed; this recursive process generated a wealth of metrics which were used to classify various different spatial pattern types. The existing computational modeling platform was used to generate relevant ESC aggregate structures which represented 7 different spatial pattern types to create a data set with enough data points to accurately fit classifiers for each pattern types. The results of this study indicated that supervised classifiers trained with the network derived metrics can accurately classify spatial patterns with an accuracy of > 95%.

4.2 Materials and Methods

4.2.1 *In Silico* Pattern Generation

In silico patterns were generated using 7 different pattern generators: Undifferentiated, Differentiated, Inside-Out, Outside-In, Random, Globular, and Snaked. The differentiated patterns were created by assigning a differentiated state to all except 5% of the cells. The undifferentiated states were created by assigning an undifferentiated state to all but 5% of the cells. The 5 % margin was incorporated to allow for some of the error creeping into the biological data sets from the network digitization algorithm. Outside-In patterns were generated using a radius parameter; all cells within the radius were set to undifferentiated while all cells outside were set to differentiated. Inside-Out patterns were generated using a radius parameter; all cells inside the radius were set to differentiated, while all cells outside were set to undifferentiated. Random patterns were generated by randomly setting a fraction of the cells to a differentiated state. Globular patterns were created using two

parameters: a seed number, and expansion number. The seed number governed how many differentiated cell clusters would be found initially, and the expansion parameter governed how many layers of nearest neighbors would be turned into a differentiated state. For example, an expansion parameter of 2 would turn all cells within 2 network connections away to a differentiated state. Snaked parameters were generated using two parameters: the number of seeds, and the elongation of each seed. The number of seeds again denoted the number of initial starting locations that were differentiated. Each of these conditions was then extended to a length of *n*, by picking a random neighbor and changing its state to differentiated and repeating the process until the target length was reached.

4.2.2 Principal Component Analysis

Principal component analysis was performed using *sklearn* for the python programming language and used primarily as a dimensional reduction technique for data visualization. The python package *Matplotlib* was used to plot all PCA plots, while the heatmaps displaying component information where created with custom written code using a combination of the python packages *numpy* and the *PIL*. All data points were mean centered and unit variance scaled as required by the PCA algorithm using the scale function from *sklearn*. The PCA algorithm relied on singular value decomposition, which can lead to multiple fitted estimators displaying the data with principal components flipped. When automatic dimension fitting was required, the method of Thomas P. Minkas was used [259].

4.2.3 Classification

The following classification methods from the python package *sklearn* were used: SGD, NuSVC, SVC, linear SVC, Decision Tree, K Nearest Neighbors. In all cases, classifiers were trained using a training data set, followed by subsequent classification of a test set for metric evaluation. The data set was split into a test and training set using a simple K-Fold splitting strategy (where k = 2). Grid searches for the sets were also performed as outlined in appendix A.4.2. This code set up the grid searches which were then subsequently run and returned the best trained classifier. These classifiers were then evaluated using the following criterion: recall, precision, f1_score, area under the curve, average precision, accuracy. Accuracy was computed as the fraction of completely accurate predictions. The average precision scores and the area under the curve were derived from the precision-recall curve, where the average precision was the average value over the curve, and the area was the integral of the entire curve. The precision score measures the ability of the classifier to not label a negative sample as positive. The recall score was the ability of the classifier to find all positive samples. The f-measure was a weighted harmonic mean between the precision and recall scores and was computed as F_1 = 2* (precision*recall) / (precision + recall).

4.2.4 Clustering

The clustering algorithms K-means, Ward Hierarchical, BDSCAN, Mean-Shift and Affinity propagation were all used from the *sklearn* distribution. The affinity propagation method was used with all default parameters (damping coefficient of .5, max iterations at 200)[260]. For the MeanShift algorithm all default parameters (automatic bandwidth estimation, automatic cluster determination) were used [261]. The Ward algorithm was

used for hierarchical clustering, and the *n_clusters* parameter was varied as needed. The K-means algorithm was used with default parameters except the *n_clusters* parameter which was varied as needed. The DBSCAN algorithm was run with all default parameters (epsilon 0.5 representing the allowable distance between samples, and min samples of 5 which represents the number of points adjacent for this to be considered a core point)[262]. To evaluate cluster efficacy 6 metrics were used: homogeneity, completeness, v-measure, silhouette coefficient, adjusted rand score, and adjusted mutual information score. Homogeneity measured the extent to which each cluster only contains members of a single data class. Completeness measured to what extent all measures of a given class are assigned to the same cluster. The v-measure was a harmonic mean between the completeness and homogeneity and was computed as follows 2*(homogeneity*completeness) / (homogeneity + completeness). The adjusted rand score took into account random labeling artifacts for data sets with small amounts of points. Similarly the adjusted mutual information score is a mutual information score normalized against chance to give a measure of the agreement between the true labels and the labels provided by the clustering.

4.2.5 Regression

Regression models were evaluated across a myriad of different approaches: linear, ridge, lasso, elastic net, least angle (LARs), LARs lasso, orthogonal matching pursuit (OMP), Bayesian, Bayesian ridge, Logistic, Stochastic Gradient Descent (SGD). The regression algorithms were all freely available through the python *sklearn* distribution. In all cases, a grid search was run to find the best regression model for the given data set during the training phase. The data was split into a training and test set using a simple K-Fold

splitting method (where k=2). The models were evaluated using two criteria: the R^2 score and the mean squared error. The R^2 score, also called the coefficient of determination, described how likely the model is to fit a future observation. The mean square error referred to how accurately the model fit the data. All functions were evaluated in python using the *sklearn.metrics* system.

4.3 Results

4.3.1 Utilizing Network Theory to Derive Quantitative Metrics

In order to identify appropriate metrics for characterizing spatial patterns, a flexible and expedient computational framework was necessary to generate a large data set comprised of various types of multicellular patterns. Previous studies reported that embryonic stem cell (ESC) aggregates contained 4 distinct transitional pattern classes associated with the loss of the pluripotency marker Oct4: outside-in, inside-out, random, and connected [263] (section 3.2.4). Here, an *in silico* training set was generated by using pattern generation algorithms to simulate spatial features of differentiation in 3D multicellular aggregates. Due to the high variance of patterns present in the connected pattern class, two additional distinct pattern populations were included within the connected group: globular and snaked. Combined with the starting (Oct4+) and ending (Oct4-) populations, the final modeling dataset contained seven distinct pattern types. The author postulated that counting cell clusters, or sub-networks, would provide discrimination between patterns; therefore, various metrics related to the number, size and distribution of these sub-networks were defined (Figure 4-1). In addition, two sub-networks (Oct4+ and Oct4-) were examined because of Oct4 transitions from high to low expression as a result of differentiation[102, 105].

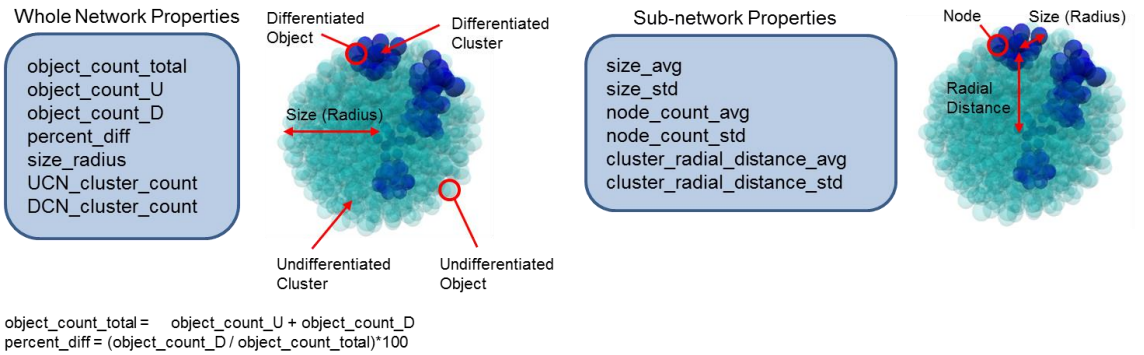

Figure 4-1: Network based metrics for characterizing spatial pattern evolution. The whole network and sub network properties are outlined and annotated to give a visual representation of what each metric represents. In this example, U refers to undifferentiated Oct4+ cells, while D refers to more differentiated Oct4- cells.

Principal component analysis (PCA) was performed to determine if the network-derived metrics captured differences in spatial patterns (Figure 4-2). PCA generated a 2D projection of all samples that revealed some overlap between different pattern classes, particularly for the globular and snaked patterns (Figure 4-2B), but in general the different classes could be readily distinguished. Furthermore, a continuum of patterns was observed (Figure 4-2A) suggesting a natural pattern evolution (starting from random and moving down towards differentiated). The model was able to explain 72.14% of the variance in the data set: 46.5% with PC-1, 13.9% with PC-2 and 11.7% with PC-3. The first principal component represented the stage of differentiation and was positively correlated with variables associated with differentiated clusters or overall differentiation, whereas the second and third principal components were associated with spatial sub-network descriptors. PCA also revealed that most metrics contributed significantly to the model (with the exception of standard deviation associated with Oct4+ cluster size, Oct4+ node count, and Oct4+ radial position), indicating that including most of the metrics were important to comprehensively describe the data (Figure 4-2B).

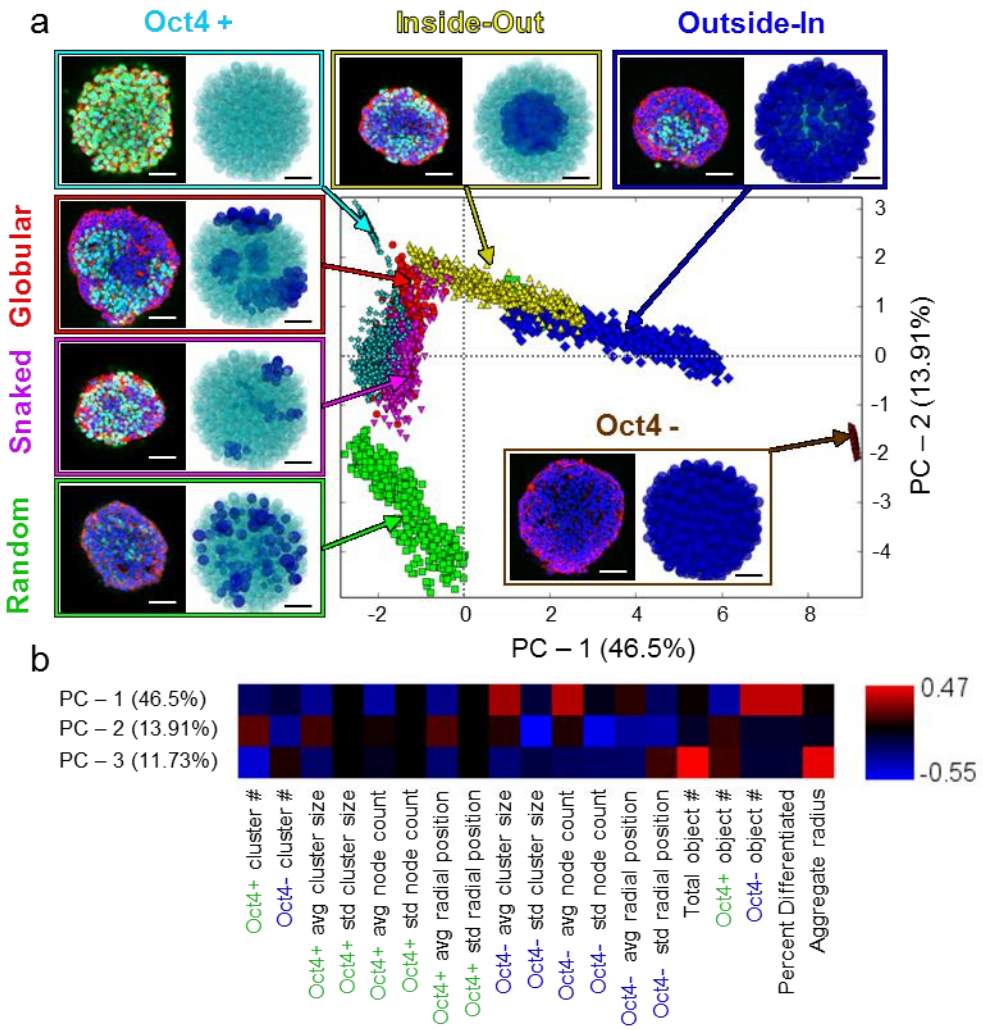

Figure 4-2: Validation of spatial network metrics for pattern classification. (A) Examples of experimental (green = Oct4 Ab, blue = DAPI, red = phalloidin) and in silico generated pattern classes (teal – Oct4+, blue Oct4 -). Scale bars are 100 μm and 35 μm for the experimental and *in silico* images, respectively. Each image is mapped onto the PCA plot of the resulting pattern space, color coded by individual pattern type. (B) PCA axis analysis showing how each metric contributes to the given principal component axis.

4.3.2 Quantitative Pattern Classification via Network Metrics

To quantitatively test how effectively network-derived metrics distinguished different pattern types, a variety of clustering and regression techniques were employed. The k-means and Ward (hierarchical) clustering methods could not sort the different classes of patterns into distinct clusters as indicated by the relatively weak (~0.5) homogeneity and

completeness scores (Figure 4-3A) and the overall cluster distributions. The predictive value of the regression models were also poor with R^2 values generally below 0.4 (Figure 4-3 B) and large mean square error values (Figure 4-3 C). Consequently, regression and clustering did not provide predictive power for pattern identification from the given metric set.

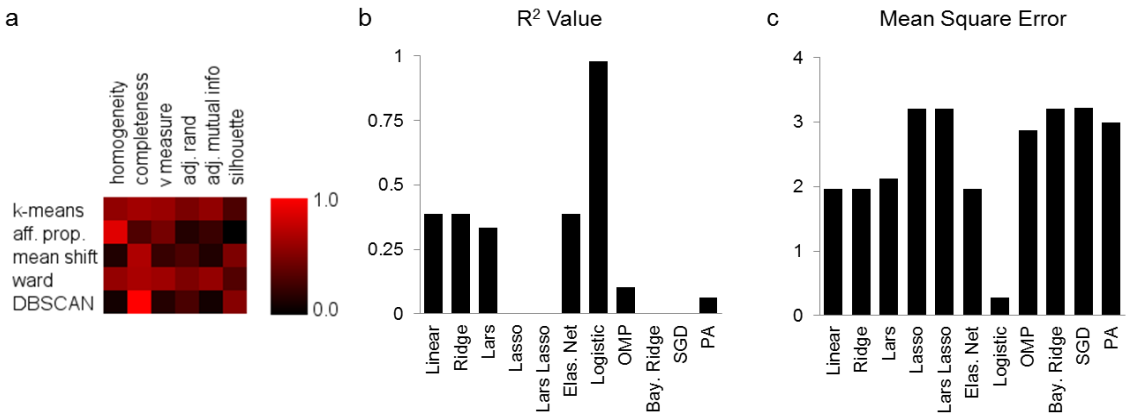

Figure 4-3: Quantitative Assessment of Clustering and Regression for Pattern Classification. (A) Scoring metrics for the efficacy of clustering to sort out different pattern types. The color scale indicates poor scores in black, and good scores in red. (B) The R^2 and (C) mean squared error for the regression techniques when asked to predict pattern types based on network value inputs.

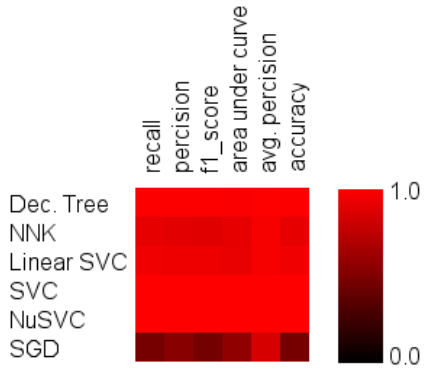

Figure 4-4: Quanitative Measures of Classification Performance. The recall, precision, f1, area under the curve, average precision and accuracy were evaluated for all classification algorithms assessed. Color scale ranges from 0 – 1 where 0 indicates a poor score while 1 indicates the best score possible.

In contrast, logistic regression performed very well (Figure 4-3 B, $R^2 > 0.95$); because logistic regression is actually a classification method, this suggested the need for multi-class prediction approaches. Various classification algorithms were assessed to determine their ability to distinguish between various pattern types: k nearest neighbor (KNN), state vector machines (SVM, NuSVC), stochastic gradient descent (SGD) and decision tree algorithms. All algorithms correctly classified patterns with an accuracy of > 0.95, except for SGD (Figure 4-4). Classifiers distinguished true negatives robustly (precision score > 0.90) while also identifying all positive samples (recall score > 0.89). Classification labeling overlapped with the true labels as assessed via PCA dimensional reduction (Figure 4-5). The SVC classifier had the highest overall accuracy, precision, and recall scores, suggesting it was the most appropriate algorithm for classifying spatial patterns. Furthermore, the probability prediction method allows the classifier to probabilistically predict what class an observation may belong to. One interpretation of these probability assignments was that the pattern could be made up of multiple trained pattern classes, thus representing novel patterns combinations. Taken together, the classification results suggested that novel network based measurements provide a robust set of metrics for classification of complex spatial patterns.

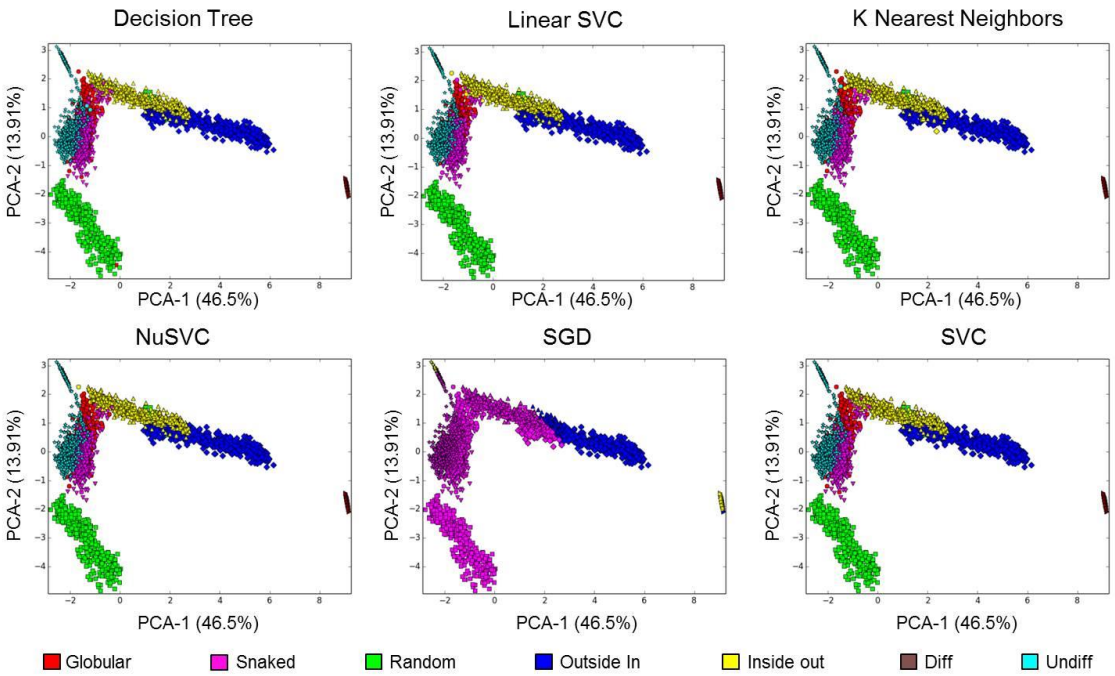

Figure 4-5: Visual Assessment of Classification Performance. The predictions of each classifier for the test set projected down onto a 2D PCA axis (Figure 4-2). Different colors represent different pattern types.

4.4 Discussion

Various different approaches exist for pattern classification depending on the system of interest. In the case of patterns generated in 3D aggregates, supervised classification schemas were able to accurately capture the rich complexity. The issues associated with generating enough representative data for training the classification algorithms were addressed utilizing the previously existing computational modeling framework to create artificial *in silico* pattern representations; which were subsequently analyzed via the network approach and used to train the different classification algorithms. This approach was also convenient as it bypassed the need to collect large amounts of experimental images. Furthermore, because single cell segmentation is performed, simple rearrangements and transformations of the images would not suffice to generate a large enough representative data set of the observed experimental patterns. The resulting

classifiers were extremely robust at identifying different pattern types, and the ultimate state vector machine (SVM) model boasted a 99.9% accuracy score.

The SVM algorithms used in python's *sklearn* toolbox also provide the ability to perform probability fits. Though this slows down the initial fitting, due to an expensive 5 fold cross validation scheme, and lowers the feasible amounts of points which can be used for training, it allows the classifiers to come up with a percent confidence or probability that a given data point belongs to each pattern class. This is based on the underlying approach of SVM, which seeks to define a set of dividing hyperplanes in n-dimensional space which maximize the distances to different sets of observations. When points are sufficiently close to the hyperplanes, the probability estimators may predict that it could belong to either of the two classes separated by that hyperplane. In this work a one-verse-all classification schemes is used, meaning that one classifier is trained for each pattern class. Snaked and globular pattern classes displayed the most overlap between the classifiers, with some slight overlap existing between the inside out and outside in patterns. Based on the principal component (PCA) results showing how the different patterns overlap in metric space (Figure 4-2), this made sense as these patterns shared the greatest amount of overlap in metric space.

The network approach is a novel technique for characterizing spatial patterns, as it extended networks to represent biologic cell representations. The ability to preserve spatial information while also maintaining single cell data is a crucial step forward in the field of biological pattern classification. This approach represents the first method capable of predicting tissue level spatial patterns by utilizing cell level measurements. With the recent advent of advanced light-sheet microscopy techniques capable of imaging through optically dense tissue[264-267] this network approach is currently the only approach capable of utilizing all that information to quantitatively analyze spatial pattern evolution and dynamics through time. Furthermore, this technique can be applied to any network containing cell data, regardless of the image source. To this end different

modalities, such as histological images, confocal microscopy, or even fluorescence microscopy could all be converted to networks and analyzed via this approach.

As mentioned previously (section 4.1), it is important to note that this approach depends on being able to discretize biologic signals. For simulated patterns, this presented no challenge as the states were easily defined by the models. However, in the case of biological signals, things are not always so discrete. The pluripotency transcription factor Oct4 can be easily ascribed to an on or off "state"; however, other signals such as phosphorylation events or expression of Nanog often display continuous behavior. Setting thresholds for the state definitions would be crucial in these cases, and assessment of the spread of the markers present in the networks using a digital flow cytometry like approach would ensure that proper cutoffs are chosen.

This network method relied of phenotypic annotated networks in which each node represented a cell, and the connections in the network represented physical connections of a cell to its neighbors. The cells are ascribed to different states depending on particular biological markers; then for each different annotated cell type or state, the network can be analyzed recursively to extract information about the amount, size, and spatial distribution of these phenotypic subnetworks. This ability makes the algorithm scalable to any number of states present in the network, as the use of subsequent data reduction methods like PCA can map these high dimensional spaces down into few dimensions for visualization. In summary this study detailed the creation of a powerful network based pattern classification system capable of distinguishing biologically relevant patterns.

CHAPTER 5 ELUCIDATING MECHANISMS GOVERNING ESC AGGREGATE DIFFERENTIATION

5.1 Introduction

Model organisms, such as *C. elegans*, *Drosophila*, and zebrafish, are frequently used to interrogate the complex sets of regulatory cues and gene regulatory networks governing morphogenic processes, like gastrulation and neurulation, due to the technical ease in manipulating and imaging [255, 268, 269] these organisms. Pluripotent embryonic stem cell (ESC) aggregates present a complementary, alternative *in vitro* platform for investigating mechanisms of morphogenesis due to their intrinsic ability to differentiate into tissues from all three germ layers and yield formation of a variety of primitive tissues including optic cups [223], human intestinal lining [226] and cerebral organoids [224]. Each of these multicellular systems is highly complex both in terms of the heterotypic cell types that comprise the tissue and the emergent spatiotemporal organization dynamics exhibited by heterogeneous cell populations.

Computational modeling of embryonic development has become an increasingly powerful tool to complement experimental investigations due largely to the fact that increased processing speed has reduced the barrier to multiscale simulations of complex multicellular organismal systems. Several different computational models have been constructed for *C. elegans* [270] and *Drosophila* [271-273] to comprehend the relationships between cell signaling and lineage development in order to gain new insights into the intricate interplay of mechanisms governing development. Stage-specific models have also been developed to examine phenomena such as gastrulation [274] and somite formation [235] at a mechanical level, while computational models of the

formation and differentiation of cells in the early mouse embryo [131] and in mouse ESC aggregates [263] explore mechanisms governing early cell fate decisions. Overall, these modeling approaches have provided a wealth of quantitative data to describe spatio-temporal events associated with morphogenesis; however it remains extremely challenging to relate spatial modeling predictions directly with experimental outcomes due primarily to the difficulty in quantifying multicellular pattern features.

The challenge of quantitatively describing emergent spatial patterns across computational and morphogenic multicellular systems has prevented high-throughput analysis of developmental processes. In experimental systems, divergent phenotypes are often characterized largely by visual inspection, thus lacking the quantitative rigor and objective criteria necessary for direct comparison with computational models. Though several techniques exist to automatically distinguish phenotypes at various spatial scales [252, 254, 275-277], they often lack the resolution of single-cell regulatory dynamics [252, 254], or are customized specifically for investigation of only specific systems [252, 277]. As a result, quantitative metrics extracted from such studies cannot be easily translated between different modes of data analysis or across various model organisms. The object of this study is to derive a portable pattern recognition pipeline capable of handling various biological and computational inputs to enable direct quantitative comparisons on previously unattainable spatial and temporal scales between different multicellular systems.

This study shows that previously defined network analysis schemata derived physiologically meaningful quantitative metrics capable of distinguishing subtleties between spatial phenotypes in a computationally tractable manner. This novel approach

extracts global metrics, such as path lengths and connectivity information, as well as local metrics based on attributes of specific clusters of cell phenotypes. Identification and comparisons of sub-networks within morphogenic systems allow for greater network quantification and significantly enrich the possible metric space by extracting specific subpopulation information. In a proof of concept study, network analysis was applied to spatial networks observed experimentally during differentiation of ESC aggregates and agent based computational modeling of ESC differentiation. This methodological tool provided the first quantitative description of spatio-temporal patterns associated with loss of pluripotency in ESC aggregates, and uncovered a paracrine mechanism capable of explaining the observed differences in spatiotemporal pattern kinetics associated with ESC aggregate differentiation. Consequently, this study described the creation of a powerful and modular pattern identification algorithm with sufficient portability to address meaningful questions about the spatiotemporal dynamics of biological pattern formation.

5.2 Materials and Methods

5.2.1 Network Reconstruction via Cell Profiler

Cell Profiler (http://www.cellprofiler.org/)[278] was used to analyze all of the 2D samples. For confocal analysis, images were imported, split into their component channels (i.e. red, blue, green) and cell nuclei were detected using a local Otsu thresholding approach to provide a binary mask. Clumped nuclei were resolved using the "intensity" module, followed with the "novel propagation function" within Cell Profiler, which led to extended objects which were termed "cells". In each object, the green signal (indicative of Oct4 expression) was measured and reported as the median and mean

value. Additionally, the number of adjacent nearest neighbors was measured and the data was then exported to a python script that reconstructed the networks by using a KDTree implementation from scipy. Images of the networks were generated using the python imaging library (PIL). Annotation was performed by thresholding the Oct4 intensity values received. In the case of multiple aggregates per image, the networks were split such that each individual network contained only one aggregate.

5.2.2 Computational Modeling

Computational modeling was carried out using a previously established framework with some slight modifications. A KDTree implementation as provided in the scipy.spatial package for python was used for detecting and resolving collisions. The form of the rule functions was changed from previous work into a more classical activation and deactivation function:

$$P(x) = \alpha \tag{5-1}$$

$$P(x) = \frac{1.0}{k1^n + {norm_d}^n}\ ,\ norm_d = \frac{d}{\epsilon} \tag{5-2}$$

$$P(x) = 1 - \frac{1.0}{k2^n + {norm_u}^n}\ ,\ norm_u = \frac{u}{\varepsilon} \tag{5-3}$$

$$P(x) = 1 - \frac{1.0}{k2^n + {norm_u}^n} \text{ and } \frac{1.0}{k1^n + {norm_d}^n} \tag{5-4}$$

Where ε denotes the number of neighbors, used in calculated the normalized differentiation (d) or undifferentiated (u) signal based on the number of neighbors for each cell. The constants k1, and k2 represent the set points for activation and inhibition respectively. As previously described, the random differentiation coefficient was kept

constant in all subsequent simulations, but for the random rule by itself α was set to 0.0075. These rules were considered to be satisfied if a randomly generated value was less than the calculated probability.

5.2.3 Modeling Diffusion

Modeling of diffusive species was carried out using a simple Euler forward integration scheme for solution. Though faster solutions exist, the forward integration scheme is stable even when solving reaction diffusion equations with spatially heterogeneous consumption and production terms. The diffusion equation was solved using three separate convolution kernels for each independent direction: x, y and z. The solution was simulated according to equation 5-5.

$$C(t) = \sum_0^t C_{t-1} + D * dt * (lx + ly + lz) + p - q \qquad (5\text{-}5)$$

Where D is the diffusion coefficient, dt the time step of integration, p the source term, and q the consumption term, and l_x, l_y, l_z are the solutions to the 1-D diffusion kernel in x, y, and z respectively. In the case of the source and sink terms, both are assumed to be independent of the concentration. The summation indicates that the solution is solved iteratively until either the upper time limit is met, or a steady state convergence is reached. Steady state was defined as no appreciable change (1E-5 of the current concentration) in the concentration gradient from a given time step to the next. All coefficients and constants were defined such that the final units of the concentration gradient where in μM, with a spatial resolution in μm. The time step dt, was subjected to the constraint in equation 5-6 for conditional stability of the integration scheme:

$$dt = \frac{.5}{D(\frac{1}{\Delta x}+\frac{1}{\Delta y}+\frac{1}{\Delta z})} \quad (5\text{-}6)$$

Where x, y, z represent the spatial resolution of the grid on which the solution was solved. All convolution was carried out using scipy.ndimage convolve function, while a specific gradient class was created for handling diffusion of different soluble species. The rules for differentiation based on diffusion were implemented using classical inhibition equations. In the case of positive induction, where the soluble factor induced differentiation, the probability was defined as is shown in equation 5-7. In the case of negative induction, where the soluble factor inhibited differentiation, equation 5-8 was used. In the case of a combined paracrine schema both equation 5-7 and 5-8 were used.

$$P(x) = \frac{1.0}{k1^n + A^n} \quad (5\text{-}7)$$

$$P(x) = 1 - \frac{1.0}{k2^n + I^n} \quad (5\text{-}8)$$

Where A and I represent the concentrations of soluble activator and inhibitor at the cell location, k1 and k2 represent the half maximal response concentrations, and n governs the width of the resultant sigmoidal probability distribution. In the case where a cell location did not lie directly on a point defined on the gird, cubic interpolation was used to infer the value at these locations. Additionally, a counter was implemented which would keep track of the time delays associated with each soluble signaling factor, which essentially represented a signal duration parameter. This helped to modulate the response time of cell to given soluble signals.

5.2.4 Parameter Ranges

For the various different rules a range of parameters was investigated to give a coherent sampling of the relevant parameter space. These parameter ranges are summarized for each rule in Table 5-1. For the paracrine rules, evaluation of the consumption to production coefficients ratio (denoted p/q) was performed separately as a test before large scale simulation were run. Taken together all of these parameter sets took ~ 2500 hours of cumulative simulation time to run, which on a 64 core cluster was equivalent to approximately 380 hours of run time (~ 15 days) and generated roughly 15 terabytes of data.

Table 5-1: Parameter Ranges For Simulations

Rule Set	Parameter Name	Range
Local Random	A	0.001,0.0025, 0.005,0.0075, 0.01
Local Negative Feedback	α k1 n1	0.001, 0.005, 0.01 0.1,0 .3, 0.5, 0.7, 0.9 10,50
Local Positive Feedback	α k1 n1	0.001, 0.005, 0.01 0.1, 0.3, 0.5, 0.7, 0.9 10,50
Local Competing Feedback	α k1 n1 k2 n2	0.001, 0.005, 0.01 0.1, 0.3, 0.6, 0.9 10,50 0.1, 0.3, 0.6, 0.9 10,50
Paracrine Positive	α k1 n1 p/q	0.001, 0.005 0.006, .008, .01, .012, .014 10, 50 10, 50, 100
Paracrine Negative	α k1 n1 p/q	0.001, 0.005 0.006, 0.008, 0.01, 0.012,0 .014 10, 50 10, 50, 100

Table 5-1: Continued

Paracrine	α	0.001, 0.005
Competing	k1	0.006, 0.008, 0.01, 0.012,0 .014
	n1	10
	k2	0.006,0 .008, 0.01,0 .012, 0.014
	n2	10
	p/q	10, 50, 100

5.3 Results

5.3.1 Converting Experimental Images to Digital Networks

Images of Oct 4+ to Oct4- transitions during differentiation of 3D murine ESC aggregates were acquired experimentally to determine how well network metrics captured an *in vitro* dynamic biological process. Since pluripotent differentiation is known to be modulated by aggregate size [250, 279, 280] , two starting cell densities were examined (250 and 1000 cells/aggregate). In order to apply the previously derived network metrics for pattern analysis, experimentally-obtained confocal images were converted into a network representation of the cells using a digital reconstruction pipeline (Figure 5-1**)**. Briefly, this involves splitting the image up into component channels, and the performing segmentation of the cells based no nuclei, which is this case is on the blue channel. In this specific case cell bodies were the inferred using a propagation algorithm included in the Cell Profiler software suite. A representative time course of differentiation in 1000-cell aggregates (Figure 5-2) demonstrates the fidelity of this process in accurately converting images into annotated networks.

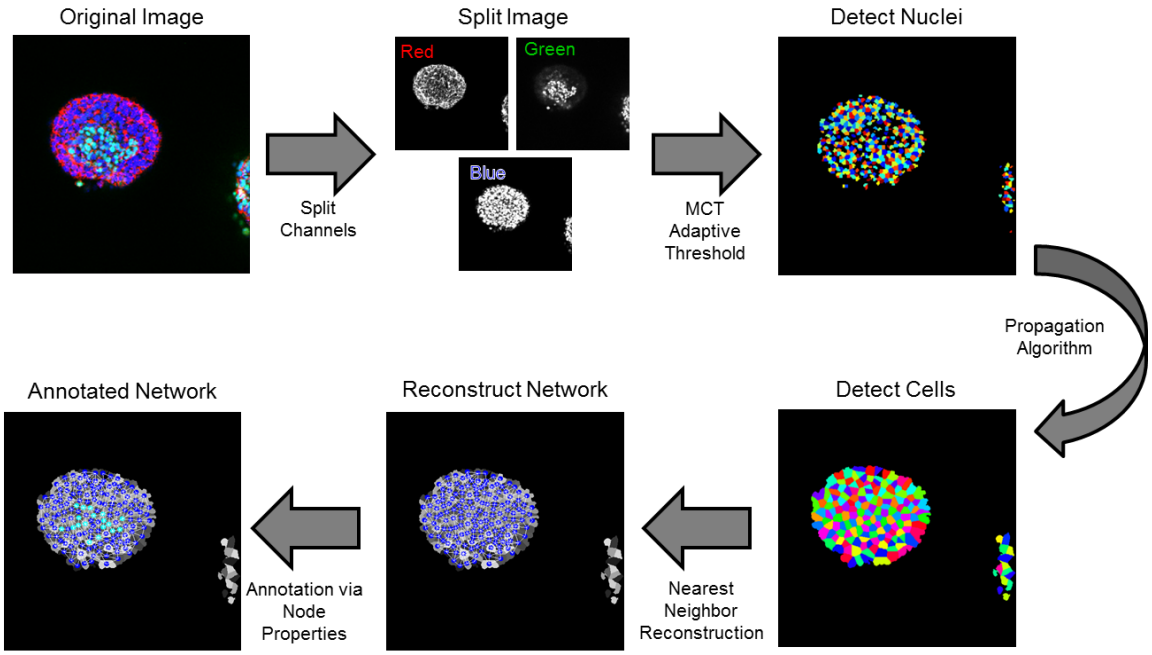

Figure 5-1: Network digitization process for 2D aggregate images. The process of splitting images and detecting cells is performed in Cell Profiler, while the subsequent steps are performed utilizing python.

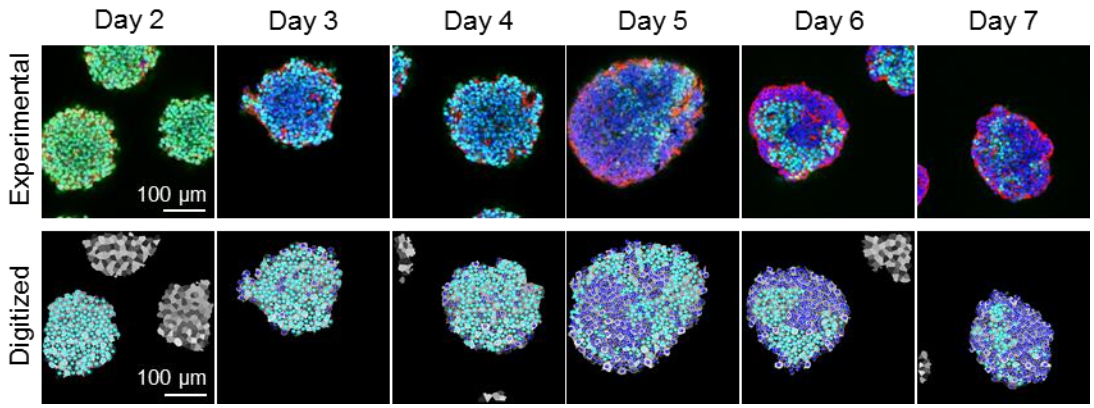

Figure 5-2: Fidelity of network digitization on experimental images. Experimental images (Oct4 green, DAPI blue, phallodin red) on top to reconstructed networks on bottom (blue, Oct4-, teal Oct4+). Scale bars are 100 µm.

Once the networks are reconstructed, a cell state was next assigned based on phenotypic information. For this process a threshold was defined, below which cells were considered negative for the marker, and above which they were considered positive. In this work the

transcription factor of interest Oct4 undergoes a binary state transition during this time period, making this assumption justifiable. However, it is important to note that not all systems behave in this manner, and thus choosing a threshold for biological markers which display a continuum of expression can be complicated. To ease this process, this network based approach also enables image based cytometry analogous to flow cytometry results in which each cell can be plotted as a function of the various markers.

5.3.2 Spatial Pattern Associated With Differentiation

After acquiring confocal images of ESC aggregates of various sizes undergoing the differentiation process (Figure 5-3 A), the resultant digitized ESC aggregate networks were analyzed with the aforementioned network metrics. To analyze the trajectories of the aggregates over time, principal component analysis (PCA) was performed. PCA revealed an average trajectory through the latent variable space for cell aggregates in which all cells began in an undifferentiated state and proceeded through a transitioning period until finally settling into a differentiated state (Figure 5-3 B). The PCA model explained 76.1% of the variance in the data: 43.6% from PC-1, 22.2% from PC-2, and 10.3% from PC-3 (Figure 5-3 C). All metrics significantly contributed to at least one principal component, suggesting that network-derived metrics capture the variance of biological spatial patterns. PC-1 represented differentiation, while PC-2 and PC-3 again correlated with spatial sub-network descriptors representing inter-pattern variation. The transitioning state was separated from the Oct4+ and Oct4- states largely by PC-2, suggesting that aggregates moving through the transitioning state had a high degree of intra-aggregate heterogeneity. The principal component metric weights for the experimental data closely mirrored the weights for the *in silico* training set, indicating

that network-derived metrics comprehensively capture the inherent biological variance that transpires during the course of ESC aggregate differentiation.

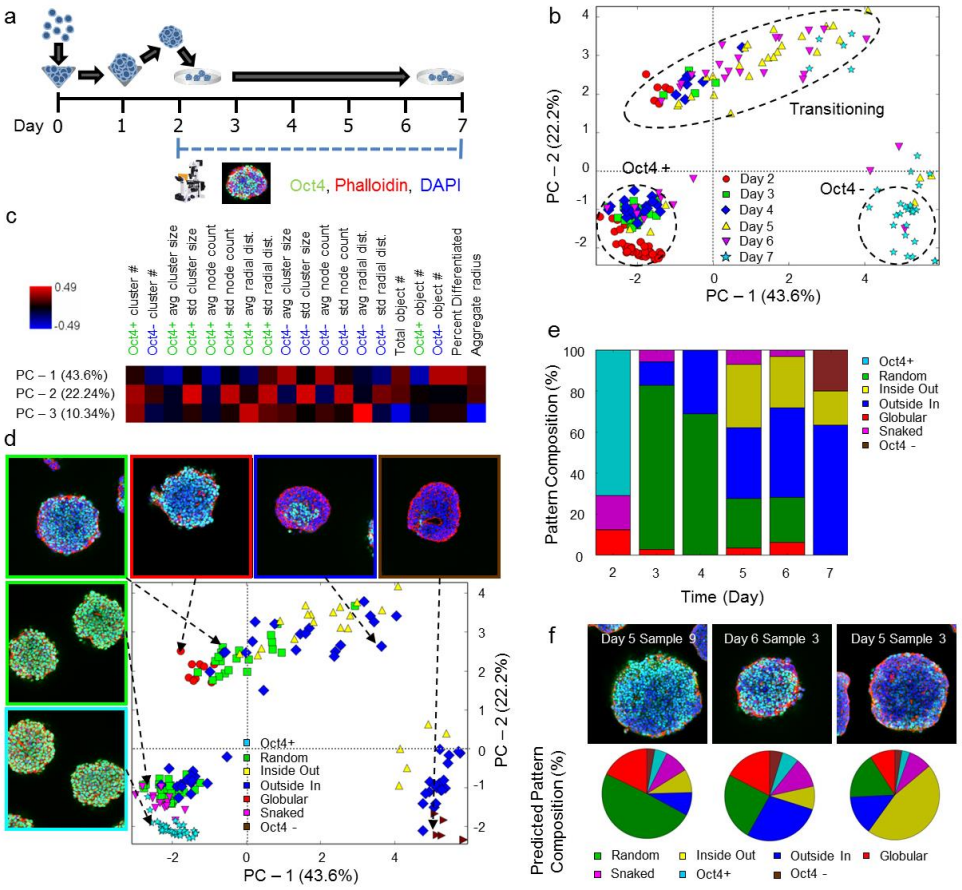

Figure 5-3: Evaluation of spatial and temporal patterns during early ESC aggregate differentiation. (A) Shows the experimental plan and timeline. Samples were removed and analyzed for confocal microscopy on days 2-7. (B) PCA trajectory for 1000 cell aggregates. Circles highlight the states present. (C) Heat map of the weights of each metric in relation to each principal component. (D) Pattern annotations for what pattern class each sample falls into. (E) The pattern composition for each day of differentiation. (F) SVC classification identifies patterns which contain aspects of multiple core pattern classes. Top row shows experimental images, bottom row shows predicted pattern composition.

While the Oct4+ and Oct4- states were quite distinct, the intermediate transitioning period displayed a great amount of variance (Figure 5-3 D). To determine if the variation

was due to differences in spatial patterns, classification was performed using the previously trained SVC classifier to characterize the distribution of spatial patterns in each state. Classification indicated that the initial state consisted largely of a mix of undifferentiated, random, and outside-in patterns, while the final state consisted of a mix of entirely differentiated and outside-in patterns (Figure 5-3, D,E). The variation with respect to each time-point peaked at days 5 and 6 (Figure 5-3B), and days 5 and 6 also displayed the most variation in spatial patterns (Figure 5-3 E), reflecting that the variation was due to the presence of diverse spatial patterns. Furthermore, SVC classification predicted that many of the aggregates at day 5 and 6 could belong to multiple pattern classes, indicating that these patterns were more spatially complex and thus displayed components of multiple different pattern types (Figure 5-3 F). Overall, spatial pattern evolution progressed in the following temporal order: undifferentiated, snaked, random, globular, inside-out, differentiated.

Interestingly, similar pattern trajectories were also observed for the 250-cell aggregates (Figure 5-4). The Oct4- state was a little more compact indicating a more complete differentiation than in the 1000-cell trajectory (Figure 5-4 A). Overall, the 250-cell aggregates underwent a much faster differentiation, over approximately a 24 hour period (starting at day 4), than the 1000–cell aggregates (Figure 5-4 C). This discrepancy in differentiation trajectories is consistent with literature showing different differentiation rates and fates based on aggregate sizes [250, 279-282]. These results represent a biological trajectory describing spatial pattern evolution in a portable quantitative fashion, but even more importantly, the analysis suggests that early differentiation in ESC aggregates progresses via quantifiable spatial patterns that do not display purely random

characteristics. Furthermore, this is the first description of biological trajectories utilizing information from single cells to capture spatial pattern complexity.

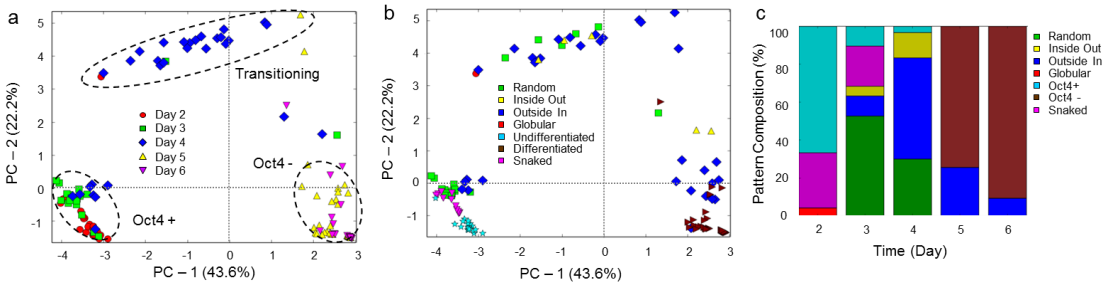

Figure 5-4: 250 cell/aggregate pattern trajectory. (A) The differentiation trajectory where each point represents an individually analyzed network. The circles denote which state the observation falls into. (B) Individual networks color-coded by pattern type. (C) Pattern distributions over the 5 day differentiation time course.

5.3.3 Simulating Rules Based on Diffusion

It has long been speculated the ESCs secrete molecules which affect the differentiation process [106, 258] and one prevalent example of this is secretion of the small peptide leukemia inhibitor factor (LIF) in ESCs[116, 283]. While various studies have shown that LIF is important in maintaining stem cell pluripotency, it is always difficult to measure the quantitative distributions of such gradients inside complex 3D structures. While techniques like FRAP can capture diffusive metrics, this requires small labeled dies whose diffusion coefficients differ by a few orders of magnitude from soluble protein species. One way to infer actions of these molecules is to use computational modeling approaches. However, while many models have been constructed to examine growth factor or nutrient penetrance into aggregates, only a few have been constructed to examine heterogeneous spatial evolution of gradients resultant from intra-aggregate paracrine signaling. This is largely due to computational limitations associated with the introduction of various spatial terms, which transforms the problem into a coupled reaction diffusion problem. Solving these problems requires simulation of each time step

iteratively as no steady state solutions exist, which represents a non-trivial computational cost.

To examine the effects of spatially heterogeneous consumption and production terms, a simple proof-of-concept study was performed. In this simulation a single cell in a small 100-cell aggregate was set to a differentiated state. A single paracrine factor was secreted which induced differentiation. The ratio of consumption to production of the factor was varied to ratios of 1:100 (Figure 5-5 A), 1:50 (Figure 5-5 B) and 1:10 (Figure 5-5 C). The resulting simulations showed that changing this ratio affected the gradient in two ways (Figure 5-5 D). The time it took the gradient to reach a 2E-5 μM concentration increased as the ratio decreased, as did the number of cells influenced or the width of gradient. The width of the gradient was defined as the area of the concentration gradient that was above 2E-5 μM. Ultimately for the subsequent simulations a 1:10 ratio was used as this demonstrated a gradient with the shortest width, which should thus impart the most spatial variability on the simulations.

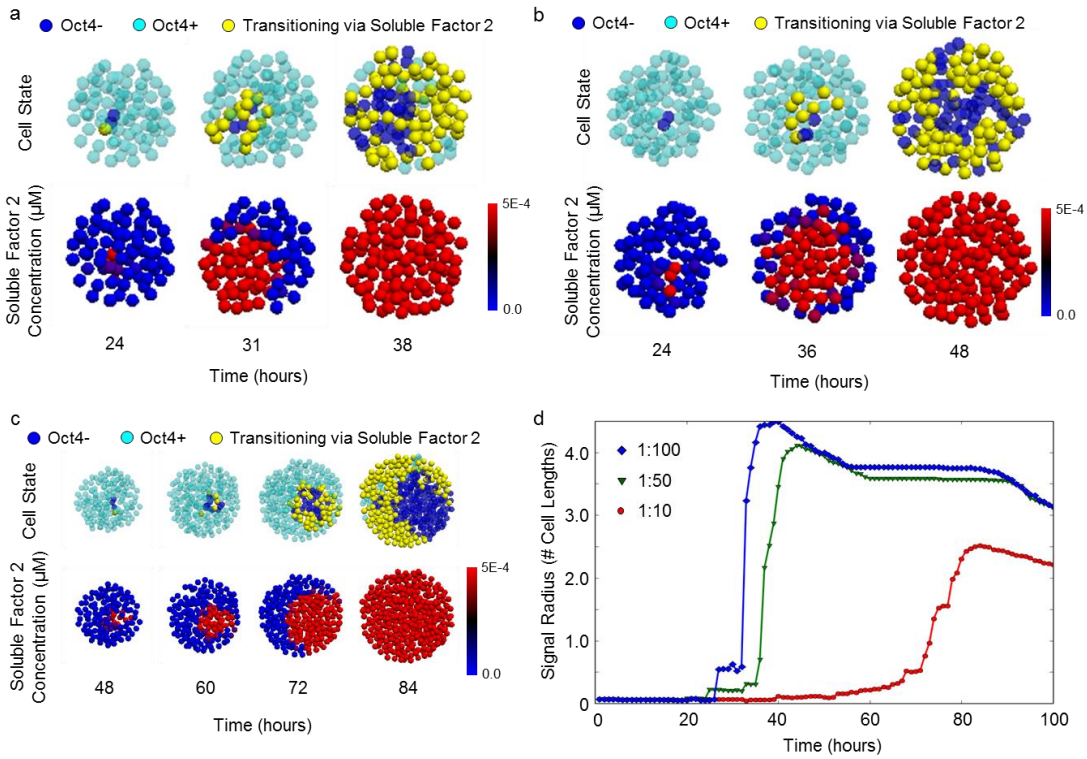

Figure 5-5: Investigation of consumption/production ratios of soluble concentration gradient evolution. Representative transition trajectories for the 1:100 (A), 1:50 (B) and 1:10 (C) consumption to production ratios. (d) The average cell length influenced by the paracrine gradient for the 1:100 (blue), 1:50 (green) and 1:10 (red) consumption to production ratios.

5.3.4 Comparing Computational with Experimental Pattern Trajectories

Next, to probe the mechanisms governing the formation of spatial patterns associated with differentiation, an agent based modeling approach was employed in which cells are allowed to proliferate, migrate, and differentiate within a 3D aggregate configuration[263]. In prior work, a simple set of rules based on local neighboring cell state(s) was used to govern changes in cells state; however, comparisons between modeling results were nearly impossible because a quantitative set of descriptors for assessing spatial patterns did not exists. In addition, comparison with experimental data could not be directly accomplished without a validated digitization strategy for

experimental data. These challenges were addressed simultaneously by enabling direct comparison between spatial patterns from computational models and experimental results via the use of these newly defined network metrics.

Seven models with different rule schemes driving differentiation were investigated: random, local positive feedback, local negative feedforward, local competing feedback, paracrine activation, paracrine inhibition, and combined paracrine activation/inhibition (Figure 5-8 A). Simulations were carried out over a six-day period with different initial cell densities: 250- and 1000-cells/aggregate. Each rule set was simulated with multiple parameters to explore the breadth of pattern trajectory space (Table 5-1). PCA using the metrics described in figures 1 and 2, captured 76.5% of the simulation variance: 48.83% from PC-1, 17.7 from PC-2, and 9.9% from PC-3 (Figure 5-6). Again, PC-1 represented differentiation, while the PC-2 was influenced by standard deviations in sub-network measurements, correlating with the formation of spatial patterns in the simulations.

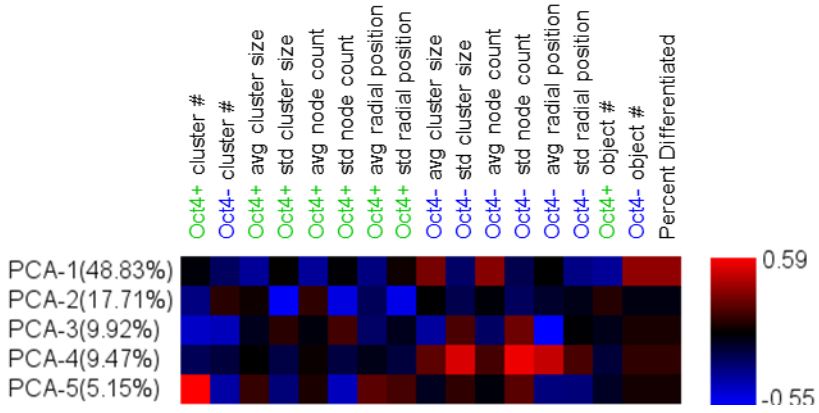

Figure 5-6: Principal Component Analysis for picking axes of largest variance. First 5 principle component axes which together explain ~ 90% of the variance present in the data. Red and blue indicate a positive or negative contribution respectively of the metric with the axes of variation.

The previously trained pattern classifiers were applied to assess the spatial patterns generated by the computational simulations and analyzed using hierarchical clustering (Figure 5-8 B). While the competing, negative feedforward, and positive feedback rules all generated similar pattern distributions and trajectories, the paracrine rules generated a more diverse set of pattern types. When hierarchical clustering was performed across all rule sets simultaneously, a wide variety of behaviors emerged (Figure 5-7). The largest difference was between parameter sets which reached a terminal differentiated state, versus those that did not. These results indicate that various complex pattern evolutions and temporal kinetics can be achieved using parsimonious, generalized rule sets.

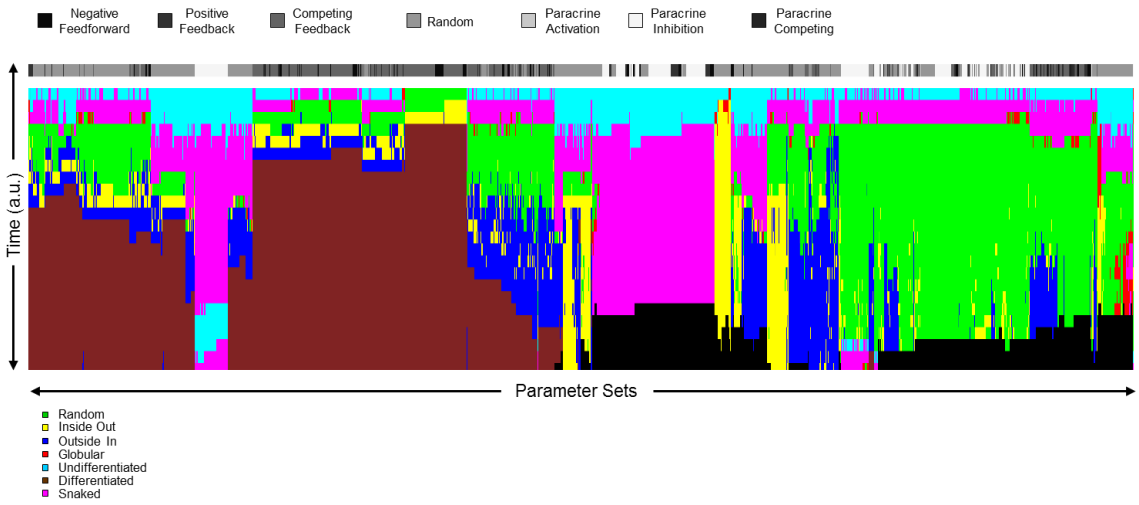

Figure 5-7: Hierarchical clustering of pattern trajectories across all rule types. The top bar indicates what rule type the simulated pattern trajectory belong to, while the interior heat-map colors indicate the pattern type which the simulation is progression through at that time.

A powerful feature of the network-based methodology is the ability to directly compare results across different platforms, which allowed the modeling and experimental data sets

to be merged into a single metric set. PCA was used to assess which metric axes were most important for describing the variation, resulting in a set of 5 axes responsible for ~ 90% of the variance along which to compare the different data sets (Figure 5-6). Previously it was postulated that a competing feedback scheme could capture the spatial pattern evolution during differentiation, but this mechanism failed to explain the kinetic differences between 250 and 1000-cell differentiation trajectories. To identify parameter sets and rules that did modulate differentiation based on aggregate size, a ratio of the differentiation rate of 250-cell to 1000-cell differentiation rate was calculated (Figure 5-8 C). This ratio confirmed that local feedback rules did not exhibit significant size dependent differences, while nearly all of the paracrine rules did. Both the paracrine activation and competing paracrine rules resulted in slower differentiation of 1000-cell aggregates than 250 cell aggregates, matching experimental observations. However examination of the paracrine activation rule revealed that this difference was due to 1000-cell aggregates not differentiating completely. By comparing these rules to the experimental data on the PCA axes derived previously, it could be determined that the paracrine competing rule set yielded the best fit (Figure 5-8 F). This rule accurately captured both the relevant time scales for differentiation (~24 hours in 250 cells/aggregate and ~48 hours for 1000 cells/aggregate) and the spatial pattern evolution (Figure 5-8 D). Furthermore, this rule suggested that in the 250 cells/aggregates, differentiation was primarily induced by the absence of activator for the pluripotent state, while at the 1000 cells/aggregate size differentiation was largely caused by the buildup of a factor which induced differentiation (Figure 5-8 E). Taken together these results demonstrate a non-intuitive paracrine mechanism that can accurately explain

differentiation of ESC aggregates and hence demonstrate the power and utility of network based metrics for elucidating new mechanisms governing biological processes.

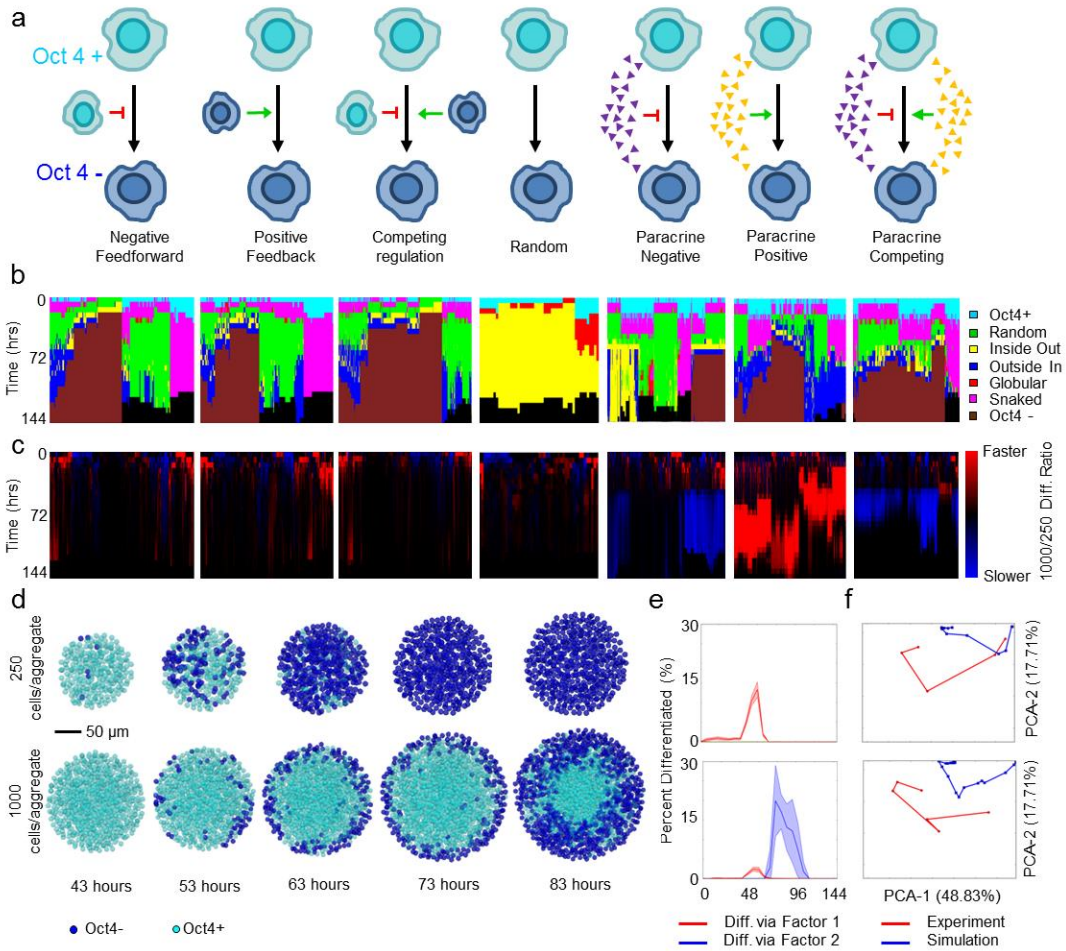

Figure 5-8: Characterization of computationally pattern trajectories and comparison to experimental patterns. (A) Rules governing modeling. Red arrows, negative feedback, green arrows positive feedback, gradient boxes indicate a soluble morphogen gradient. (B) Hierarchal clustering using a ward linkage algorithm of the pattern trajectories for each parameters set, annotated by rule (gray scale bar). (C)The ratio of the 250-cell to 1000-cell differentiation rate. Blue represents slower differentiation in 1000-cellular aggregates, while red represent faster differentiation. (D) Representative images showing the differentiation trajectories predicted by the competing paracrine rule. Cells in blue are Oct4-, cells in teal are Oct4+. (E) Percent differentiation for representative 250 (top) and 1000-cell (bottom) simulations. Red represents differentiation induced via absence of soluble factor one, while blue represents differentiation induced via the presence of soluble factor 1. (F) PCA projections showing the fits of the 250 (top) and 1000-cell (bottom) experimental data (red) to the best fit simulations (blue).

5.4 Discussion

Network and information theory provide a powerful tool for the analysis of many complex systems ranging from social[284, 285] to biological networks[286, 287]. For the first time, our work applies the principles of network theory to the study of morphogenic biological systems in a spatial manner. Increasingly, examples of emergent spatial patterns are being reported from initial pluripotent states, leading to organoids such as optic cups, cerebral tissues, or others[121-123], however, quantitative descriptions of multicellular patterns are still lacking. The quantification could be used in a variety of ways: increased QC of organoid structures, elucidation of cellular dynamics during structure formation, comparison between different structures. Ultimately this technology and approach lends itself well to high throughput screening methods for

While a few strategies to examine spatial patterns in ESC aggregates have been examined, the extracted metrics are generally not sophisticated enough to capture complex structure formation. For example Warmflash *et al*. recently used radial distance to delineate organization of differentiated phenotypes within micro-patterned ESC colonies [223, 224, 226, 288]. While this radial distance worked to some extent for capturing the different cell type distributions in their system, this was an artificially patterned system 2-D in which ring-like region so of differentiation emerged. Thus, the method would not be extendable to formation of germ layers happening in a more heterogeneous system such as a cellular aggregate. Herberg *et al.* used a similar method to compare spatial distributions of proteins in ESC colonies to computational models[258]. Here they calculated the variation on a per pixel basis for each colony of stem cells, and use this radial variation metric to calculate spatial heterogeneity within the colony. While this approach works somewhat well for 2D cultures, and could be adapted for 3D methods, it has some flaws. First, such methods will not give any information

about interacting cell types and will not scale well to multiple different interacting biological stimuli. Second, this approach is not sensitive enough to pick out more complex spatial patterns. The patterns distinguished are generally radial, and the calculated metric is more a measure of combined intensity than it is of any spatial distribution of these factors. For each colony, a single metric is extracted and subsequently used to distinguish between different pattern types. This does not readily allow for the use of multivariate approaches to classify different types of patterns, and it debatable whether the metrics presented have the resolution to do so. However, Herberg *et al.* do take their approach a step further than Warmflash *et al.*, and compare the extracted metric to a computational model to show a good fit, which demonstrates that in theory the comparison of extracted model and image metrics is possible. To the author's knowledge, this analytical approach uniquely enables the first direct quantitative comparison between computational modeling and complex emergent spatiotemporal patterns during multicellular lineage commitment in 3D ESC aggregates.

This method reconstructs cellular locations as interacting networks that can subsequently be further subdivided into biologically relevant sub-networks. As mentioned previously this network-based approach circumvents problems associated with traditional classification methods that rely solely on standardized images [252] and use of individual pixel classification methodologies [236, 289]. While some systems exist for classifying spatial patterns in zebrafish [277], *C. elegans* [275] and *Drosophila* embryos [252], previous approaches require specifically orientated and annotated images, are specific to the organism of interest, and/or often do not have single cell resolution. This method is the first approach capable of integrating single cell information to quantitatively describe tissue level patterns. Single cell information integrated with the power of network based theory allows the extraction of both spatial and temporal data about the evolution of patterns.

This work shows the derivation of spatial pattern trajectories associated with experimentally observed loss of Oct4 in ESC aggregates, and computational models of Oct4 loss in ESC aggregates. This network analysis approach highlights a potential novel mechanism in ESC differentiation by suggesting that ESC aggregate spatiotemporal pattern kinetics can be explained by a combined paracrine signaling methodology. A large body of literature exists which suggests that differentiation is heavily modulated by ESC size [204,234,243-245]. However, as of yet, no one has been able to determine the mechanisms which govern this size-dependent regulation. The two paracrine process proposed here can explain these differences (a secreted factor responsible for maintaining pluripotency, and a factor which induces differentiation), and mirrors the known properties of soluble LIF and FGF4 signaling respectively [102, 106, 110, 290]. In fact stem cells are known to secrete both factors in the pluripotent state, and FGF4 has been shown to reduce Nanog expression, causing cells to enter a "primed" state, which makes them more responsive to differentiation cues.

It is worth noting that while the local rules are capable of capturing the spatial pattern evolution, they can't explain the temporal kinetic differences associated with differentiation of different size aggregates. This apparent lack of local neighbor-to-neighbor regulation of phenotypic state, as analyzed by our methodology, suggests that transmission of cell state information by intercellular cues, such as Notch, may impact later stages of differentiation than the time period examined here. However, in the mathematical analysis of diffusion consumption and production ratios (Figure 5-5) it was noted that a wide variety of behaviors can be attained via paracrine signaling. When the consumption to production rate is high (Figure 5-5 A), the concentration gradient influences a wider range of cells, and thus tends to unify patterns over space. However, at low levels (Figure 5-5 C) paracrine effects only reach a few cell lengths, leading to local pattern formation which mimics the contact mediated interactions.

In summation, this novel pattern classification pipeline permits entirely new forms of quantitative analysis based upon the fundamental interconnectivity of multicellular networks, which could revolutionize the characterization of biologically complex spatiotemporal phenomena.

CHAPTER 6 APPLICATIONS OF NETWORK ANALYSIS TO OTHER BIOLOGICAL SYSTEMS

6.1 Introduction

The work in previous chapters focused primarily on the Oct4 transition, as this is one of the first cell fate decision made during development, to try and understand biological mechanisms regulating this transition. In particular, the network based analysis technique developed in Chapters 4 and 5 provided unique biological insight into how soluble cues interact to form complex spatial patterns. However, while Oct4 transitions are important in the context of early cell fate decisions, ultimately the scalability and biological relevance of network based pattern analysis needed to be addressed. Along these lines, several key questions remained: could this approach scale to larger aggregates of cells, was analysis of markers which display gradient patterns of expression possible, does the method scale with the number of markers, and can it provide novel biological insight in the systems examined? To address these questions, three different model systems were analyzed.

The first system investigated at gastrulation in cichlid fishes. Specifically expression of distless homeobox gene (*dlx3b*) and BMP signaling via downstream phosphorylation of SMAD were monitored to examine formation of the neural plate during gastrulation. Though it was known that dlx3b was involved in the clearance of BMP from the neural plate, the complex spatial interplay of the two signaling molecules was still unclear. This work challenged the network based architecture by examining a marker (phosphorylated Smad) which displayed expression in a gradient fashion, as well as scaling the network size to thousands of cells. Furthermore, this approach yielded novel insight into the evolution of *dlx3b* gene expression with respect to BMP induced signaling.

The next system investigated formation of neural progenitor domains via a process known as neurulation which occurs after gastrulation during development. This progenitor domain will first form motor neurons and then subsequently differentiate towards supporting glial cell types. This transition from motor neuron to glial cell types is commonly referred to as the glial switch. In this work, an ESC aggregate system was used to model the formation of neural progenitor domains, and subsequent differentiation towards motor neurons and glial cells. By examining the spatial patterns associated with this switch like behavior this work examined whether network analysis was sensitive enough to distinguish between soluble and secreted cues in the context of a complex differentiation pathway. Ultimately this approach yielded novel insight and suggested that a soluble factor secreted from motor neurons, which negatively regulated motor neuron differentiation from progenitor populations, was responsible for the switch to glial cell production.

The final system investigated was the formation of mesenchymal like regions in ESC aggregates which were thought to occur via an EMT-like process. This study examined whether network based analysis and properties could aid in the classification of various different cell structures present in histological images. Network based metrics were shown to be extremely useful for classification of several different cell types including mesenchymal like cell types, and dense epithelial regions. Furthermore, representation of aggregates as single data points in a reduced metric space allowed novel comparisons across experimental systems, as well as examination of intra and inter-aggregate heterogeneity. Ultimately this work showed that network based spatial pattern analysis was capable of analyzing complex spatial pattern dynamics to provide unique biological insight across multiple systems in a portable fashion.

6.2 Gastrulation in Cichlid Fish – The role of dlx3b and BMP signaling

6.2.1 Background

To assess the applicability of network-based analysis on a tightly regulated biological process involving multiple biological signals, gastrulation of east African cichlid fish embryos was analyzed. East African cichlids represent a fascinating organism due to the adaptive radiation of the species, and the variety of isolated environments in which they live [301-305]. Specifically the cichlids in Lake Malawi have evolved into species which can be grouped into two larger groups: rock and sand. The sand dwellers feed primary at the surface, while the rock dwellers feed closer to the bottom. Due to the different cues present in the two environments, the brains of the fish develop quite differently. Work in the Streelman lab has shown that differences in brain development can be traced back to differences in early developmental stages, specifically formation and specification of the neural plate during gastrulation [306].

Gastrulation begins from a relatively undifferentiated cell aggregate that undergoes coordinated multicellular movement and differentiation to yield three tissue layers, a neural plate and rudimentary gut [307]. This fundamental developmental process occurs under the tight spatial and temporal control of morphogens, such as bone morphogenic protein (BMP), and subsequent activation of downstream Smad signaling via phosphorylation (designated as pSmad). BMP signaling moves in a dorsal-to-ventral direction across the entire embryo during gastrulation to eventually form a pSmad gradient [307] and the subsequent amount and rate of BMP removal correlates with expression of the distless homeobox gene, *dlx3b*[308]. As with pSmad activity, expression of the *dlx3b* gene went from being ubiquitous across the majority of the

embryo to specifically and strongly expressed in the neural plate boundary. The gene *dlx3b* is known to be important for specification of the preplacodal endoderm, and is crucial in establishing the boarder and evolution of the neural plate [309]. While the spatial localization of *dlx3b* is known throughout development, it is currently not well understood how the spatial temporal relationship between *dlx3b* expression and BMP expression varies in different species of cichlid fish. Assessing the temporal and spatial patterns of multiple correlated signals during morphogenic processes represents a powerful new application of network analysis which could provide insight into the species specific regulation of these cues during gastrulation. In this work, the properties of dlx3b gene expression were assessed with the hope of establishing a pipeline for quantifying spatio-temporal gene expression differences in cichlid species.

6.2.2 Methods

6.2.2.1 Three Dimensional Network Reconstruction

First the confocal images were read by ImageJ, merged into a single channel and then saved as an image sequence. The image sequence was read using python and converted into a memory mapped array via the numpy package, which allows for analysis of large arrays that would normally exceed the amount of memory present in the computer's RAM. Images were split into respective red, green, and blue components and then denoised using a Gaussian filter from scipy's ndimage package. Initial thresholding was performed on the blue channel using a global Otsu approach from the python package skimage to identify grouped nuclei. Local maxima detection was performed to segment nuclei using the python package skimage and once detected, a merging step was performed to identify local maxima that were too close to each other, and these were

merged into a single new local maxima point. The local maxima points were then converted into seeds for nuclei detection and served as the subsequent nodes in the network. Connections were formed using a nearest neighbor approach using a KD Tree implementation from scipy, in which only neighbors within a certain distance (twice the cell radius) were connected to each other. The cells were then annotated by computing the average red and green values within radii around the points. A global threshold over all images was established for the red and green channels using an Otsu thresholding approach over all of the nodes. The networks were then filtered to remove unconnected subnetworks, and further filtered using a quality metric that excluded all nodes with a low blue signal in order to produce the final annotated network.

6.2.2.2 Feature Elimination

Feature elimination was performed utilizing a combination of techniques. A first approximation of important features was extracted by creating box plots of all metrics across all embryos examined. If a feature did not display a sufficient amount of variation, then it was removed for subsequent analysis. In this specific case, the cichlid markers tended to display as single, well connected clusters, making most of the metrics measuring variation between clusters irrelevant. Thus these metrics did not display enough variability across the embryos and were subsequently removed. The technical term for this process is variance thresholding, and it is one of the simplest feature selection tools available. To back up these results, univariate feature selection was also applied. In univariate feature selection the SelectPercentile function in SKlearn was applied. Recursive feature elimination, which uses recursion and linear fits, was also performed.

6.2.2.3 Cichlid Maintenance and Culture

Cichlid maintenance and culture was performed by Jonathan Sylvester and Amanda Ballard. Several species of East African Cichlids were kept as brooding populations in 40 gallon tanks, in two species per tank configurations, with total individuals equaling 30 to 40. Male to female ratio is typically 1 to 4 up to 1 to 10, depending on species. Fishes were allowed to spawn naturally, then broods were taken from the female's mouth approximately 24 hours post fertilization (hpf). A brood consisted of 20 to 80 eggs, depending on the species. Broods were grown in 150 mL flasks, in a mixture of tank water from which the mother lives, and methylene blue, to prevent fungal growth. A subset of individuals was taken from each brood at 36 hpf, and at 4 hour intervals until 48 hpf to cover the entire duration of gastrulation. Embryos were fixed in 4% paraformaldehyde (PFA) at each time point of interest.

6.2.2.4 Cichlid Embryo Staining and Confocal Imaging

Immunostaining and imaging was performed by Jonathan Sylvester and Amanda Ballard. After fixation, expression of *dlx3b* was visualized using whole mount *in situ* hybridization, using a modification of previously published methods [310]. The gene was visualized using Fast Red (naphthol chromogen, Roche Diagnostics). After *in situ* hybridization, embryos were immunostained for pSmad 1,5,8 protein, using previously published protocols [311]. Embryos were then bathed in Vectashield (Vector Labs) containing DAPI and placed in a specially built mold that holds the embryo upright. Embryos were then scanned using a Zeiss LSM 700-405 confocal microscope and processed using LSM 700 software and Image J.

6.2.2.5 Path Finding Analysis

Path finding algorithms were used to assess the likely flow of information through the system. Briefly, a set of waypoint nodes were chosen for the algorithms to pass through in a given order. For cichlid fish analysis, 3 waypoint nodes where chosen, a start node, an end node, and one in the middle of the gastrulation process. Simulations were run to find likely paths or flows of information by assigning a probability function to designate which node should be chosen next. In this case the PDF was directly related to the distance in the PCA reduced metric space. Thus, the resultant paths were reduced to the shortest distance representations in metric space constrained by passing through all waypoint nodes.

6.2.3 Results

Gastrulation in cichlid fish proceeds is coincident with the formation of the neural plate via dlx3b, and the subsequent clearing of the neural plate of phosphorylated Smad protein. To capture the complex spatio-temporal dynamics associated with this process, embryos at various stages of gastrulation were harvested; immunostained for phosphorylated Smad (pSmad) with *in situ* hybridization was performed against dlx3b. Next the embryos were oriented using a microfluidic embryo trap, and imaged using confocal microscopy. This produced 3D confocal image stacks for each embryo.

As many of the previous analysis performed were in 2D, the inherent 3D structure of cichlids presented a challenge, thus a 3D segmentation algorithm was implemented to output annotated spatial networks (Figure 6-1). Using the previously defined metrics, three separate sub-networks of cells were analyzed: pSmad+, dlx3b+ and pSmad+/dlx3b. In addition, several new metrics were added, such as the ratio of pSmad+/dlx3b+ nodes

to dlx3b+ nodes and pSmad+ nodes, as well as circularity and eccentricity measures for cell clusters in an effort to capture the additional spatial complexity of this system. This approach was subsequently validated by the software package IMARIS.

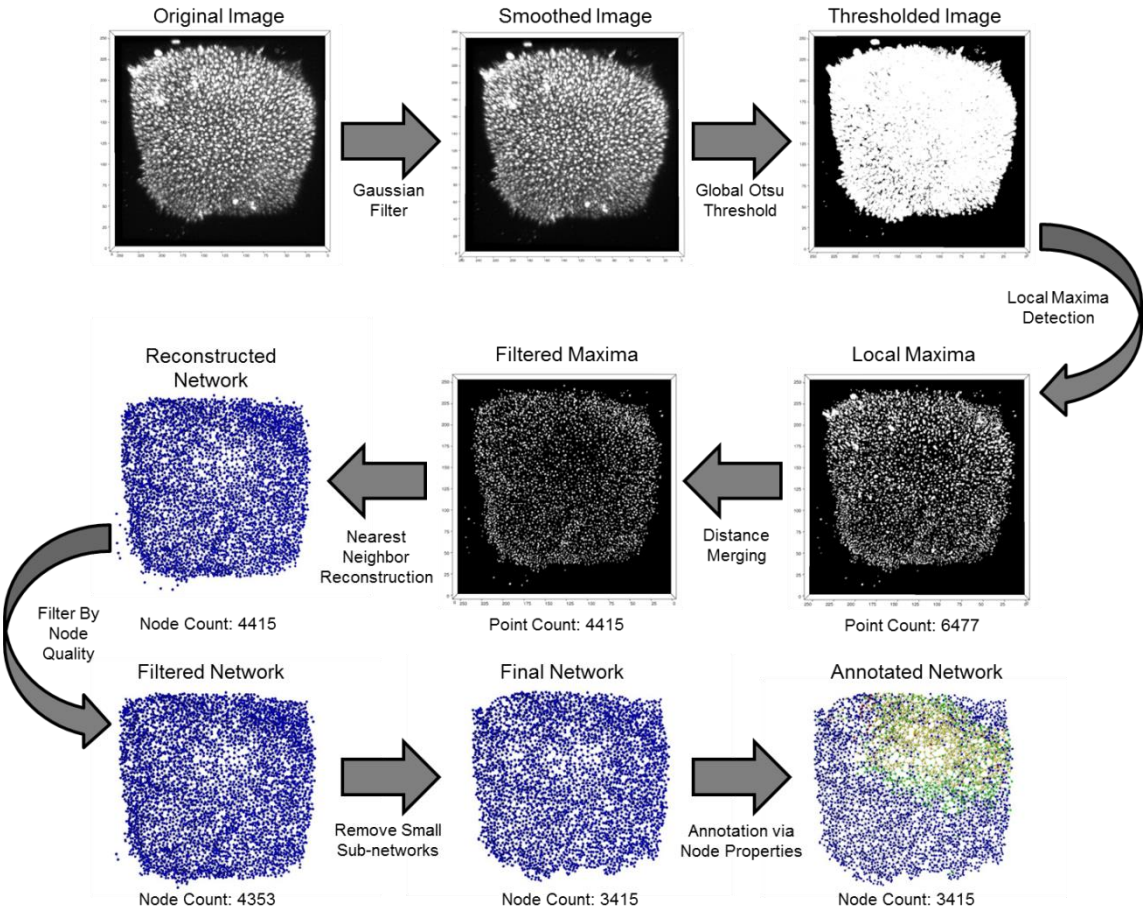

Figure 6-1: Network digitization procedure for 3D confocal cichlid images. A Gaussian smoothing filtered was applied to filter out noise in the image. Next the image was thresholded using a local Otsu method. Next local maxima were detected using a local maxima filter. A simple thresholding and watershedding segmentation approach is used to separate individual cells using the maxima points as seeds. Next the network is reconstructed using a KD-tree method to infer cell connections. Final networks were generated by filtering out small unconnected nuclei.

One problem with large multivariate data sets is choosing metrics which accurately describe the variation in the population. Metrics which do not change across the

population can confound subsequent analysis including classification, dimensional reduction, or clustering. For this data set many of the metrics which measure variation between clusters were not applicable as there was only one primary cluster identified for each signal type. To identify which features should be excluded from the data set a simple variance threshold method was applied, which only keeps metrics displaying a certain amount of variance within the population (Figure 6-1 A,B). This method confirmed the exclusion of metrics related to the standard deviation in clusters.

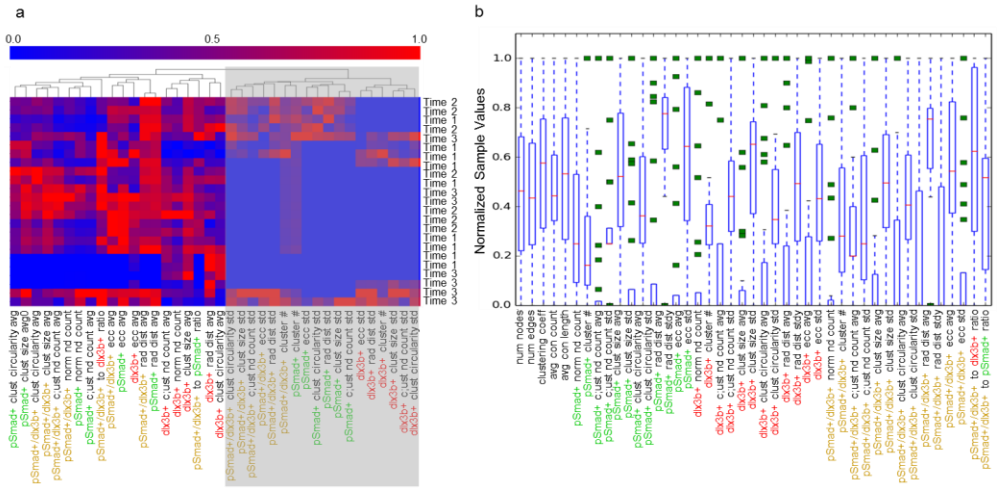

Figure 6-2: Feature elimination in cichlid fishes. (A) Investigating clustering of metrics for metric elimination. The grayed out region shows metrics which all clustered together. This gray cluster consists of metrics related to the standard deviations in measured values. Data were normalized by each metric such that the max value along each metric was 1 (red), while the minimum value was 0 (blue). (B) Identification of informative features by examination of metric distributions. Red lines represent the means, while blue boxes represent the first quartile, blue dashed lines represent the second quartile, and green squares represent outliers.

Initial hierarchical clustering analysis of the resulting metrics revealed segregation of the data set into three main clusters, but the majority of the data set fell into a single large

cluster that was difficult to interpret (Figure 6-3 A). This mirrored results with other tissues, mainly the ESC aggregates discussed in Chapter 5 (

Figure 5-3). Thus, in order to analyze the clusters further, a PCA model was created which explained the majority (83.8%) of the variance: 51.3% from PC-1, 22.9% from PC-2, and 9.6% from PC3 (Figure 6-3 C,D). PC-1 correlated highly with metrics associated with the shape of pSmad+ and pSmad+/dlx3b+ clusters, while PC-2 was strongly inversely correlated with dlx3b+ cluster metrics, and PC-3 correlated with the eccentricities metrics.

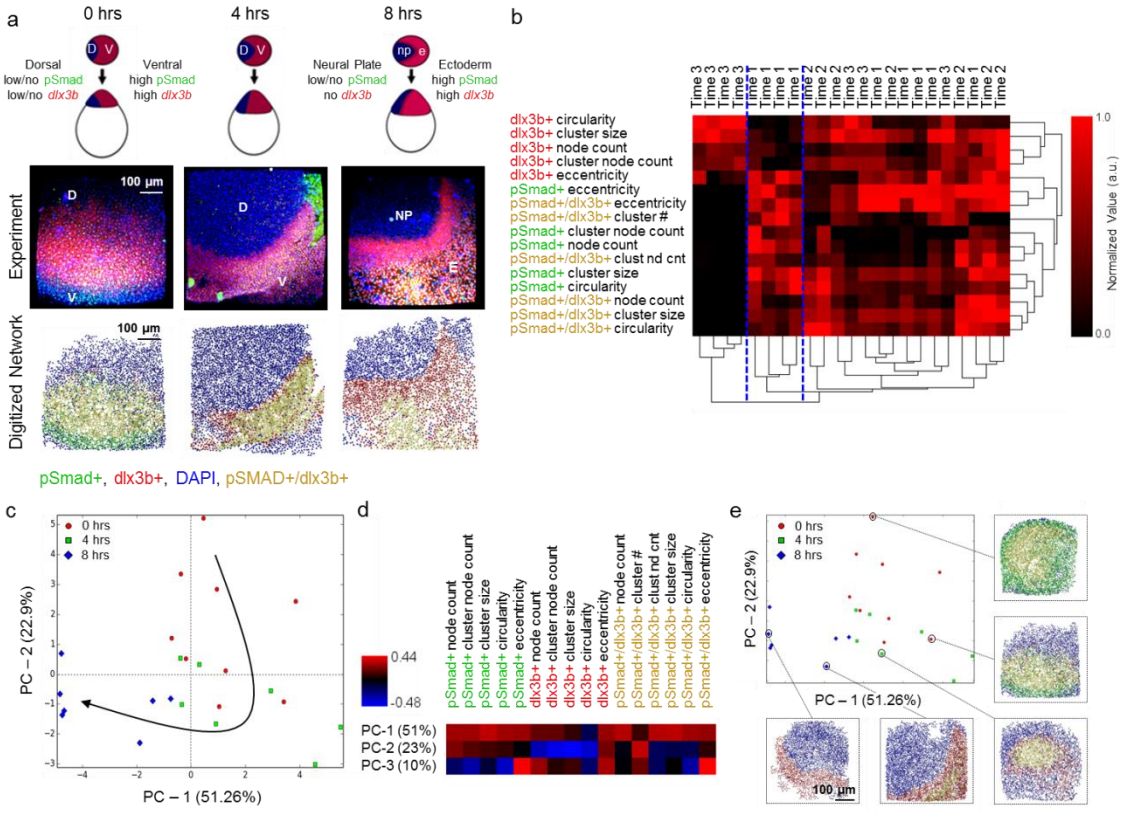

Figure 6-3: Pattern analysis during gastrulation of cichlid fish. (A) Representative schematic of gastrulation in cichlids on top, experimental confocal data in middle (red – dlx3b, green – pSmad, blue DAPI, yellow pSmad+/dlx3b+) followed by network reconstructions on the bottom (red – dlx3b, green – pSmad, blue DAPI, yellow pSmad+/dlx3b). Scale bars are 100 µm. (B) Hierarchical clustering of the data set yields three distinct sub-clusters (highlighted by dividing blue dashed lines). (C) PCA of the

resulting data set reveals (n >= 7 per time point) a distinct trajectory informed by time (0 hours – red circle, 4 hours – green square, 8 hours -blue diamond). (D) Heatmap of how each metric contributes to the principal component model. Blue indicates a strong negative contribution, while red indicates a strong positive contribution. (E) Annotated PCA plot showing the evolution of pattern formation over time during gastrulation in cichlid fish. All images have the same scale bar of 100 μm.

This final PCA model not only revealed the initial and terminal states detected by hierarchical clustering, but more importantly, resolved the remaining data along a clear trajectory (Figure 6-3 C). Selecting various points along the trajectory revealed a set of patterns that matched the known biology, while also identifying subtle transition states between discrete time points (Figure 6-3 E). Early development time points (0 hours – 4 hours) were characterized by a shrinking pSmad+ region with an increase in dlx3b+/pSmad+ regions, as indicated by the shift primarily in early time points along PC-2. The midpoint of gastrulation (~4 hours) exhibited an important switch in the formation and shape of the dlx3b+ region, and the final developmental stage (4 hours – 8 hours) was heavily influenced by the emergence of a crescent of dlx3b expression, as indicated by its progression along the PC-1 axis.

To test how well these metrics predicted the evolution, a set of path finding simulations intended to find the most likely flow of information through a given process were performed (Figure 6-4 A).The average trajectory for these simulations showed the 0 hours samples peaking first, followed by a peak in the 4 hour, and 8 hour samples (Figure 6-4 B). Analysis of the clusters indicated that early samples were marked by a high level of pSmad+ and pSmad+/dlx3b+ regions, followed by a gradual increase in the presence of solely dlx3b+ clusters (Figure 6-4 C). Taken together, these results indicate that the biological trajectory produced by this approach can distinguish the precise state of gastrulation of a biological sample regardless of when it was acquired during the process

(Figure 6-3 C,E, Figure 6-4 B), further demonstrating the unique strength of a quantitative network-based pattern classification approach for analyzing morphogenic processes.

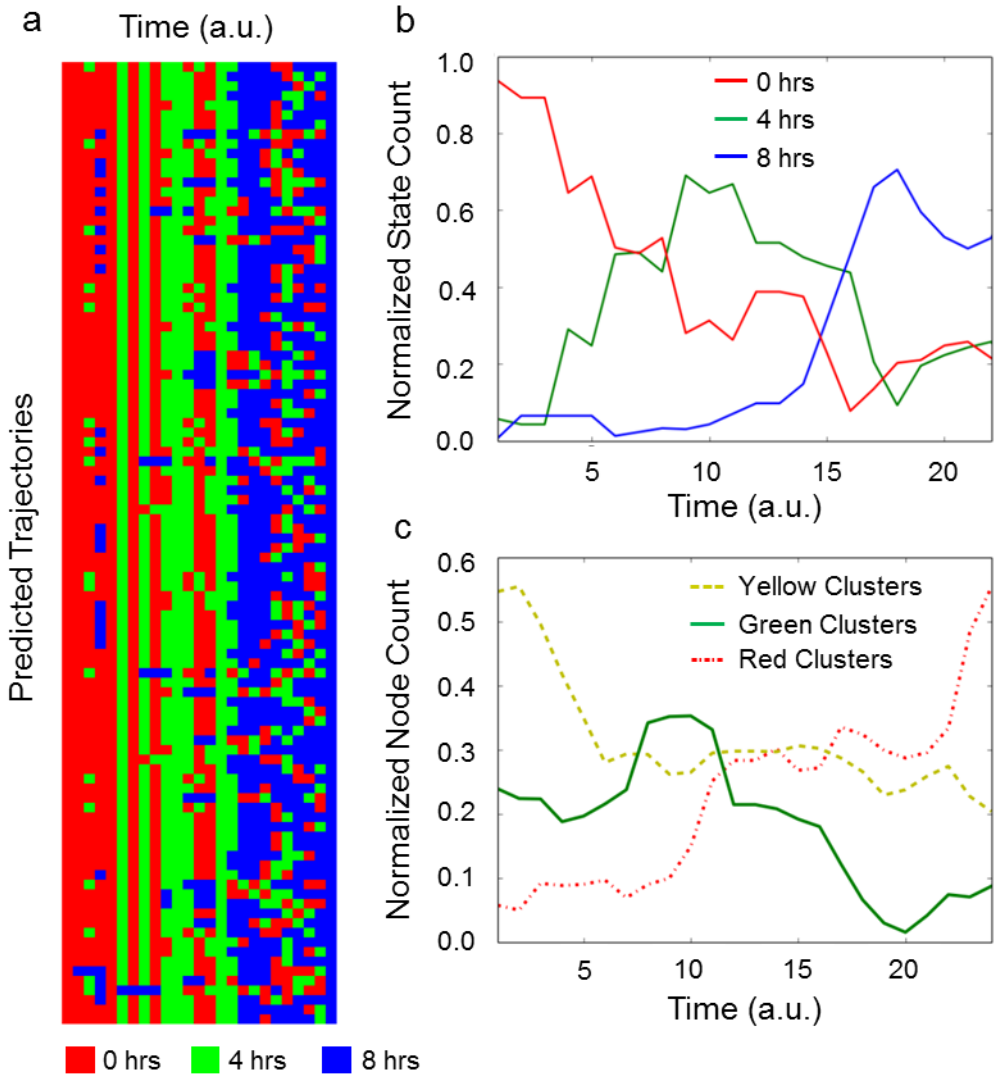

Figure 6-4: Quantifying cichlid differentiation via path-finding methods. (A) Predicted pattern trajectories through time show temporal ordering of early (0 hours, red), middle (4 hours green) and late gastrulation (8 hours blue). (B) Averages over all trajectories show state evolution from early, to middle to late gastrulation. (C) Average over all trajectories showing the evolution of different cluster number during gastrulation.

6.2.4 Discussion

Gastrulation represents a complex stage during development in which the blastula undergoes a dramatic transformation in which the three germ layers, endoderm, ectoderm, and mesoderm are formed. The ectoderm tissue eventually gives rise to the neural plate which forms the neural tube and neural crest cells in a process called neurulation. These cells go on to form all of the tissue expressed in the nervous system, which underscores the importance of understanding the process of gastrulation and subsequent neurulation. As mentioned previously, prior work by the Streelman lab has shown that gene expression events occurring as early as gastrulation can be important for specifying downstream neural development. Thus, pipelines for extracting quantitative single cell spatial information from images would be crucial for understanding these regulatory processes.

This work provides a pipeline for analysis of spatial protein and gene expression in cichlid fish embryos. Though it did not directly address the question of differences in *dlx3b* spatial expression patterns between cichlid species, this study showed the network based metric system accurately captured *dlx3b* and pSmad dynamics during gastrulation. This work highlighted the creation of a principal component (PCA) model which utilizes the network derived metrics to segregate embryos based on stage of gastrulation. The model highlights the evolution of pSmad+/*dlx3b*+ regions as key measurements for distinguishing early stage embryos, while the evolution of the subsequent *dlx3b*+ crescent is important for distinguishing later stage embryos.

The spatial expression levels of *dlx3b* were quite binary, the gene was either on or off, however, pSmad expression represented more of a challenge as it was present in a gradient. Gradient expression levels were difficult to digitize into a binary on/off state. For the context of this work, the threshold for pSmad positive clusters was set extremely low to capture a majority of the pSmad signaling present in the embryo. In the future more sensitivity could be achieved by splitting the pSmad region up into two or possible

even three regions with different expression levels. The relative size and spatial location of the different regions would provide further resolution into the dynamics of the BMP4 induced pSmad gradient.

Taken together this work provided a proof of concept approach demonstrating that network analysis was powerful enough to classify early expression patterns during gastrulation in cichlid fish. The networks analyzed here were on the scale of tens of thousands of nodes, with tens to hundreds of cellular connections. Furthermore, the patterns analyzed in this study were typically a single cluster of dlx3b or pSmad expression which challenged the sensitivity of the network based method to detect subtle differences from single large clusters of cells. In future collaboration with the Streelman lab this network based method will be used to classify embryos from different cichlid species to determine if there are differences in dlx3b expression during cichlid fish gastrulation.

6.3 Neural Differentiation and the Glial Switch: Turning on the Light

6.3.1 Background

The glial switch remains an enigma in neuronal developmental biology. After the formation of the neural tube and crest cells (outlined in section 6.2.1), the ventral portion of the tube undergoes specification into neural progenitor domain under the control of a Sonic Hedgehog (Shh) gradient which is secreted from the notochord [312, 313], a group of cells formed during the gastrulation process. The notochord is responsible for the formation of the neural tube and crest during development. During development the Shh signaling gradient led to the formation of a progenitor pool (marked by Olig2) which created motor neurons at early stages (marked via HB9) and oligodendrocyte precursor cells (marked by Olig2[314] and Nkx2.2[315]) at later stages.

In the context of embryonic stem cell aggregates, motor neurons and glial cells proceed through a progenitor cell phenotype (marked via Olig2)[316] making this an appropriate *in vitro* system in which to study the glial switch. To elucidate possible molecular mechanisms governing the spatio-temporal patterns associated with the glial switch *in vitro*, network based analysis in combination with simulations were used. Different signaling paradigms like soluble signaling through TGFβ and local cell-cell interaction via Notch [317, 318] have been implicated in regulating progenitor cell populations. In this work a previously developed computational model was used (Chapter 3) to ask questions about the factors regulating transitions between cell states. By comparing images to computational simulations via a network analysis approach, this study highlights several novel findings about biological regulation of the glial switch. First, the classical linear pathway in which a cell progressed through a nascent dividing progenitor state before they differentiated could not explain the cellular dynamics observed in this system. Second, the model suggested a soluble negative feedback mechanism existed which governed the differentiation of motor neurons from progenitors. Taken together these results highlight the power of network based analysis in combination with simulation to elucidate mechanisms governing complex spatial evolution of biological processes.

6.3.2 Methods

6.3.2.1 Network Reconstruction for 3D Networks

First, confocal images were loaded into the commercial software package IMARIS. A blob finding routine for identifying nuclei or spots was run. This routine involved utilizing a Gaussian smoothing step with a filter size of 6.5 um, following by a subsequent local maxima detection step to identify seeds for nuclei. These seeds were then expanded via a region growing algorithm, and resulting clumps of nuclei were segmented using an intensity watershed algorithm based approach. For the purposes of

these studies this yielded sufficiently annotated nuclei, however, more sophisticated cellular segmentation techniques exist, particularly those using Gaussian Mixture Models (GMMs) to approximate nuclei. Then all of the properties of each spot were calculated (size, shape, and intensity information) and exported to a file which was then processed using a custom python script. In this case, IMARIS does not export neighbor connectivity information, so this information was inferred by reconstructing via a KDTree implementation (available under scipy.spatial.KDTree) with a distance cutoff of 22.5 um. This distance was chosen as it represented the one and a half times the average diameter of a cell, and it was inferred that this distance would capture most cellular interactions. Each node was then annotated according to the information form the respective channels, blue (channel 1) for Nuclei, green for HB9 (channel 2), red for Olig2 (channel 3), and cyan for Nkx2.2 (channel 4). For subsequent cluster analysis, cells were only considered to fall into the glial class if they stained positive for both Oli2 and Nkx2.2, while progenitors were identified as cells which only stained positive for Olig2. Though motor neuron progenitors (Olig2+/HB9+) could be identified via this analysis, they were not analyzed in the following models. Cells staining positive for Nkx2.2 were considered artifact and also not included in the analysis.

6.3.2.2 Cell Culture and Aggregate Formation

Mouse embryonic stem cells (mESCs) were thawed at 10^6 cells per 10 cm dish in 10 mls media. Media was composed of a 1:1 mix of Advanced DMEM/F12 and AB2 (ArunA Biomedical) with Knockout Serum Replacement (Life Tech, 10%), L-Glut (life tech, 2 mM), PenStrep (life tech, 1x), beta mercaptoethanol (100 um). Aggregates of cells were formed via rotary culture as described previously. Briefly after cell seeding ESC aggregates are formed via rotary culture at 50 RPM in an incubator at standard culture conditions (37 C and 5% CO2.) Aggregates were subsequently maintained in this

environment with exchanges of medium on days 1, 2, 3 and 6. To differentiate down the neural lineage retinoic acid and purmorphamine (Smoothened Hedgehog (SHH) agonist) were added to the medium at a concentration of 1 uM on days 2,3 and 6. To induce maturation of neural cells types brain derived neurotrophic factor (BDNF) and glial-cell line derived neurotrophic factor (GDNF) were added to the medium on day 6 at a 10 ng/ul concentration. The notch response inhibitor DAPT was added at day 3, 5 or 7 (and everyday thereafter) at a concentration of 1 uM which established 4 treatment groups, no DAPT treatment, DAPT at day 3, DAPT at day 5, and DAPT at day 7.

6.3.2.3 Immunostaining and Confocal Microscopy

ESC aggregates were collected for staining and fixed in 10% formalin for 45 minutes. EBs were permeabilized for 30 minutes in 1.0% TritonX-100, re-fixed in formalin for 15 minutes, and blocked in blocking buffer (2% bovine serum albumin, 0.1% Tween-20 in PBS) for 3 hours. Olig2 (1:500, Millipore) and Nkx2.2 (1:5, DHSB) antibodies were applied overnight on day one. Secondary antibodies (Alexa Fluor 546 for Olig2 and 647 for Nkx2.2, Life Tech) added for 4 hours, followed with a subsequent Anti-GFP antibody (conjugated to Alexa488) overnight staining step. After staining the aggregates were counterstained with Hoescht (1:100) for 25 minutes. At least 10 aggregates for each time point and treatment condition were imaged on a Zeiss LSM 710 Confocal microscope using Zen software. The 40X oil objective was used to obtain sufficient spatial resolution with a 2-μm z-resolution for 80-100 μm stacks total.

6.3.2.4 Agent Based Computational Modeling

ESCs were modeled using a simple agent based modeling approach. Individual cells were allowed to proliferate, and change state. The state changes were defined via sets of probabilistic equations. Four different rule sets were considered for this analysis: linear model random, linear model feedback, fast model random and fast model feedback. The equations along with parameter values used are shown for each scenario below. In all

cases *num_mn* refers to the number of motor neurons present in the simulation, *mn* denotes motor neuron, *np* denotes neural progenitor and *g* denotes glial.

6.3.2.4.1.1 Linear Model – Random Differentiation

$$P_{np} = \alpha \tag{6-1}$$

$$P_{np \rightarrow mn} = \beta \tag{6-2}$$

$$P_{np \rightarrow g} = \gamma \tag{6-3}$$

Where α = .0005, β = .1 and γ = .001.

6.3.2.4.1.2 Linear Model – Motor Neuron Feedback

$$P_{np} = \alpha \tag{6-4}$$

$$P_{np \rightarrow mn} = \left(\frac{\beta}{\left(1 + \left(\frac{num\_mn}{k1}\right)^{n}\right)} \right) \tag{6-5}$$

$$P_{np \rightarrow g} = \gamma \tag{6-6}$$

Where α = .0005, β = .1, γ = .001 and k1 = 200.

6.3.2.4.1.3 Fast Model – Random Differentiation

$$P_{np} = \alpha \tag{6-7}$$

$$P_{mn} = \beta \tag{6-8}$$

$$P_{np \rightarrow mn} = \gamma \tag{6-9}$$

$$P_{np \rightarrow g} = \delta \tag{6-10}$$

Where α = .0001, β = .001, γ = .01, δ = .001.

6.3.2.4.1.4 Fast Model – Motor Neuron Feedback

$$P_{np} = \alpha \tag{6-11}$$

$$P_{mn} = \left(\frac{\beta}{\left(1 + \left(\frac{num\_mn}{k1}\right)^n\right)} \right) \tag{6-12}$$

$$P_{np \rightarrow mn} = \left(\frac{\gamma}{\left(1 + \left(\frac{num\_mn}{k2}\right)^n\right)} \right) \tag{6-13}$$

$$P_{np \rightarrow g} = \delta \tag{6-14}$$

Where α = .000099, β = .001, γ = .01, δ = .01, k1 = 50, and k2 = 200

6.3.2.5 3D Aggregate Computational Modeling

Aggregate based ABM modeling was carried out in a similar manner to what was described in section 5.2.2. The only difference was the inclusion of either a gradient of soluble signaling factor X responsible for inhibiting differentiation, or a term regulating differentiation via direct cell-cell contact. Diffusion of factor X was solved as previously described in section 5.2.2.

$$P(x) = 1 - \frac{norm\_mn^n}{(k1 + norm\_mn)^n} \tag{6-15}$$

$$P(x) = 1 - \frac{X(x,y,z)^n}{(k1 + X(x,y,z))^n} \tag{6-16}$$

Where *X(x,y,z)* represents the soluble factor X concentration at the cellular location (denoted as x, y and z), *n* represents the hill coefficient, and *norm_mn* represents the number of motor neurons immediately ascent to a cell normalized by the total number of

cellular connections. These simulations were carried out with parameters shown in Table 6-1. In all cases simulations were carried out for a 6 day time period *144 hours) or until an EB size of 60,000 was reached.

Table 6-1: Parameter values for 3D ESC aggregate neural simulations

Soluble Signaling		Cell-Cell Signaling	
Parameter	Value	Parameter	Value
a	0.001, 0.005	*a*	0.001, 0.005, 0.01
k1	0.01, 0.025, 0.04	*k1*	0.05, 0.1, 0.2
k2	0.04, 0.055, 0.07, 0.09	*k2*	0.2, 0.3, 0.4
n	30	*n*	29
c	0.001, 0.01	*c*	0.001, 0.01

Where *a* represents a basal differentiation rate for emergence of neural progenitors, *k1* and *k2* govern the emergence of motor neurons either from the stem cell or progenitor pool, *n* is the hill coefficient in both cases, and *c* represents the glial differentiation probability from neural progenitors. As mentioned above the only difference between soluble and cell-cell interactions was the form of the equation used for state transition probabilities (equations 6-15 and 6-16 respectively).

6.3.3 Results

All of this work was carried out in conjunction with the Stice lab at UGA, where Raymond Swetenburg carried out the experiments for differentiation ESC aggregates towards their respective neural lineages. The first step in this process was to ascertain the spatio-temporal dynamics associated with motor neuron and glial cell formation using the *in vitro* ESC aggregate system. To do this network analysis was performed on 3D confocal stacks taken from days 3-9 of differentiation. The DAPT molecule (an inhibitor of notch response, specifically γ-secretase) was applied at days 3, 5 and 7 leading to a total of 4 different groups for analysis (Figure 6-5 A). Confocal stacks were digitized

with aid of the commercial software package IMARIS to generate robust network reconstruction and digitization of the images (Figure 6-5 B). The reconstructed image showed that in the basal treatment (Figure 6-5C Tx (Day 0)) there is an initial burst of motor neurons, followed by a decline in motor neuron (HB9+) production, concomitant with an increase in progenitor (Olig2+) cells, and glial populations (Nkx2.2+/Olig2+) (Figure 6-5 D – top row). Interestingly, when DAPT treatment is administered early (Day 3 or 5) the cultures regressed, and not progenitors were observed at later time points. However, when DAPT was administered at day 7 induction of a robust glial response was observed, indicating that notch may regulate the glial switch (Figure 6-5 D– bottom row).

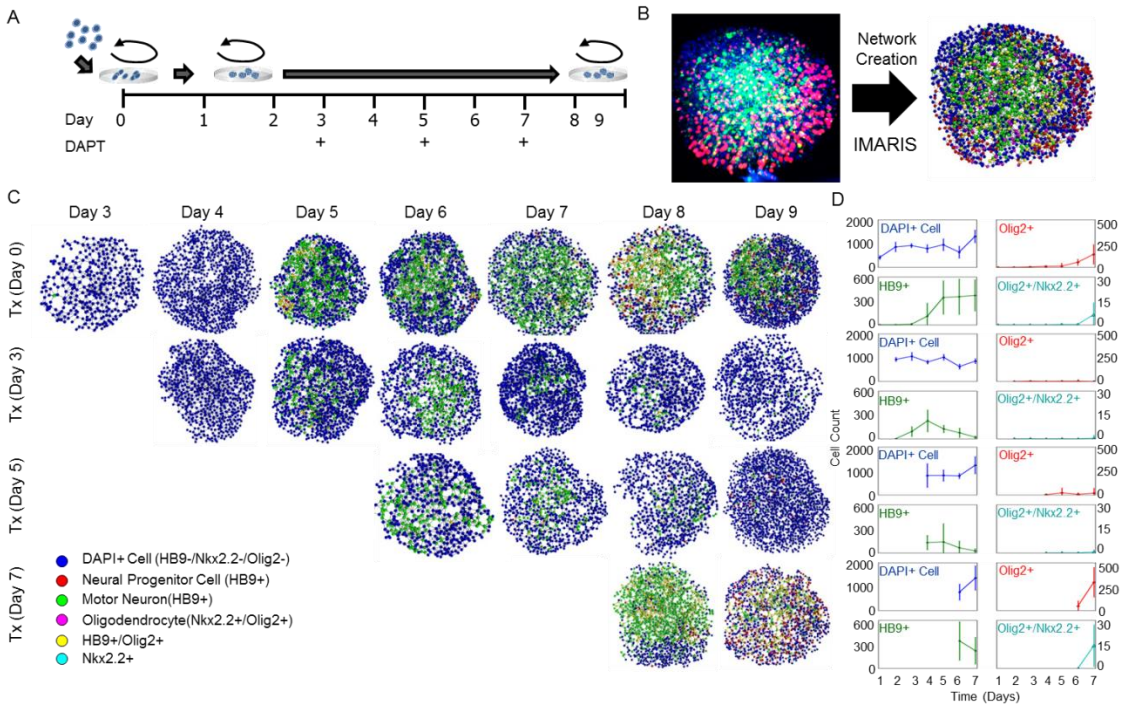

Figure 6-5: Analyzing neural differentiation kinetics in ESC aggregates. (A) The experimental outline for EB formation culture and treatment showing DAPT administration on Day3, 5 and 7. (B) The digitization of ESC aggregate images into networks via IMARIS. (C) Representative reconstructed networks showing and their dynamic changes with time (D) for each treatment. For each time point and treatment 10 samples were analyzed. In (D) averages are shown +/- one standard deviation of the mean.

To investigate what rules were implicated in governing the temporal dynamics of neural differentiation in this ESC aggregate system, a simple stochastic cell model was implemented. Originally two rule sets were investigated (random and negative feedback Figure 6-6 A,B), however neither of these models could explain the plateau in motor neuron production. Several parameter sets utilizing the 3D agent based aggregate model were also investigated and were unable to recapitulate the experimental dynamics (data not shown). Thus, subsequent models were investigated in which motor neuron differentiation could proceed rapidly through the Olig2+ progenitor state which never allowed the progenitor population to become established. In this case even the random model was able to capture the dynamics of motor neuron production (Figure 6-6 C), and subsequent introduction of negative feedback via motor neurons inhibiting their own production lead to a near exceptional fit of the experimental data (Figure 6-6 D).

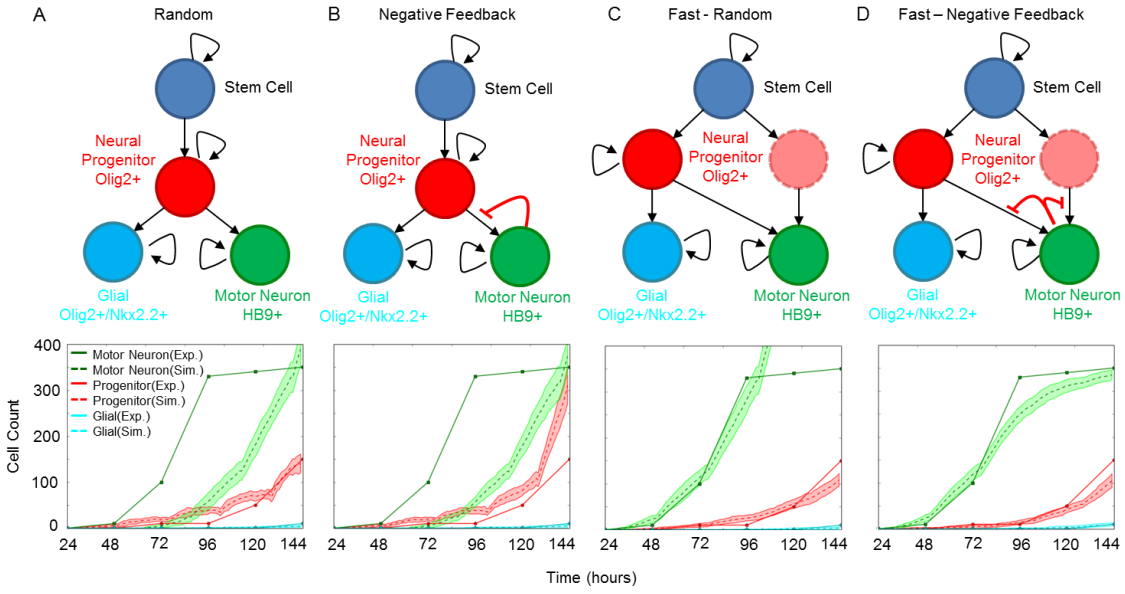

Figure 6-6: Computational models of differentiation kinetics in ESC aggregates. (A) The random rule for a linear cascade and (B) a feedback rule for motor neuron are unable to capture experimental dynamics for motor neurons (green), progenitors (red) and glial cells (cyan). (C) Representative reconstructed networks showing and their dynamic changes with time (D) for each treatment. For each time point and treatment 10 samples were analyzed. In all cases experimental data is represented by the dashed lines, while

simulations the mean is graphed a solid line with a shaded region corresponding to +/- one standard deviation. In all cases 20 simulations were performed.

To analyze the nature of the signal provided by motor neurons, spatial network analysis was employed. Here network based analysis was used to determine if the regulation by motor neurons was a soluble or local cell-cell interaction based effect. To do this the spatial expression signatures of Olig2 + (progenitors), Olig2+ /Nkx2.2+ (glial progenitors), HB9 + (motor neurons) and unmarked cells were quantified. Principal component analysis (PCA) of the variable space revealed a definitive set of axis capable of discriminating motor neuron and glial differentiation (Figure 6-7). Robust motor neuron differentiation in cultures is indicated by positive translocation along principal component (PC) 1 (33.17%), while glial differentiation is roughly indicated by translocation with principal component axis 2. Together these two axes only explained ~ 50% of the variance in this system; interestingly 8 principal components were needed to describe 85% of the variance in this system, suggesting that other component axes may capture important differences in the data.

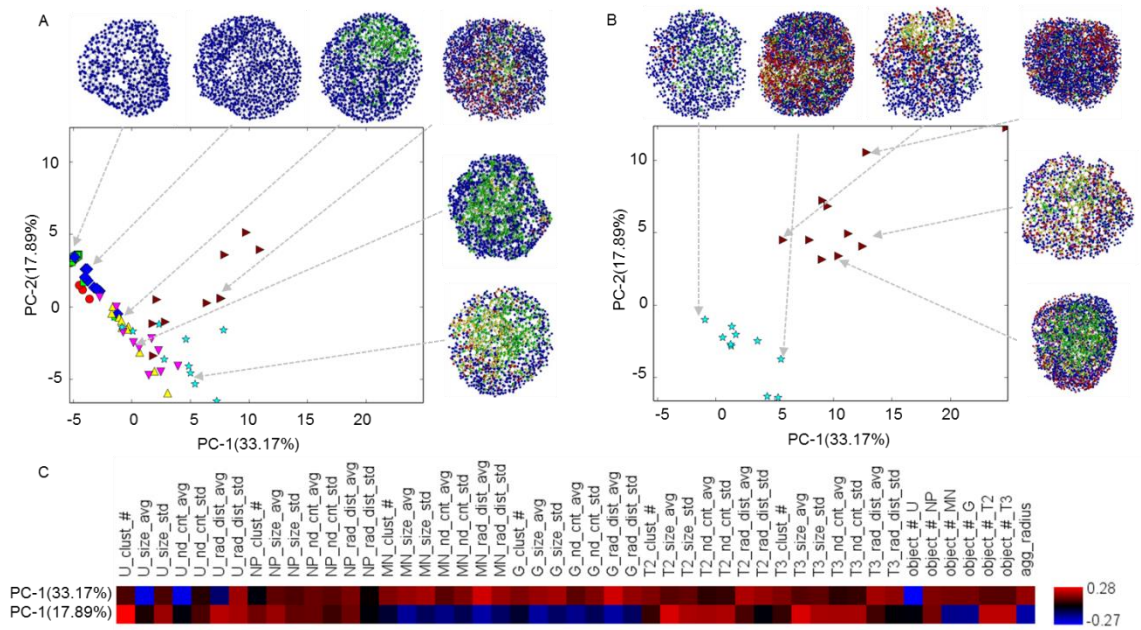

Figure 6-7: Analyzing neural differentiation kinetics in ESC aggregates. (A) PCA of the no DPAT control group, surrounded by representative images. (B) PCA of treatment with DAPT on day 7. (C) Heat map showing the contribution of each metric to principal components 1 and 2 where red indicates a positive correlation, and blue indicates a negative correlation. For each time point 10 samples were compared.

Glial cells (Nkx2.2/Olig2+) and neural progenitors (Olig2+) were rare at early time-points, and only around day 7 of differentiation as rapid induction of these cell types observed. This rapid switch-like behavior seemed to implicate a soluble factor as the motor neuron signal, however to quantitatively assess this computation modeling simulations were performed. Two simulation rule sets were simulated, one in which motor neuron feedback was implemented as a soluble signal, and one in which it was implemented as a local cell interaction based signal.

Both parameter sets where produced trajectories which were able to match the experimentally observed data (Figure 6-8 A,B). The local rule produced a motor neuron trace which was similar to the experimental data, however, it failed to produce switch-like transitions in the production of Olig2+ progenitor cells and Olig2+/Nkx2.2+ glial

cells (Figure 6-8 C,D). However, the soluble rule set was capable of producing stark rapid "switch-like" changes in the progenitor and glial cell populations (Figure 6-8 E,F) providing further evidence that the signal produced by motor neurons is likely transduced via a soluble factor. It is important to note that the top simulations from each rule set produced similar scores on the current objective function, suggesting that an alternative approach is needed to provide further resolution in distinguishing parameter sets which lead to switch like behaviors.

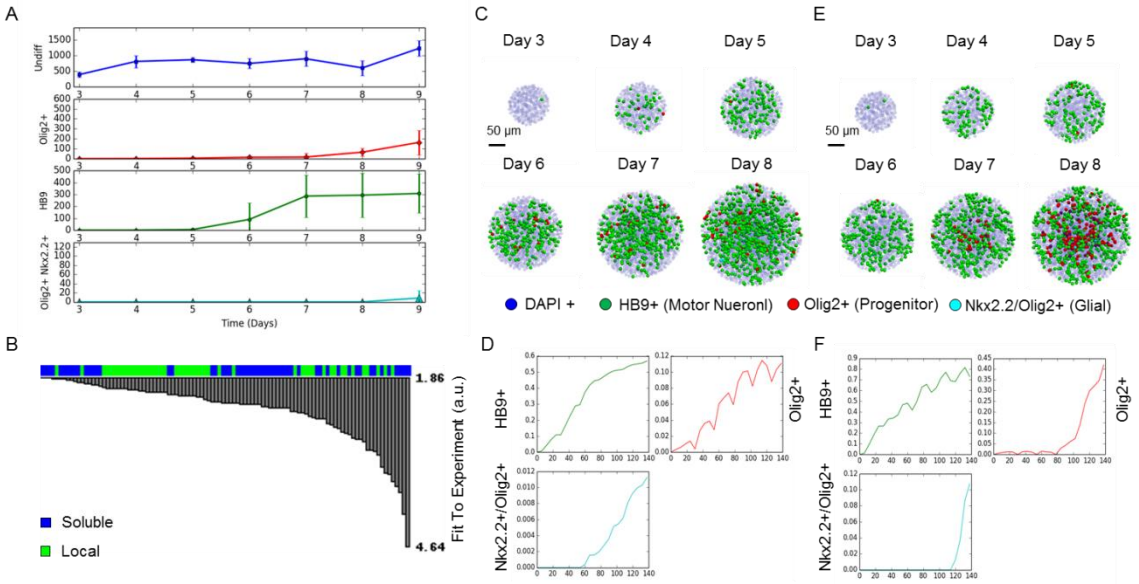

Figure 6-8: Comparing spatial dynamics of neural differentiation to computational models. (A) The experimental data traces for DAPI + (blue), Olig2+ (red), HB9+ (green), and Olig2+Nkx2.2+ (cyan) cells. Lines represent mean with error-bars of +/- 1 standard deviation. (B) Distance from the experimental data to the simulations where blue represents soluble rule sets, and green represents local. (C) Representative simulation images of the best local-rule parameter set and the quantification of cell types over time (D). (C) Representative simulation images of the best soluble-rule parameter set and the quantification of cell types over time (D).

6.3.4 Discussion

Utilizing network based analysis approaches this study uncovered several novel biological insights into regulation of the glial switch. First, the differentiation schema of the *in vitro* system was not strictly linear. Motor neurons appeared spontaneously in culture without observation of established progenitor populations. The model predicted the only way this could arise was if the motor neurons were able to rapidly move through the progenitor phase during early time points in culture. This drives the establishment of a population of motor neurons in the *in vitro* system which was observed seemingly before neural progenitors. The second important insight was that motor neurons must inhibit their differentiation in some manner. Computational models predicted that only when motor neurons negatively fed-back on their own production would the progenitor population be able to proliferate, which ultimately made the environment permissive for glial cell differentiation. It is important to note that the amount of glial cells arising in culture was consistent with a normal stochastic process that did not require any feedback to drive its production. Finally, spatial pattern analysis revealed that the signal secreted by motor neurons was likely a soluble factor. This is not without precedent in the literature where terminal neurons have been shown to feedback on progenitor cell populations and inhibit their proliferation [319, 320], but would be the first report of motor neurons inhibiting their own production in such a manner.

Another interesting aspect of this work is the system behavior upon treatment with DAPT. DAPT is a notch inhibitor, and it is thought that progenitor cell populations are established by local notch interactions between progenitor cell types which inhibit differentiation and promote self-renewal. This study indicated that upon DAPT treatment progenitor populations underwent subsequent differentiation towards glial cells at later time points. This was interesting because it agreed with the modeling hypothesis that the motor neurons secreted a factor which made the environment permissive for glial differentiation. Furthermore, by inhibiting self-renewal of the progenitors the resulting

progenitor population was primed for differentiation to glial cell types. This leads to a variety of future studies which Raymond is currently performing in the Stice lab. The first study will address if motor neurons really do secrete a factor which inhibits progenitor differentiation via a conditioned medium study. In this study conditioned medium from motor neurons will be placed on neural progenitors in order to assess subsequent differentiation of motor neurons for the neural progenitor population. Additional studies will look at the ability of progenitors to bypass mitosis. The central idea for these studies is that progenitors at low density will not possess enough Notch signaling via cell-cell contact to remain in a progenitor state. To test this hypothesis cell division rates at various cell destines will be monitored in real time via an Olig2-GFP reporter line in the Stice lab. Taken together these future experiments will provide enhanced information about factors influencing the glial switch *in vitro*. In summation, this study highlights the power of spatial network comparisons between experimental and computational systems to elucidate non-intuitive novel biological relevant regulatory mechanisms.

6.4 Analysis of sources of intra and inter-aggregate variation via histology

6.4.1 Background

Pluripotent stem cells have the ability to differentiate into all three germ layers, and when cultured as aggregates, can recapitulate various aspects of development [223-227]. These aggregate cultures produce cell and tissue types with markedly different morphology, even after a week in culture [263, 321]. Though population assays have been successful in capturing some phenotypic differences, questions about intra- and inter-aggregate variation still remain. While technologies such as light sheet and confocal microscopy provide a method to assess 2D/3D spatial heterogeneity, they require the use of reporter

lines or destructive staining analysis, are inherently low throughput, and require access to complex equipment [263, 268]. One alternative which has been around for decades is the use of histological staining for assaying complex tissues. Histology is inherently lower cost, and in many cases does not require assessment of specific markers to identify interesting morphological structures in tissues, which makes this technique popular in assessing various pathologies [322-326].

Histological staining preserves native local tissue conformation and spatial information; however a major limitation of histological analyses is that visual differences are extremely hard to measure quantitatively. Quantitative histology typically requires an expert to identify regions of tissue in order to extract semi-quantitative information that is inherently prone to user error and not high throughput. Previous efforts to identify cells with similar properties within histological sections using machine learning or pattern recognition approaches have attained limited success because these techniques often rely on specific markers or solely on cell shape information[203, 322-326]. This makes these methods inherently susceptible to shape-based bias from automated cellular detection algorithms [203, 322, 327]. This study describes the creation of a novel histological classification pipeline designed to improve single cell classification while also harnessing the power of networks to classify more complex cellular structures with simple rules.

6.4.2 Methods

6.4.2.1 ESC Aggregate Formation and Culture

To initiate ESC differentiation, a single cell solution was obtained via dissociation in 0.05% trypsin / 0.53 mM EDTA solution. ESC aggregates were formed via forced aggregation of single cells into 400 μm diameter PDMS microwells (AggreWell), with

approximately 1000 cells per well. After 20 hours of formation, EBs were removed from the microwells and maintained in suspension on a rotary orbital shaker at 65 rpm, with approximately 1500 EBs per plate[250]. Aggregates were fed via gravity-induced sedimentation and exchange of 90% of the media every other day throughout the remainder of the 14 day differentiation period. Aggregates were formed in the standard growth media, without LIF, and were subsequently cultured in N2B27 serum-free media once transferred to the rotary. The differentiation toward mesoderm lineages was accomplished via supplementation of basal N2B27 media with BMP4 (10 ng/mL)[321].

6.4.2.2 Cell Culture

A murine embryonic stem cell line (D3) was used. The undifferentiated cells were cultured on 100 mm tissue culture plates coated with 0.67% gelatin in Dulbecco's modified Eagle's medium (DMEM) supplemented with 15% fetal bovine serum(FBS) (Hyclone, Logan, UT), 2mM L-glutamine (Mediatech), 100 U/ml penicillin, 100 ug/ml streptomyocin, and 0.25 ug/ml amphotericin (Mediatech), 1x MEM nonessential amino acis solcuiotn (Mediatech), 0.1 mM 2-mercaptoethanol (FisherChecmical, Fairlawn, NJ), and 10^3 U/ml leukemia inhibitory factor (LIF) (Chemicon Internation, Temecula, CA). Cells were passaged every 2-3 days prior to reaching 70% confluence.

6.4.2.3 Histology

For histological analysis, aggregates were collected via gravity-induced sedimentation, washed with PBS, and fixed in 10% formalin for 45 minutes under rotation. Fixed samples were subsequently encapsulated within Histogel and processed via a series of graded ethanol and xylene rinses and embedded in paraffin. Paraffin-embedded samples were sectioned using a rotary microtome (Microtom HM310) to produce 5 µm sections,

which were mounted on glass slides. Histological analysis was subsequently conducted via deparaffinization in graded xylene and ethanol, followed by staining with hematoxylin and eosin (H&E). Stained sections were imaged using a Nikon Eclipse 80i microscope with a SpotFlex digital camera.

6.4.2.4 PCA

Principal component analysis was performed using *sklearn* for the python programming language and used primarily as a dimensional reduction technique for data visualization. The python package *Matplotlib* was used to plot all PCA plots, while the heatmaps displaying component information where created with custom written code using a combination of the python packages *numpy* and the *PIL*. All data points were mean centered and unit variance scaled as required by the PCA algorithm using the scale function from *sklearn*. The PCA algorithm relied on singular value decomposition, which can lead to multiple fitted estimators displaying the data with principal components flipped. When automatic dimension fitting was required, the method of Thomas P. Minkas was used [259].

6.4.2.5 Network Reconstruction via Cell Profiler

Cell Profiler (http://www.cellprofiler.org/)[278] was used to analyze all of the 2D samples. For histological data, a spectral unmixing algorithm was applied using Cell Profiler to extract the nuclear staining and separate the cell cytoplasm stained with eosin, from the nuclei stained with hematoxylin. A global MCT algorithm was used to threshold each of the samples and nuclei were detected via an intensity based method in the "identify primary objects" module in Cell Profiler. The detected nuclei were used as seeds for cell detection using the thresholded eosin images and the "identify secondary

objects" module in Cell Profiler. Shape information (e.g. area, eccentricity, perimeter) was then analyzed using the measure object properties module. The number of nearest neighbors was computed using the "measure object neighbors" module and exported to the python script that reconstructed the networks as described above. For annotation of cell fate, a classification system was employed to use the cell and nuclei measured metrics along with network properties to classify mesenchymal cells.

6.4.2.6 Classification

Classification was performed utilizing the algorithms available in the sklearn package for python. In the case of the mesenchymal classifiers the following functions were used: linear SVC, decision tree, K nearest neighbors, and SVC with a RBF kernel. A 2 fold cross validation step was used for splitting data into test and training validations sets. The image yielded approximately 6000 cells for comparisons. All classifiers were trained using the grid search function in sklearn to optimize the resulting classifier. Classifiers were optimized for the precision score to eliminate as much false positive labeling as possible. To evaluate ground truths for training, images were annotated with masks showing which cells belonged to which type of structure. All classifiers were trained using a set of 15 images which contained various features of interest: mesenchymal regions, super-dense regions, aligned regions, lumen like structures, and rosettes.

6.4.2.7 Scoring

Generally all scoring functions were calculated as the distance of the point from a given plane. For the mesenchymal score, the following points were used to define the plane: p1 = (0, -4), p2 = (-4, -5). In the case of the complexity scores the following points were used to define the plane: p1 = (0,-2), p2 = (-2, 0). In the case of the Lumen score the

following points were used: p1 = (0,0), p2 = (1E5, -2). In the case of the Super-dense score the following points were used: p1 = (0,10), p2 = (40,-10). All scoring functions were developed using control images to calibrate the scoring hyper-planes.

6.4.3 Results

6.4.3.1 Network Based Classification Improves Mesenchymal-like Region Classification in ESC aggregates

Epithelial to mesenchymal transition or EMT is a striking example of a cellular transition which results in a change of cells at both the local and global scales. EMT like phenomena occurring in stem cells were examined during two separate treatments, one in which exogenous BMP4 was applied [250], and a second in which material in the form of microparticles were added to the system [328]. This work sought to classify the degree of mesenchymal morphogenesis present in histological images (Figure 6-9 A) utilizing a novel network metric based approach. Briefly, by converting the aggregate structure into a network, information about individual cell size and shape, as well as local connectivity information was used to describe complex phenotypes (Figure 6-9 B). This allowed the identification of complex structures by linking the power associated with single cell classification methods into a tissue based regime. Classification was attempted with several methods: decision trees, state vector machines (linear, and RBF kernels), and k-nearest neighbors. When supplied with only the metrics for cell size and shape, the algorithms were not able to correctly distinguish mesenchymal cells robustly (Figure 6-9 C). The low scoring was is in part due to errors introduced during the segmentation phase. Additionally, the classifiers tended to over classify cells as mesenchymal (Figure 6-9C) leading to a low (< 75%) precision score, which is a measure of false positives

(Figure 6-9 D). However, network properties (number of neighbors or local nuclear density) conveyed a striking correlation with the phenotypes of interest (Figure 6-10). Classifiers trained with the addition of these network properties identified mesenchymal cells with greater accuracy than those without, leading to a more accurate classification (Figure 6-9 E) and a significantly increased precision score (~ 91%). This result demonstrated that network derived metrics can improve classification results when assessing individual cells.

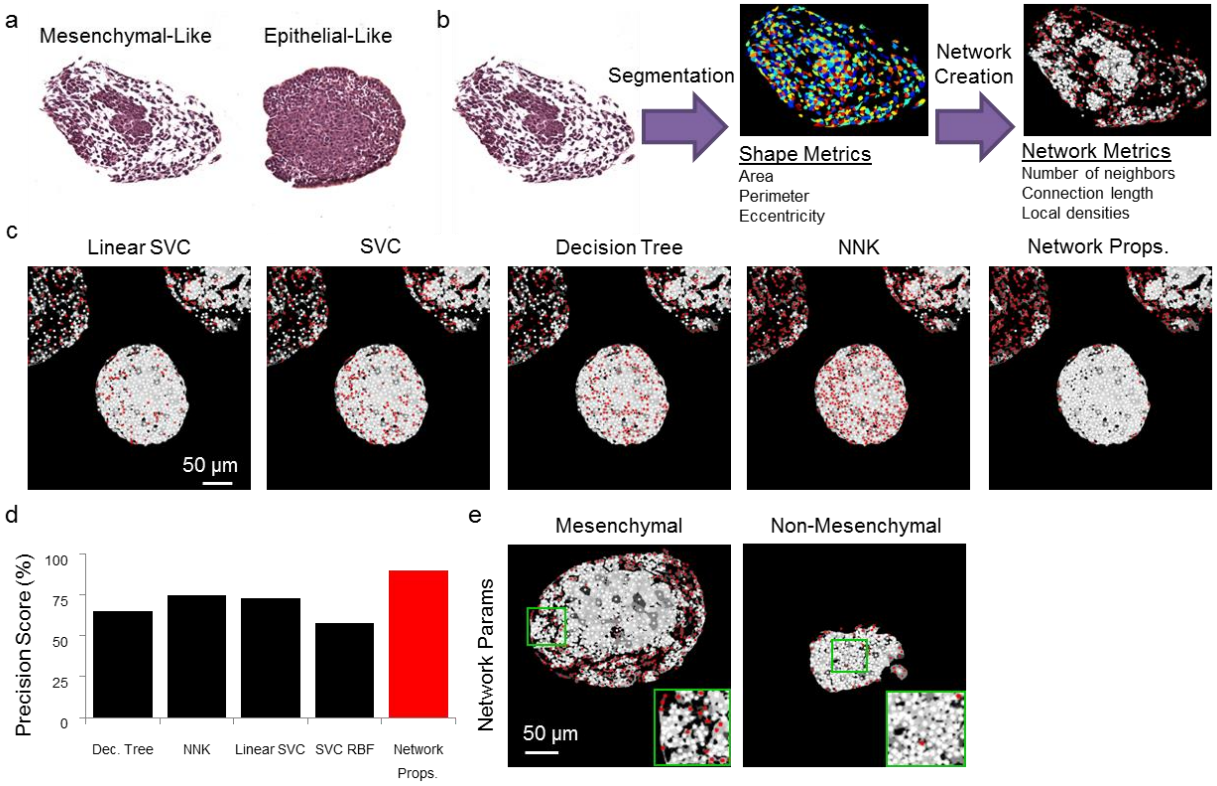

Figure 6-9: Network based metrics increase classification of mesenchymal-like regions in ESC aggregates. (A) Mesenchymal and epithelial like structures observed in ESC aggregates. (B) The pipeline for extracting quantitative per cell metrics to use in classification of cells. (C) Representative images showing the classification of mesenchymal cells (red) over epithelial like cells (white). (D) The precision scores of different classification methods, where precision is measuring false positive discrimination. (E) The network based classification algorithms applied to representative mesenchymal, and non-mesenchymal aggregates. Scale bars in all images are 50 µm.

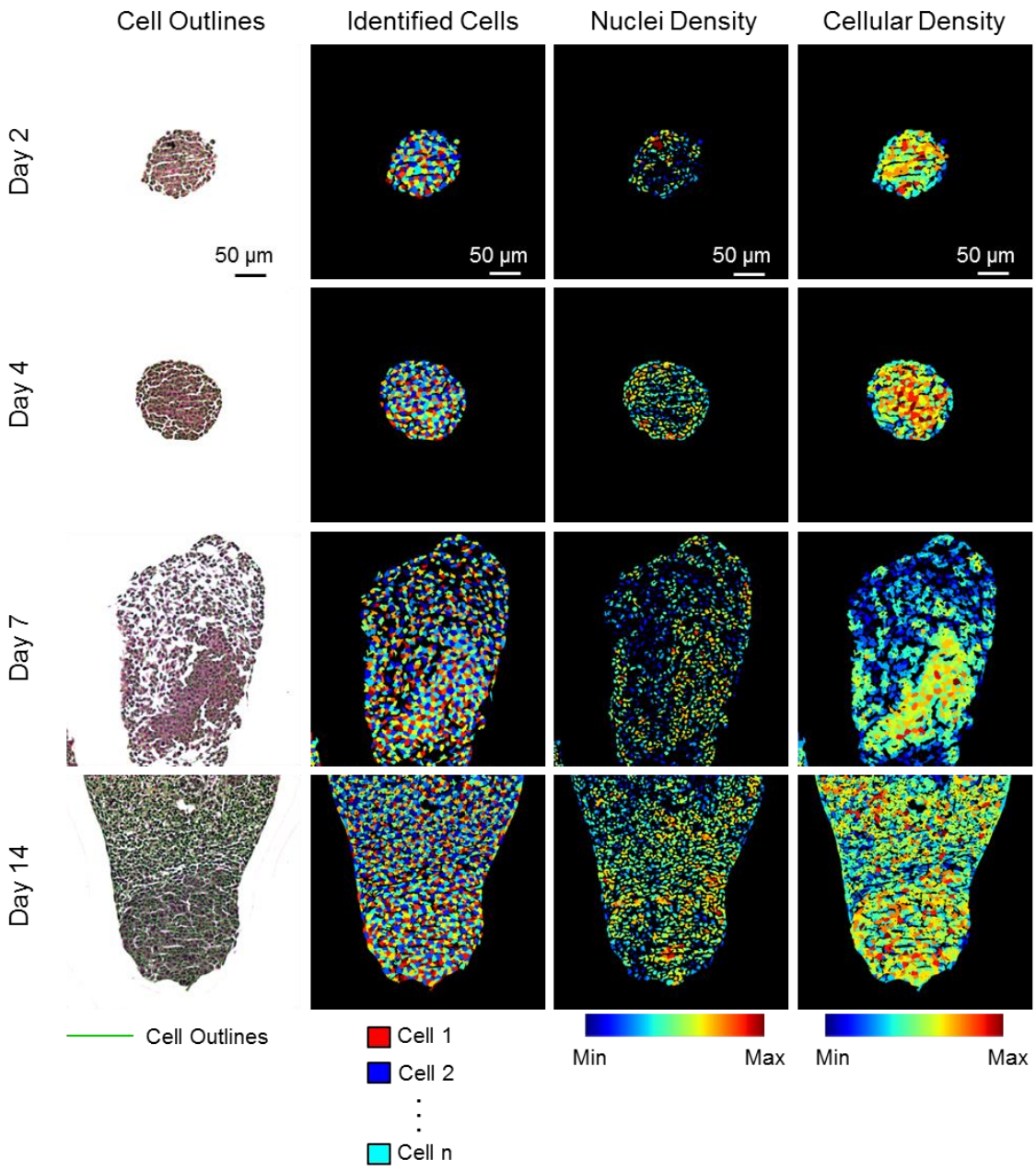

Figure 6-10: Cell profiler segmentation of histological images reveals features which correlate with mesenchymal morphologies. Representative cellular segmentation of histological images are shown for Day2, 4, 7, and 14. Identified cells are randomly color coded to provide maximal contrast between neighboring cells. The nuclei and cellular density colormaps both go form a minimum values in blue, to a maximum value in read to indicate relative packing densities, and illustrate their correlation with mesenchymal area.

6.4.3.2 Examining Spatio-Temporal Mesenchymal Morphogenesis in ESC Aggregates

ESC aggregates were stained with hematoxylin and eosin after culture in basal serum-free medium or BMP4 treatment which yielded differences in cellular phenotypes and organization between the treatment conditions (Figure 6-11A,B). As noted previously, with BMP4 treatment, the cells adopted a mesenchymal morphology compared to the compact epithelial organization of ESC aggregates in basal conditions [321]. Histological images were initially processed with Cell Profiler and information about the geometries of the cells and nuclei were used to train classifiers that distinguished epithelial and mesenchymal morphologies.

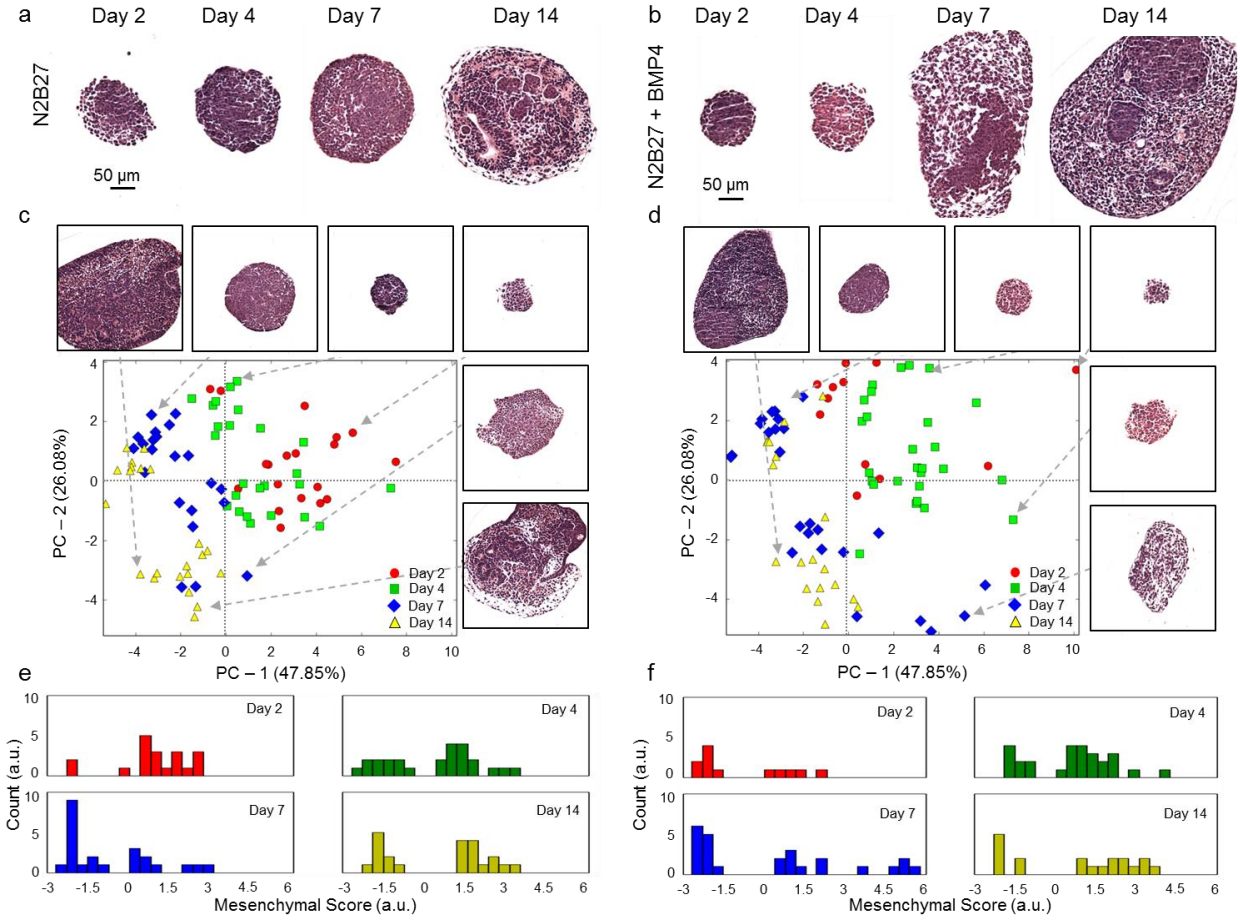

Figure 6-11: Mesenchymal-Like morphogenesis is observed during soluble BMP4 treatment of ESC aggregates. (A) The progression of differentiation over 14 days in culture with basal N2B27 medium. (B) The progression of differentiation over 14 days in culture with BMP4 treatment. (C) Principal component analysis (PCA) of the network derived spatial metrics describing inter and intra-aggregate variability in basal N2B27 treated aggregates (D) and BMP4 treated aggregates. (E) The EMT score calculated for the N2B27 aggregates and (F) bmp4 aggregates.

Using the previously trained mesenchymal classifier, the spatial and temporal distributions of epithelial and mesenchymal sub-networks were analyzed using network based analysis (Figure 6-12) to assess inter-aggregate variability. Individual subnetworks of mesenchymal cells identified with the classification algorithms were analyzed and a set of 19 metrics describing the mesenchymal-like and epithelial like regions were extracted. To examine the distribution of these metrics, principal component analysis (PCA) was employed to reduce the dimensionality of the data set. The PCA model explained 73.93% of the variance: 47.85% from PC-1, and 26.08% from PC-2 (Figure 6-11C,D). PC-1 largely described the extent of differentiation, and was anti-correlated with the total object number and radii whereas PC-2 was anti-correlated with mesenchymal object number and correlated with epithelial object number making it a representative axis of the epithelial/ mesenchymal phenotypic spectrum (Figure 6-13).

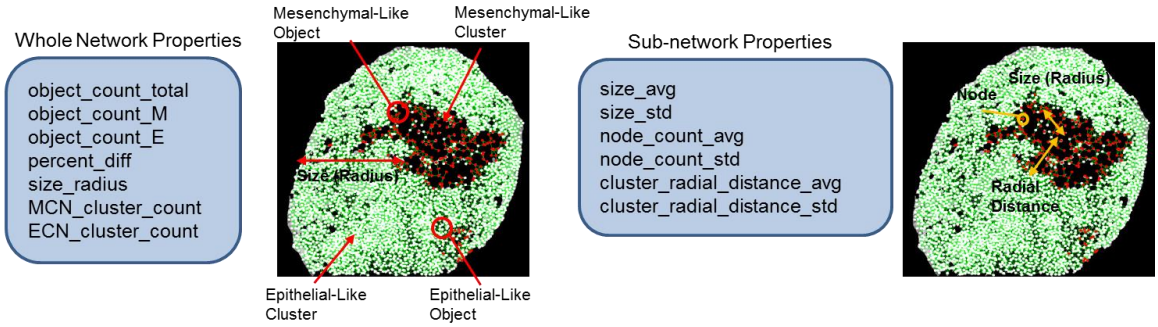

Figure 6-12: Network metrics of mesenchymal regions in ESC aggregates. Identified mesenchymal cells (red) and epithelial like cells (green) are connected into networks (white lines). Different network based metrics are extracted from these on a whole network (left) or subnetwork (right) basis.

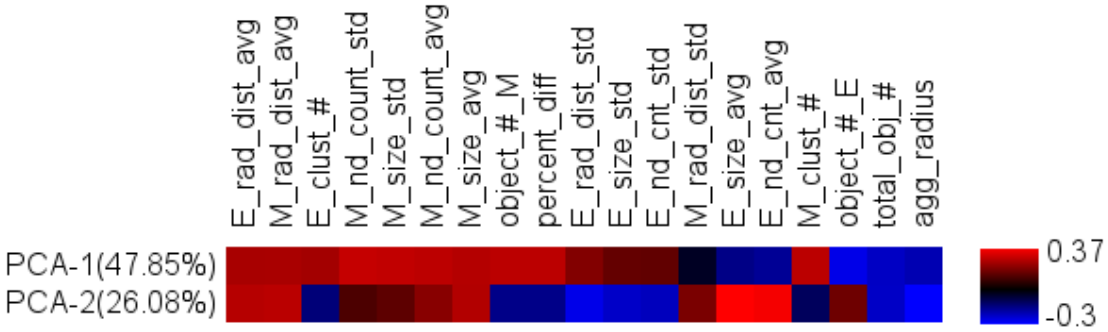

Figure 6-13: Principal component (PCA) metric heat maps. Networks extracted from the networks and their positive (red) or negative (blue) contribution to each principal component.

To further quantify this phenomenon, an EMT score was developed utilizing combinations of the important metrics derived from the PCA model. The EMT score was calculated such that 0 represented an aggregate which had equal amounts of mesenchymal and epithelial like regions, while a negative score corresponded to more epithelial-like regions, and a positive score correlated to more mesenchymal like regions. Using this score a clear evolution of multicellular patterns was evident for both the untreated (Figure 6-11 C) and BMP4-treated conditions (Figure 6-11 D). While the untreated ESC aggregates remained epithelial-like with time (Figure 6-11 C,E), the

BMP4-treated aggregates displayed an increasing mesenchymal population that peaked at day 7, but then subsequently regressed by day 14 (Figure 6-11 D,F). Interestingly at day 2 untreated aggregates displayed slightly less packed morphologies than their BMP4 treated counterparts. These results indicated that, the response to BMP4 treatment appeared to be highly variable, with only a small subset of aggregates (5 of 23 or ~%21) displaying a robust mesenchymal response, suggesting that most of the variation in EMT response is due to inter-aggregate variability.

6.4.3.3 Comparing Mesenchymal Responses induced via Incorporation of Microparticles and Soluble BMP4 Treatment

BMP4 treatment is not the only system which can lead to EMT like events in ESC aggregates. Previously others had reported that introduction of GMA microparticles can also lead to EMT like transition. In this study, three different groups were assessed, a group with microparticle treatment, a group without microparticles, and a group in which the microparticles and the SMAD inhibitor (SB1534) was employed (Figure 6-14 A). For contrast purposes, only the BMP4 results at day 7 were shown as these presented the most robust EMT response in the previous study. The mesenchymal classifiers performed quite robustly on this data set (Figure 6-15), and the resulting data points were added to the previous data set. After classification and network analysis, principal component analysis (PCA) was performed to analyze the resulting variation within the data set. PC-1 was able to explain 47.1% of the variance while PC-2 was able to explain 22.73% (Figure 6-14 B). The variables had similar components weights to those observed previously.

To quantify the resulting mesenchymal morphogenesis, the EMT score was used. In contrast to the day 7 BMP4 treated group where roughly 21% of aggregates responded, the microparticle treated groups displayed robust mesenchymal morphologies (8 out of 16 or 50%) (Figure 6-14 C). A slight response was observed in the basal group (1 out of 15 or 6.66%) and no EMT like events were observed in the inhibited group (Figure 6-14 C). The variation in responses was lower in the inhibited and basal groups than in the microparticle treated groups. Though more aggregates responded in the microparticle groups, the variation appeared to have a similar magnitude as that observed in the BMP4 treatment. Taken together these results suggested that microparticle treatments induced a more robust EMT like response than soluble BMP4 treatment. Furthermore, the variation present during the EMT process was likely due to inter-aggregate variation, as a homogenous EMT response was not observed in either treatment group. Finally, these results highlighted the utility of network based analysis for quantitatively describing spatial and temporal variation present during EMT like events in ESC aggregates.

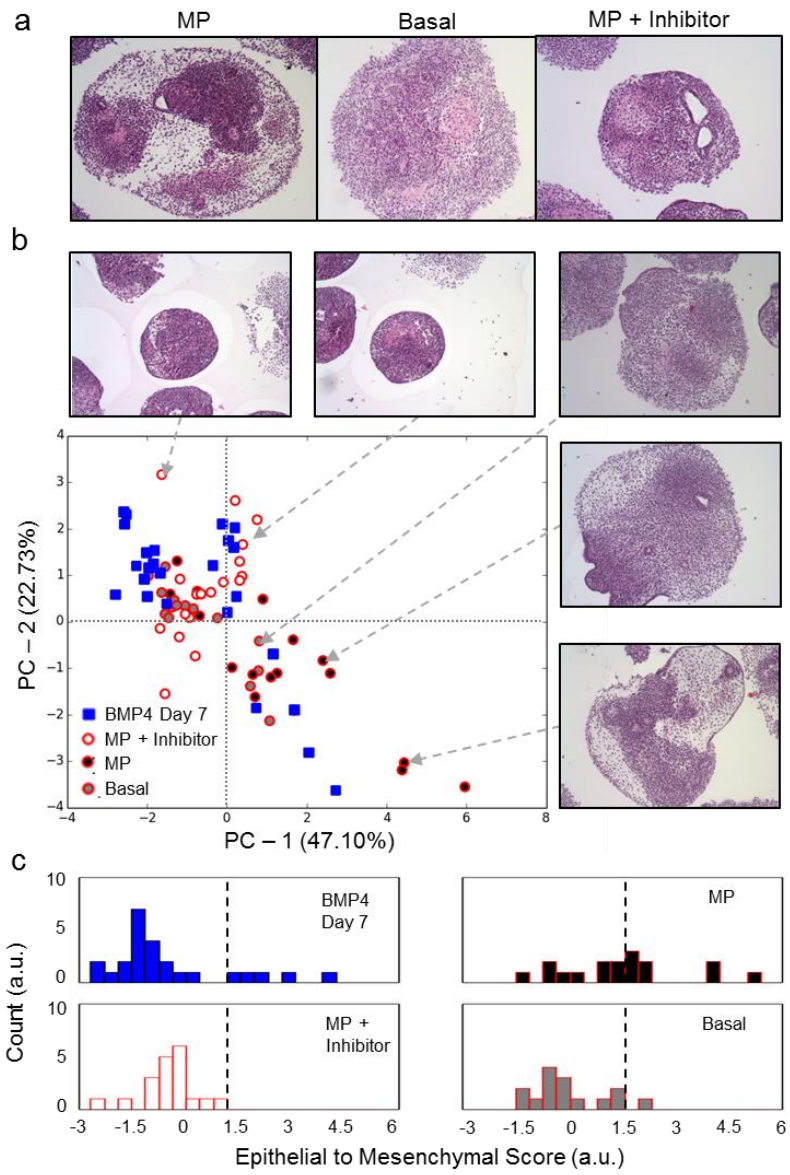

Figure 6-14: Incorporation of microparticles leads to a more pronounced mesenchymal response than soluble BMP4 treatment. (A) Representative images for microparticle (MP) and basal, and inhabited groups (MP + inhibition). (B) Principal component analysis (PCA) of the network derived spatial metrics describing inter and intra-aggregate variability. (C) The EMT score calculated for Day 7 Bmp4 treatment in contrast with the microparticle, inhibition, and basal conditions.

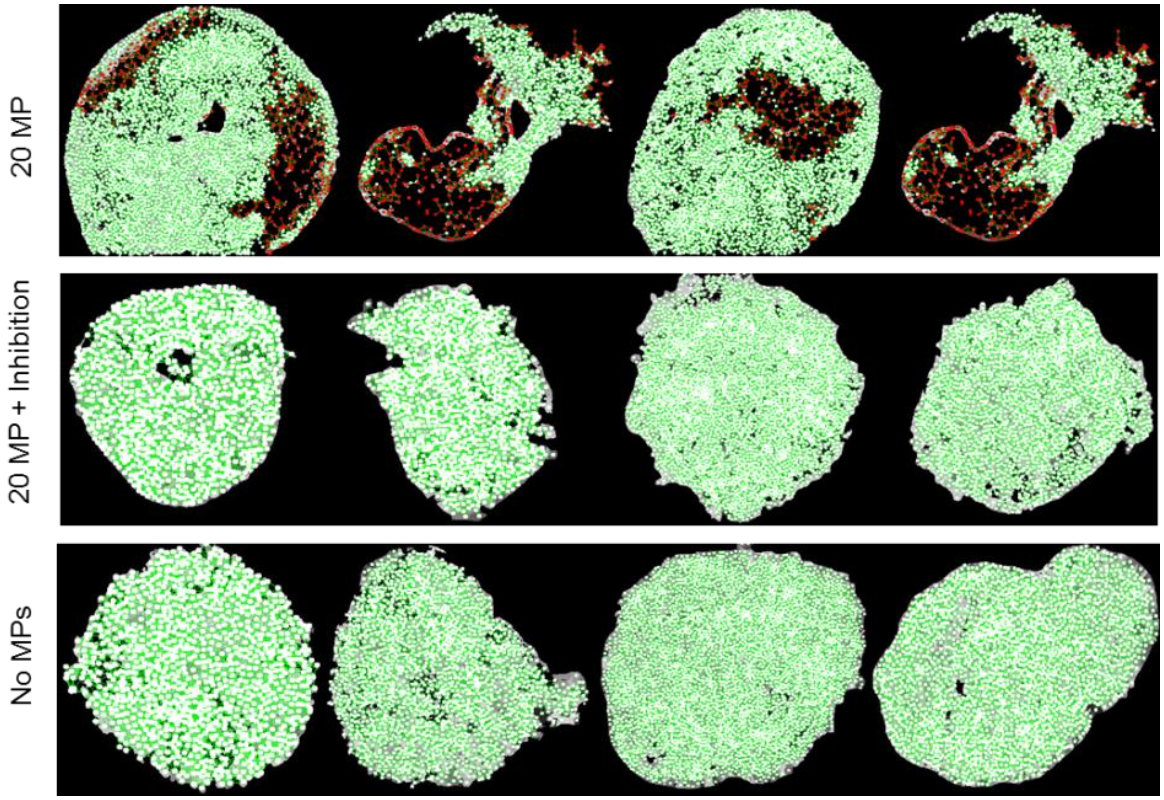

Figure 6-15: Representative images of mesenchymal based classification for ESC aggregates treated with microparticles. Shows mesenchymal regions (red) and epithelial like regions (green) for the three treatment groups: ESC aggregates treated with microparticles, ESC aggregates treated with microparticles and inhibitor, and no microparticle controls.

6.4.3.4 Examining Sources of Complexity in ESC Aggregates

While the EMT score was able to explain some of the heterogeneity in the data set, it was also observed that quite a bit of variation existed along the orthogonal axis (Figure 6-16 A). To assess variation along this axis and how it could explain differences between individual aggregates as well as between different treatment groups, a complexity score was defined. Complexity increased during the differentiation process for both basal (Figure 6-16 B) and BMP4 treated conditions peaking at day 14. In the context of the microparticle studies (Figure 6-16 C), the basal and inhibited groups had a lower complexity score then the microparticle group and also displayed a tighter distribution,

indicating decreased variation between aggregates. Aggregates with high complexity scores contained regions of densely packed cells, lumen, rosettes, and epithelial like membrane structures (Figure 6-16 D). Thus, the complexity score was roughly a measure of inter-aggregate complexity, with larger aggregates containing more structure earning a larger complexity score.

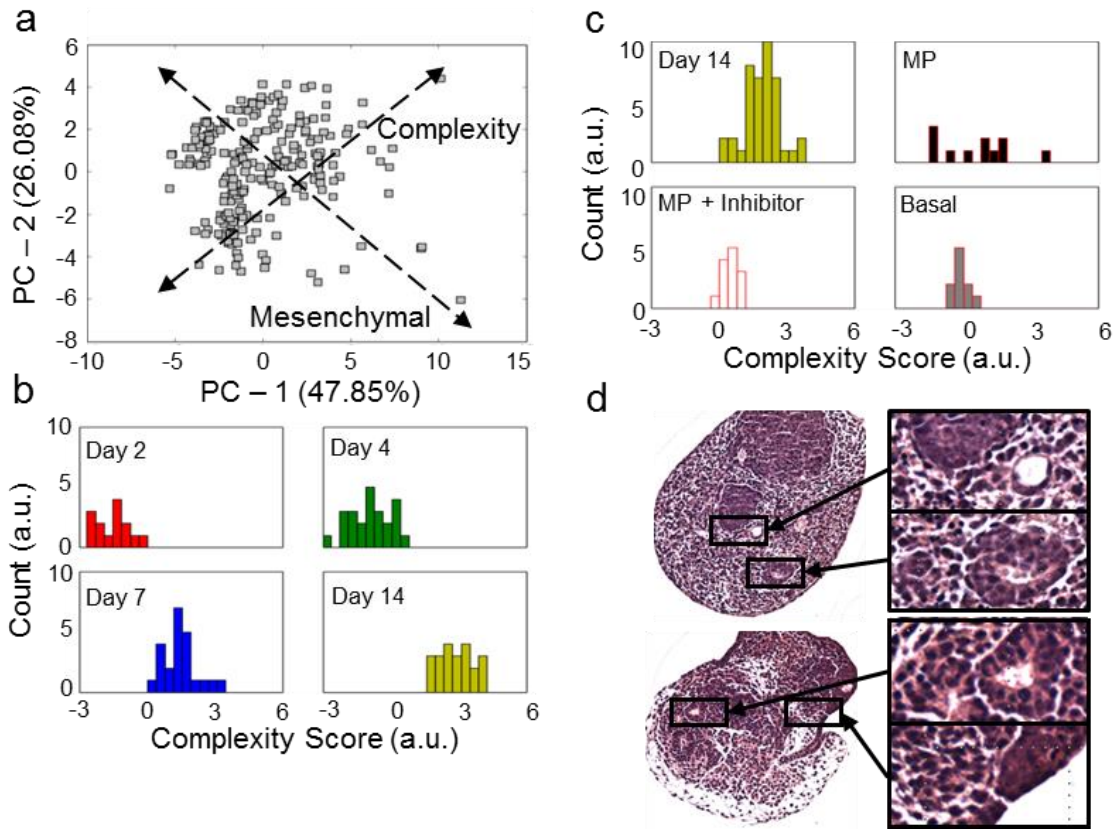

Figure 6-16: Exploring sources of complexity in ESC aggregates. (A) While the mesenchymal axis was clearly defined, another axis could be draw perpendicular to that which also captured a substantial portion of heterogeneity, denoted as the complexity axis. (B) Complexity scores for differentiation in basal N2B27 medium and (C) microparticle studies. (D) representative images showing possible complex structures which could contribute to intra-aggregate variability.

6.4.3.5 Lumen-Like Structures were Enriched in Microparticle Treatment Groups

Lumen classification was performed to assess the degree to which lumen formation contributed to the observable variation in the groups. Briefly lumen like regions were classified using shape descriptors (mainly the area, and form factor) to identify circular regions devoid of cells within aggregates. This worked quite well even in the presence of less dense mesenchymal regions (Figure 6-17 A). To examine the variation by group, the average lumen area (in pixels) and lumen count were plotted (Figure 6-17 B). Generally, aggregates contained many smaller lumen-like structures (Figure 6-17 Ci) as opposed to single larger lumen (Figure 6-17 Cii), however several aggregates had no lumen-like structures at all. To quantify lumen formation a lumen score was assigned which took into account the total area of the lumens present in the aggregate. The microparticle and basal treatment groups displayed relatively high lumen formation, while inhibition generally resulted in less lumen formation. This correlated well with the decrease in complexity score noted earlier (Figure 6-16). The microparticle group displayed higher lumen formation then the basal group, and lumens were observed in all microparticles aggregates as opposed to ~ 50% in basal treatment, again correlating well with the increase in microparticle complexity scores. There was no difference in lumen formation in the basal and BMP4 treated groups at day 14, and a large portion of the aggregates (> 50%) did not have any lumen formation at all, suggesting that the observed complexity in those aggregates must also be coming from another source. Taken together, these data suggest that the variation in complexity data in the microparticle studies could be due to changes in lumen composition between treatment groups.

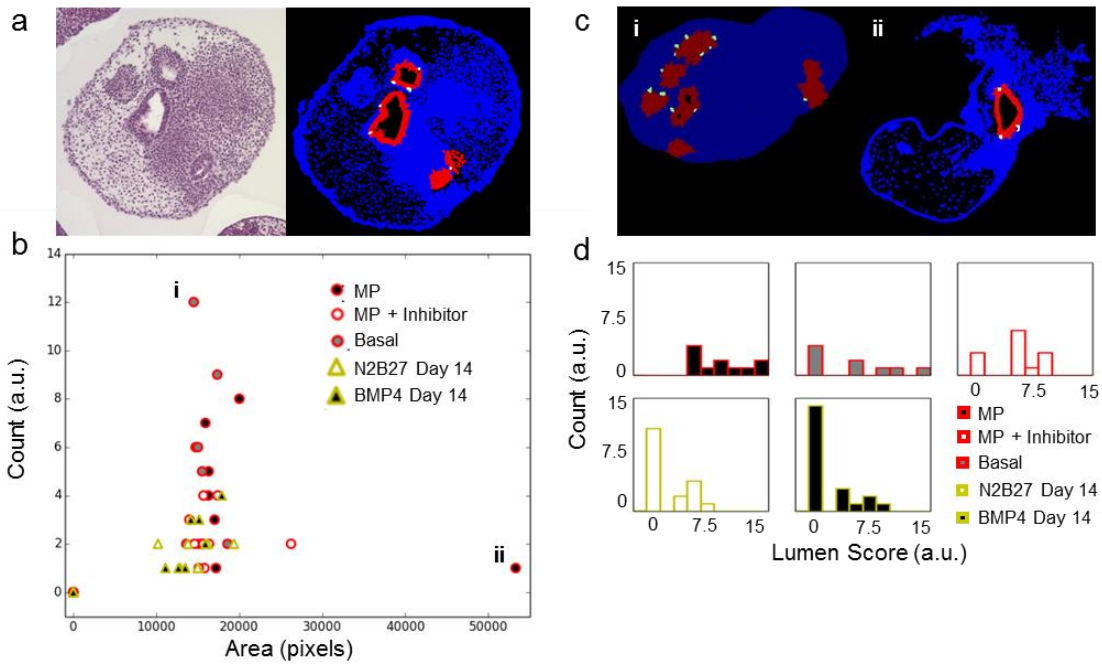

Figure 6-17: Lumen like structures are enriched during microparticle induced differentiation. (A) Representative image of classification of lumen like structures. (B) Lumen count vs average lumen area on a per aggregate basis. (C) Representative images showing an aggregate with (i) many small lumens or (ii) one single large lumen. (D) Lumen score plotted as histograms showing relative lumen distributions across all treatment conditions.

6.4.3.6 Dense Cell Regions were Enriched in Soluble BMP4 Treated Groups

In addition to lumens structures, dense highly aligned and compact nuclear regions reminiscent of polarized epithelial membranes were also observed in the ESC aggregates. To classify regions of dense cells, simple rules for the area, local cellular density, and number of neighbors were applied. This resulted in classification of locally dense highly aligned structures, which were typically present on the outside of aggregates (Figure 6-18 A). Plotting the number of dense structures against the average size of these structures (Figure 6-18 B) revealed a substantial difference between the day 14 BMP4 and N2B27 treated groups (Figure 6-18 C i) in contrast to aggregates cultured in the microparticle study (Figure 6-18 C ii). To quantify these differences further a score for the epithelial

density was calculated. While no appreciable differences could be observed between the individual groups within the microparticle studies, a drastic difference existed between the microparticle and BMP4, and a slight shift could be observed in the BMP4 treated group over the basal control (Figure 6-18 D) .This difference correlates with the increase seen in the complexity score of the Day 14 samples over the microparticle studies (Figure 6-16), suggesting that these aligned epithelial like membrane structures were significantly enriched at day14 in the BMP4 studies, which correlated with the increased complexity observed at that time point. Taken together, these results indicated the BMP4 and microparticle studies produced differences not only in the degree of mesenchymal morphogenesis observed, but also in the types of subsequent complex structures formed even under basal culture conditions.

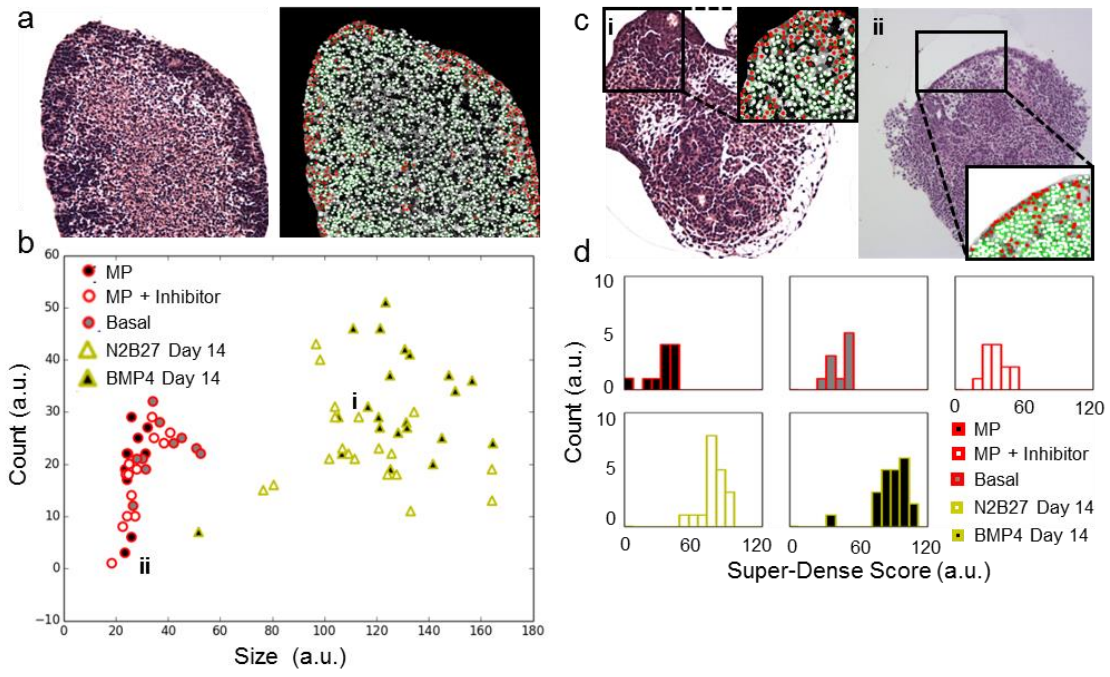

Figure 6-18: Dense epithelial like structures are enriched during microparticle induced differentiation. (A) Representative image of classification of lumen like structures (white nodes represent negative cells, while red represents positive cells. (B) Dense epithelial count vs average area on a per aggregate basis. (C) Representative images showing an

aggregate with (i) a dense epithelial region or (ii) smaller thinner area. (D) Super dense region scores are bar plot showing the score for multiple different groups.

6.4.4 Discussion

Quantitative histology has long been a specialized field requiring a trained eye to pick out structures of interest. Though various efforts have been made to automatically detect cell types and lumen like structures via histology this has yet to be extended to recognizing more complex multicellular structures within ESC aggregates. Utilizing a network based approach this study addressed many of the challenges associated with quantifying complex structures. This work outlined the use of network analysis for classifying single cell phenotypes, as well as more complex aggregate structures in ESC aggregates. Furthermore, this approach allows quantitative assessment of sources of variation within these populations. The ability to derive universal metrics allows the comparison of histological samples across multiple different studies. To this end, the authors specifically excluded intensity information from the classifications schema, eliminating the need for image standardization before analysis. This method was applied to multiple systems of interest including ESC aggregates differentiated in the presence of BMP4, microparticles, and Smad inhibitor. In the BMP4 and microparticle cases mesenchymal differentiation was observed, although it appeared to a greater extent in the microparticle treated groups. Mesenchymal differentiation displayed a large inter-aggregate variation, with large differences in response present in the populations. Some of this heterogeneity could be explained by the formation of other complex multicellular structures within the aggregates. In the case of BMP4 differentiation a lot of the increased complexity observed at day 14 which could not be explained by mesenchymal differentiation was

explained by an increased in epithelial membrane like structures within the ESC aggregates. For the microparticle groups, differences in complexity could be explained by changes in the size and number of lumens formed during the differentiation process. Taken together these results highlight the utility of network based analysis to automatically detect complex multicellular features in histology, and also explain the inter-aggregate variation present across multiple studies of interest.

CHAPTER 7 PROBING MECHANISMS GOVERNING PATTERN FORMATION: A MULTISCALE MODEL APPROACH

7.1 Introduction

The core regulatory mechanisms governing ESC differentiation have been extensively studied, particularly for interactions in monolayer culture. The classic transcription factors Oct4, Sox2, and Nanog have been shown to interact with a larger web of factors including Klf4 and Rex1 all of which cooperate to regulate pluripotency[102, 103]. Furthermore, it is known that factors like Leukemia Inhibitory factor (LIF), and FGF are crucial in regulating self-renewal in ESCs. Specifically, LIF activates JAK/STAT signaling, which activates Sox2 and Klf4, both of which activate Oct4 transcription [67, 291]. On the other hand, FGF4 signaling acts through MEK/ERK, inhibits Nanog expression, and ultimately leads to cells exiting pluripotency [106, 258, 292]. However, even though much is known about the interactions between Oct4, Sox2, and Nanog specific signaling molecules, it is extremely difficult to understand how all of these factors work together to regulate pluripotency. Particularly, the role of Nanog during differentiation has been quite controversial, with a variety of different theories as to how Nanog ultimately controls pluripotency [68, 102, 103, 106-108, 113].

Nanog overexpression studies show that Nanog is sufficient to maintain pluripotency in the absence of external signaling cues [106]. Under culture in LIF conditions, mouse ESCs display high levels of Sox2 and Oct4, but a heterogeneous population of Nanog can be observed [68, 102, 108, 109]. It is generally agreed that stochastic fluctuations in Nanog expression are the cause of the observed Nanog high and low states, however, the source of this stochasticity is unknown. A recent paper by Ochiai

et al. [109] showed that the contribution of intrinsic cell noise (represented by differences in cell-cell protein level) was about 45% as compared to extrinsic noise (related to environmental factors) in the total levels of Nanog heterogeneity. While Sasai *et al.* was able to explain the evolution of ESC pluripotency through a model implementing stochasticity at the epigenetic level, this has yet to be corroborated by experimental studies [293]. In fact, no epigenetic modifications which correlate in differences in Nanog expression have been identified [68, 109]. Other groups have shown that adding stochastic terms into Nanog expression was able to capture the relative proportions of cells in low and high states during various stages of differentiation [125, 130, 132]. However, investigations of the extrinsic factors governing this process have been largely ignored.

FGF4 signaling acting through MEK/MAPK pathways inhibits Nanog expression [68, 112, 113, 294]; thus FGF4 could be one such extrinsic factor responsible for heterogeneity in ESC differentiation. Furthermore, FGF4 is known to be secreted by ESC cells in the pluripotent state which mirrors the role of FGF4 during development where it is responsible for specifying early mesoderm and endoderm progenitors over ectodermal differentiation pathways. In particular, Nanog is known to repress expression of early endoderm transcription factor GATA6 which in term, repress expression of Nanog [113]. This is a classic example of a bi-stable toggle switch, which has long been specified to be important in governing key cell fate decisions at the transcriptional level. By inhibiting Nanog levels, FGF4 pushes ESCs towards differentiation down the endoderm lineage. Furthermore, FGF4 is thought to inhibit the expression of Nanog via activation of

repressors, while Nanog's relatively short half-life (~ 2 hours) clears Nanog from the system.

Though FGF4 represents a promising soluble factor which could be responsible for this heterogeneity in Nanog expression, it is difficult to study the soluble gradients formed by such factors during the differentiation process. Furthermore, it is not clear exactly which receptor FGF4 is signaling to during this process, further confounding the results. Therefore, one promising approach to examine Nanog heterogeneity induced via FGF4 signaling is to investigate this phenomenon via computational modeling strategies. While various computational models have been constructed to look at ESC differentiation they suffer from a variety of problems, mainly instantaneous fate switches, and incorporation of scaling factors to induce differentiation [113, 125, 130, 294]. Herberg *at al.* addressed this problem recently using a multiscale modeling approach to explain Nanog expression heterogeneity in ESC cultures [106]. However, in this approach FGF4 diffusion was not directly simulated; instead the authors opted to model FGF4 signaling as proportional to the contact area of neighboring cells. Results from Chapter 5 (and appendix A. 2) suggested that modeling soluble signaling versus local signaling can impact the overall pattern formation, which question the assumption of modeling FGF as a local cell based interaction, particularly in a freely diffusible 2D environment.

The objective of this study was to investigate the roles of inherent stochasticity and intrinsic FGF4 signaling in explaining heterogeneous Oct4 expression patterns observed in ESC aggregates. It was hypothesized that inclusion of inherent Nanog stochasticity with responsiveness to FGF4 would lead to the formation of complex spatial patterns. This study showed that Nanog heterogeneity, coupled with FGF4 expression

was capable of capturing spatial patterns associated with loss of Nanog, and ultimately the eventual loss of Oct4/Sox2. Taken together these data represented a significant advancement in multiscale ESC aggregate models by integrating stochastic ODE and soluble PDE species together, and provided insight into the role of Nanog signaling in governing pluripotency.

7.2 Materials and Methods

7.2.1 Network Reconstruction of Confocal Images

Cell Profiler (http://www.cellprofiler.org/)[278] was used to analyze all of the 2D samples. In this case, four different channels were captured in the confocal images to measure Sox2 (white), Oct4 (red), Nanog (green) and nuclei (blue). Therefore the confocal images were split into the component channels using an ImageJ macro resulting in 4 images for each original confocal file. In cell profiler the images were imported, and the nuclei were thresholded using a local MCT approach. The resulting binary mask was applied to all other channels: Oct4, Sox2, and Nanog. Clumped nuclei were separated using the intensity with a Laplacian of Gaussian (LOG) filter. Intensity information was extracted from each nuclei identified via the segmentation for all four different species. Additionally, the number of adjacent nearest neighbors was measured and the data was then exported to a python script that reconstructed the networks by using a KDTree implementation from Scipy. Images of the networks were generated using the python imaging library (PIL). Digital flow cytometry was performed by extracting the intensity values for the different colors, and normalizing them by the intensity of the nuclei in that area. This normalization was performed to negate local aberrance in the images

(equivalent to background subtraction). This normalized intensity value was also used for network annotation during the subsequent reconstruction of networks.

7.2.2 Flow Cytometry

ESCs were washed in PBS, and then trypsanized in .025% Trypsin with EDTA. Cells were then pelleted by centrifugation at 5 min at 1000 RPM. After exchanging the supernatant, the cells were fixed in 10% formalin for 15 min at 4 ^{o}C. Next cells were diluted into a working solution (.3% BSA 0.001% Tween by volume), then pelleted at 4000 RPM for 2 min. After extracting the supernatant, the resulting cells were resuspended in 1 ml of permeabilizing solution (.5% Triton X in sterile H_20). After centrifugation and aspiration, the cells were placed in blocking buffer (10% normal Donkey Serum (Jackson Scientific)) for 1 hour (vortexing every 15 minutes to prevent settling). After pelleting and removing the blocking buffer the cells are split into tubes, placed into 100 ml working buffer, and incubated with primary antibodies. In this case, a tri-staining assay for pluripotency was performed. The primary antibodies used were from eBiosciences (1:100 dilution, rat-α-mouse monoclonal, eBioMLC-51) for Nanog, Santa Cruz (1:100 dilution, goat-α-mouse N-19) for Oct 4, and Santa Cruz (1:100 dilution, polyclonal rabbit-α-mouse H-65) for Sox2. After incubation samples were diluted into a 1 ml working buffer volume, pelleted, and then resuspended in 100 ml working buffer for secondary staining. Secondary antibody staining was performed (1:200 dilution, all secondary's purchased from Invitrogen) and allowed incubate for 30 min at 4 ^{o}C. The secondary antibody combinations used were donkey-α-rat AlexaFlour-488 for Nanog, donkey-α-goat AlexaFlour-547 for Oct4 and donkey-α-rabbit AlexaFlour-633 for Sox2. After incubation, cells were pelleted, and washed in working buffer for two cycles before a final resuspension in PBS. Cells were run through the Accuri© flow cytometer using high fluidics speed, and measured on the Fl-1 (488), Fl-2 (546), and Fl-4(633) channels respectively. Color compensation between the Fl-1 and Fl-2 channels

was determined using single stained controls, and samples were adjusted in the Accuri software using the set color compensation function, and as was consistent with the literature, compensation was only necessary between the red (546) and green (488) channels.

7.2.3 Immunostaining and confocal microscopy

ESC aggregates were collected for staining and fixed in 10% formalin for 45 minutes. EBs were permeabilized for 30 minutes in 1.0% TritonX-100, re-fixed in formalin for 15 minutes, and blocked in blocking buffer (2% bovine serum albumin, 0.1% Tween-20 in PBS) for 3 hours. For Nanog, Oct4, Sox2 tri-staining the following antibodies were used: eBiosciences (1:100 dilution, rat-α-mouse monoclonal, eBioMLC-51) for Nanog, Santa Cruz (1:100 dilution, goat-α-mouse N-19) for Oct 4, and Santa Cruz (1:100 dilution, polyclonal rabbit-α-mouse H-65) for Sox2. The secondary antibody combinations used were donkey-α-rat AlexaFlour-488 for Nanog, donkey-α-goat AlexaFlour-547 for Oct4 and donkey-α-rabbit AlexaFlour-633 for Sox2 all at a 1:200 dilution. It is important to note that for this protocol, Sox2 rarely worked with the 633 secondary, suggesting an alternative Sox2 antibody may be in order or a new secondary combination may be necessary. Counterstaining with Hoechst (1:100) was performed for 25 minutes. Samples were washed, resuspended in blocking buffer, and imaged using a Zeiss LSM 710 Confocal Microscope. A single image was taken at the top of the EB and at a depth of 45 μm into the EB. For each time-point, 20 images were obtained.

7.2.4 Stochastic ODE Modeling

Stochastic differential equation systems were modeled using a custom simulation code written in python. A previous stochastic model for Nanog fluctuations containing three species (Oct4-Sox2, Nanog, and FGF) was modified to contain an additional LIF soluble

signal, as well as regulation of the Oct4-Sox2 complex via LIF (through downstream activation of STAT3), and a modulation of FGF secretion by Nanog. Furthermore to specify the differentiation state, Gata6 was added to the model. Gata6 has been shown to inhibit Nanog, and Oct4-Sox2-Sox2 expression and serves as one of the earliest toggle switches in ESC differentiation. Additionally, Stat3 activation via LIF via JAK signaling molecules has been shown to be inhibited in the presence of FGF signaling via MEK/ERK, thus this was incorporated into the model [295]. This led to the system of equations shown in 7-1 through 7-5.

$$\frac{dOS}{dt} = \frac{a_{OS\_OS}OS}{\left(\frac{1}{k}+OS+i_{Gata6\_OS}*Gata6\right)} + \frac{a_{LIF\_OS}*LIF}{\left(\frac{1}{k}+LIF+ i_{FGF\_LIF}*FGF^2\right)} + \frac{a_{N\_OS}*N}{\left(\frac{1}{k}+N\right)} - k_{\deg\_OS} * OS + \gamma(\sigma_{OS}) * OS \quad (7\text{-}1)$$

$$\frac{dN}{dt} = \frac{a_{N\_N}(N*OS)}{\left(\frac{1}{k}+N*OS+i_{FGF\_N}*FGF\right)} - \frac{i_{Gata6\_N}*Gata6}{\left(\frac{1}{k}+Gata6\right)} - k_{\deg\_N} * N + \gamma(\sigma_N) * N \quad (7\text{-}2)$$

$$\frac{dLIF}{dt} = k_{prod\_LIF} - k_{\deg\_LIF} * LIF \quad (7\text{-}3)$$

$$\frac{dFGF}{dt} = \left(\frac{OS}{\left(\frac{1}{k}+OS\right)}\right) * k_{prod_{FGF}} + \left(\frac{Gata6}{\left(\frac{1}{k}+Gata6\right)}\right) * k_{prod\_FGF} - k_{\deg\_FGF} * FGF \quad (7\text{-}4)$$

$$\frac{dG6}{dt} = \left(\frac{a_{LIF\_OS}Gata6}{\left(\frac{1}{k}+Gata6+i_{N\_Gata6}*Nanog+i_{LIF\_Gata6}*LIF\right)}\right) + \gamma(\sigma_G) - k_{\deg\_Gata6} * Gata6 \quad (7\text{-}5)$$

Where N represents Nanog and OS represents an Oct4-Sox2 complex. Stochastic noise is generated as denoted by the γ which indicates a zero mean Gaussian process with a standard deviation of σ. It was assumed that appreciable noise is only associated with Oct4-Sox2, Nanog and Gata6 species, and no stochastic noise is associated with soluble

factor production. Parameters and initial conditions for the three different test cases, LIF, no LIF, and no FGF are shown in Table 7-1 through Table 7-6.

Table 7-1: Parameter values for LIF simulations

Parameter	Value	Parameter	Value
a_OS_OS	40	i_FGF_N	13
a_LIF_OS	50	i_FGF_LIF	50
a_N_O	1	kdeg_OS	0.8
a_N_N	29	kdeg_N	1
a_G_G	20	kdeg_FGF	0.1
i_N_G	20	kdeg_LIF	0.1
i_G_N	20	kdeg_Gata6	1
i_Lif_G	0.01	σ_N	0.3
i_OS_G	20	σ_OS	0.15
prod_FGF	1	σ_Gata6	0.15
prod_LIF	1		

Table 7-2: Initial Conditions values for LIF simulations

Parameter	Value	Parameter	Value
LIF	100000000	Nanog	1
FGF	0	Oct4	50
Gata6	0.1		

Table 7-3:Parameter values for no LIF simulations

Parameter	Value	Parameter	Value
a_OS_OS	40	i_FGF_N	13
a_LIF_OS	50	i_FGF_LIF	50
a_N_O	1	kdeg_OS	0.8
a_N_N	29	kdeg_N	1
a_G_G	20	kdeg_FGF	0.1
i_N_G	20	kdeg_LIF	0.1
i_G_N	20	kdeg_Gata6	1
i_Lif_G	.01	σ_N	0.3
i_OS_G	20	σ_OS	0.15
prod_FGF	1	σ_Gata6	0.15
prod_LIF	1		

Table 7-4: Initial Conditions values for LIF simulations

Parameter	Value	Parameter	Value
LIF	0	Nanog	1
FGF	0	Oct4	50
Gata6	0.1		

Table 7-5: Parameter values for no FGF simulations

Parameter	Value	Parameter	Value
a_OS_OS	40	i_FGF_N	0
a_LIF_OS	50	i_FGF_LIF	0
a_N_O	1	kdeg_OS	0.8
a_N_N	29	kdeg_N	1
a_G_G	20	kdeg_FGF	0.1
i_N_G	20	kdeg_LIF	0.1
i_G_N	20	kdeg_Gata6	1
i_Lif_G	0.01	σ_N	0.3
i_OS_G	20	σ_OS	0.15
prod_FGF	1	σ_Gata6	0.15
prod_LIF	1		

Table 7-6: Initial Conditions values for no FGF Simulations

Parameter	Value	Parameter	Value
LIF	100000000	Nanog	1
FGF	40	Oct4	50
Gata6	0.1		

7.2.5 Multiscale Computational Modeling

Multiscale computational modeling was accomplished utilizing previously described agent based model, coupled with the stochastic ODE system described above (section 7.2.4). Briefly, cells are allowed to divide, and move based on rules governing their behavior. Physical constraints were applied using a KDTree collision detection based method with springs to assume inelastic collisions. Over each hour, a stochastic ODE simulation was run for each cell based on the internal variables (Section 7.2.4). The values of the Oct4, Nanog, and Gata6 species are stored for each time (over 144 hour time span) and used to determine the state of the cell. The spatial equations defining the

constants associated with LIF and FGF gradients were based on the previously described methods (Section 5.2.3). The parameter governing stochasticity in Nanog was varied between 0.3 and 0.7 (0.3, 0.4, 05,. 0.6, 0.7), and the parameter governing FGF production was varied between 1 to 1000 (1, 10, 250, 500, 1000) to investigate potential roles of each parameter in governing spatial patterns.

7.2.6 Sensitivity Analysis

Sensitivity analysis was performed for each parameter (n = 22) and calculated according equation 7.5.

$$S = \sum_{t=0}^{t=n} \frac{F(x_o)-F(x_1)}{x_o-x_1}\frac{1}{n} \quad (7\text{-}5)$$

Where x_0 and x_1 were the values of the parameters and $F(x)$ was the ODE function evaluated for the parameter value x. The resulting sensitivity was calculated with respect to each of the species in the model: Oct4, Nanog, Sox2, FGF, LIF and Gata6. An average sensitivity was also calculated by taking the average values of all species for each parameter. All sensitives are reported on a logarithmic base 10 scale.

7.2.7 Monte Carlo Simulation

In order to assess the topology of the parameter space and ultimately different models of model behavior, Monte Carlo simulations were used. A set of 200,000 parameters was generated to span the parameter space associated with the stochastic model described in section. Parameters where chosen randomly accordingly a normal distribution with the ranges specified in Table 7-7.

Table 7-7: Parameter Ranges for Monte Carlo Simulation

Parameter	Range	Parameter	Range
a_OS_OS	1-100	i_FGF_N	1-50
a_LIF_OS	1-100	i_FGF_LIF	1-100
a_N_O	0.1-10	kdeg_OS	.5 - 2

Table 7-7: Continued

a_N_N	1-50	kdeg_N	0.5 – 2
a_G_G	1-50	kdeg_FGF	0.05-0.25
i_N_G	1-50	kdeg_LIF	0.05 - .25
i_G_N	1-50	kdeg_Gata6	0.5 - 2
i_Lif_G	0.0001 -0 .1	σ_N	0.1-0.7
i_OS_G	1, 50	σ_OS	0.01-0.03
prod_FGF	0.1-10	σ_Gata6	0.01-0.03
prod_LIF	0.1-10		

To achieve enough confidence in the simulation result, 100 simulations were run for each parameter set. The mean values for Oct4, and Nanog were calculated over time, and compared back to the basal simulation values as a means of visually representing the data. Additionally, a difference for the residence of Nanog low vs. Nanog high states was also computed to add a third dimension for comparison.

7.3 Results

7.3.1 Capturing Nanog/Oct4 Dynamics During Differentiation

To confirm the presence of Nanog heterogeneity in mouse ESC culture, ESCs were cultured in 2D in the absence of LIF for a period of 4 days. Every 12 hours, cells were removed from the plate, stained, and then run through flow cytometry to assess the Nanog, Oct4 and Sox2 expression during loss of pluripotency (Figure 7-1 A). A Nanog high population could be observed in ~27% of the cells at early time-points suggesting the presence of Nanog heterogeneity in the ESC population (Figure 7-1 B-C). To quantify population shifts, the geometric mean of the population was calculated and plotted as an average over three replicates. Nanog was lost earlier than Sox2 and Oct4, falling to a minimum value by day 1.5 while Oct4/Sox2 fell to minimum values at ~ day 2 (Figure 7-1 E). This difference in kinetics is consistent with experimental observations showing loss of pluripotency proceeds through an Oct4+/Sox2+/Nanog- intermediate before

subsequent loss of Oct4 and Sox2. High correlation between Oct4 and Sox2 levels was also observed, consistent with the literature (Figure 7-1 E). Based on this observation and the literature it was assumed for subsequent modeling efforts that Sox2/Oct4 can be simplified to a single active species.

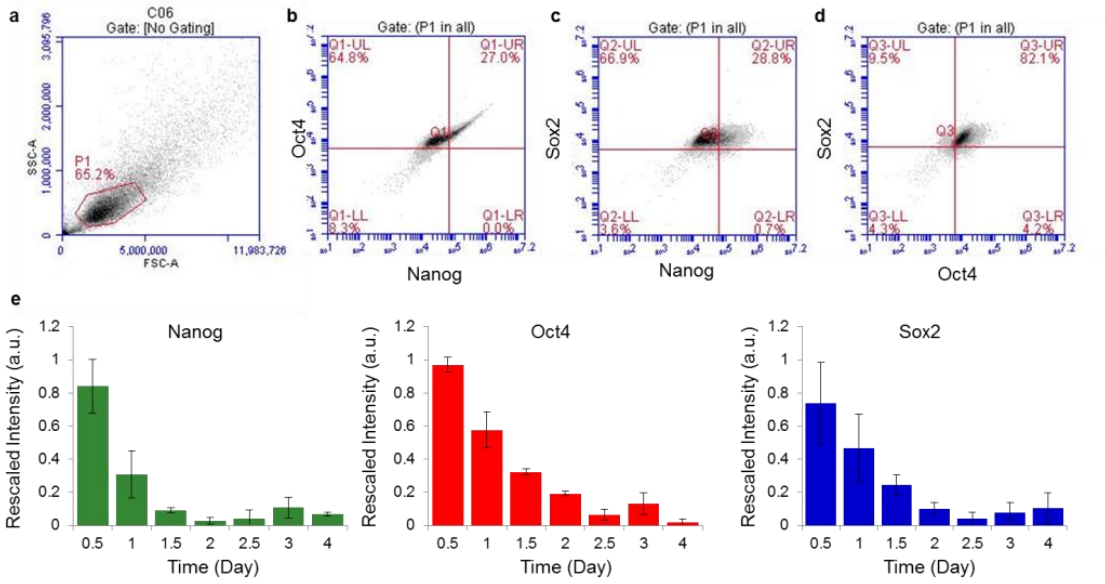

Figure 7-1: Nanog, Oct4 and Sox2 dynamics in mESC monolayer culture. (A) Representative side-scatter vs. forward-scatter flow cytometry plot showing the gate applied for subsequent analysis. (B) Nanog vs. Oct4 showing Nanog high and Nanog low states. (C) Nanog vs. Sox2 showing Nanog high and Nanog low states. (D) Oct4 vs. Sox2 shows good correlation of Oct4 to Sox2. (E) Extracted geometric mean over time for Nanog (green), Oct4 (red) and Sox2 (blue) (n=3).

To confirm that similar distributions of Nanog were also preset in 3D aggregates, ESC aggregates were cultured over 9 days to observe the loss of pluripotency. Samples were removed at Day 1, 4, 7, and 9 and assessed via confocal microscopy. To quantify individual cellular levels, digital flow cytometry was performed on the image stacks (Figure 7-2). At day 1, roughly 20% of the cell population was in the Nanog high state, while 80% was in the Nanog low state (Figure 7-2). This suggested that a substantial amount of heterogeneity already existed in this population even during early phases of

differentiation. Furthermore, the variance of the Nanog distribution was much broader than subsequent days, which was again consistent with experimental observations by others (Figure 7-2). By day 4, the population had shifted towards a mostly Oct4+/Nanog- population, although a small portion (~10%) of Oct4+/Nanog+ cells remained. By day 7 the population had shifted towards Oct4-/Nanog-, but approximately 30% of the cell remained in the Oct4+/Nanog- state. However, by day 9, 77% of cells were Oct4-/Nanog- suggesting that loss of Oct4 had finally occurred. This overall population shift was consistent with prior reports of monolayer culture for ESCs[106, 294] and suggested a model in which Nanog was lost early in differentiation, while Sox2/Oct4 were lost later. Interestingly Oct4+/Nanog+ cells were observed even during subsequent differentiation events throughout the culture period, which suggested the maintenance of a resident pluripotent population during subsequent differentiation stages.

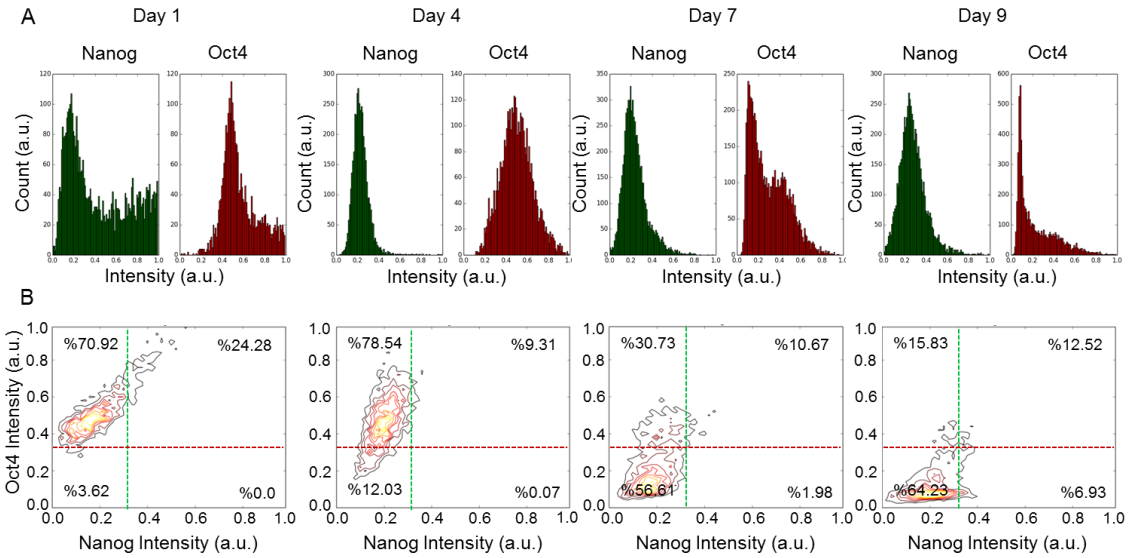

Figure 7-2: Oct4 Nanog differentiation dynamics in ESC aggregates assessd via digital flow cytometry. (A) Histograms of Nanog (green) and Oct4 (red) populations at day1,4,7 and 9 of differentiation. (B) Contour plots of Oct4 vr. Nanog populations, showing the progession of Nanog from a high to low state, followed by a subsequent change in Oct4 expression.

7.3.2 ODE Models of Nanog Fluctuations

To investigate the observed Nanog fluctuations, a computational model in the literature capable of capturing Nanog fluctuations was implemented. This model consisted of a single Oct4/Sox2 species which encompassed the dynamics of both Oct4 and Sox2. This was a valid assumption for a number of reasons: first Oct4 and Sox2 dimerize to form a transcriptionally active dimer which activate Oct4/Sox2 expression in a positive feedback loop [36, 78, 79, 125]. Second, data in Figure 7-1 suggested that Oct4 and Sox2 co-vary which was consistent with literature findings [102, 106, 258]. This model included FGF signaling as a repressive factor of Nanog signaling, and also included stochastic noise in Nanog protein levels. However, in order to update this model based on findings in Chapter 5, the activation of Oct4/Sox2 via LIF (which activates STAT3) was added to the model. Additionally, Stat3 activation via JAK has been shown to be inhibited by MEK induced phosphorylation of ERK. Furthermore, several papers have shown that Nanog also activates Oct4/Sox2 transcription. These modifications resulted in the regulatory scheme shown in Figure 7-3.

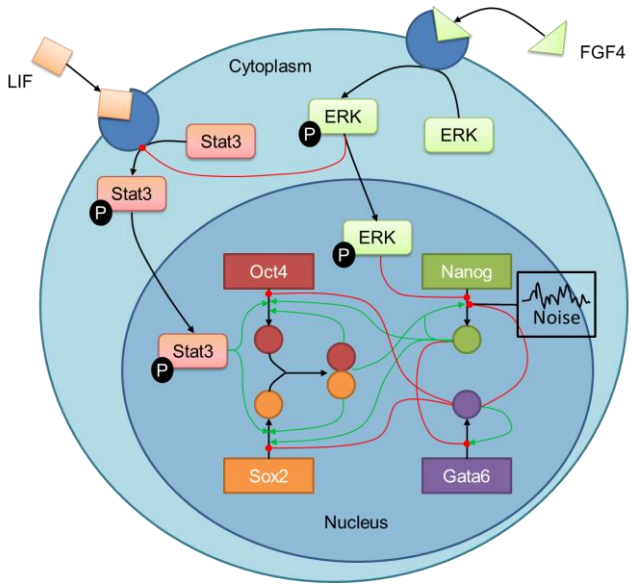

Figure 7-3: Signalling Network Regulating ESC pluripotency. The core transcription factor network for Oct4, Sox2 and Nanog is shown in the nucleus. Binding of FGF4 to

it's receptor leads to activation of ERK via phosphorlyation, which in term inhibits Nanog expression via an unknown mechanism. Binding of LIF to it's receptor leads to phosphorylation of Stat3 which translocates to the nuclues and actiavtes Sox2 and Oct4 gene expression. Introduction of noise represents the inherent stochsticity shown to be associated with Nanog transcription.

To test if this network architecture was capable of recapitulating differentiation, several scenarios where tested: a scenario in which LIF was absent which should induce differentiation, a scenario in which LIF was present which should induce heterogeneous Nanog expression, and a scenario in which ESCs were maintained in 2i which inhibited FGF4 signaling via ERK, thus reducing Nanog heterogeneity.

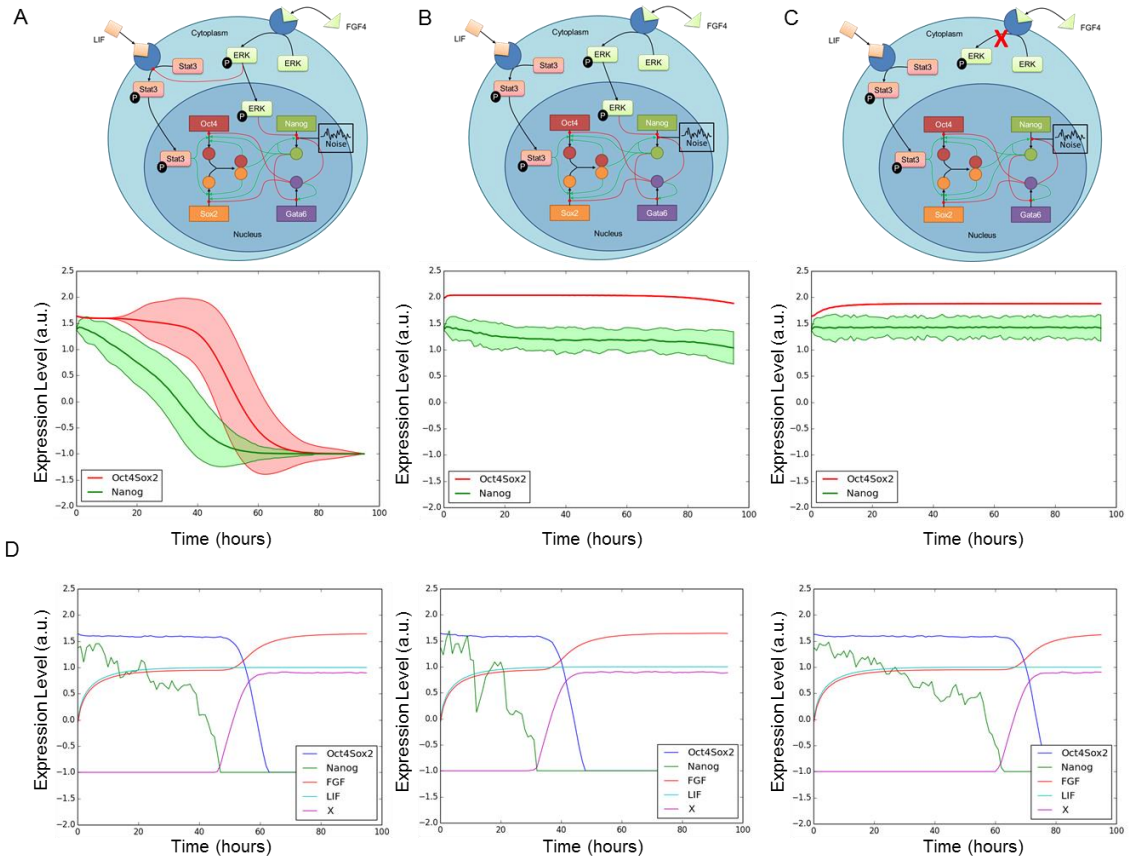

Figure 7-4: Stochastic Network Modeling of ESCs. (A) The model architecture for LIF culture conditions (top) and the resulting simulation average trajectories (bottom) for Oct4 (red) and Nanog (green) (n=1000). (B) Model for 2i culture conditions (top) and the resulting simulation average trajectories (bottom) for Oct4 (red) and Nanog (green)

(n=1000). (C) Model for differentiation conditions (top) and the resulting simulation average trajectories (bottom) for Oct4 (red) and Nanog (green) (n=1000). Middle lines represent means and outer lines represent +/- one standard deviation. (D) Individual simulations showing the heterogeneity present in differentiation time over a 2 day period.

In the case of differentiation (i.e. no LIF present) the model was fit to differentiation data from early ESC monolayer cultures (Figure 7-1), and rapid loss of Nanog could be observed. Once Nanog was lost Oct4 would follow within a time-frame of about 24 hours, which was consistent with the previous monolayer culture data (Figure 7-1). To test the validity of the fitted parameters, the model was challenged with two hypothetical scenarios, culture in LIF, and culture in 2i. In the case of LIF maintained pluripotency, the model rapidly moves to a steady state level of Oct4/Sox2 complex (Figure 7-4 B). However, various amounts of Nanog heterogeneity could be observed, which to matched the 20% high versus low population present in ESC aggregate culture (Figure 7-2). When FGF signaling was inhibited (as with culture in 2i), the observed heterogeneity in Nanog signaling is abolished (Figure 7-4 C), which is consistent with other studies [106, 109, 258, 294]. Importantly, the parameters here suggest that while the intracellular network components were important in stabilizing the various states, the switches between states were governed by robust signaling via soluble factors.

Once the simulation parameters were determined, sensitivity analysis was performed to investigate influences of each parameter (Figure 7-5). Each parameter was varied and the effect on each of the output variables was measured. The aggregate sensitivity for each parameter was performed by taking an average of the sensitivities over all species. From this exercise several key parameter emerged: p1, p5, and p16 (Nanog degradation rate) all strongly modulated Nanog expression, while p1 and p5 (Oct4/Sox2 activation parameters) as well as p13-16 (protein degradation parameters)

were all Oct4 regulators. In contrast, for Gata6 the positive feedback (p6) was the only parameter which drastically affected Gata6 regulation. Generally these parameters (p1, p5, p6, p13-16) correlated to the internal inhibition and degradation constants present in the network, suggesting that the network architecture needed to be finely tuned to recapitulate differentiation behavior. Interestingly, the parameters governing stochasticity for Nanog (p18) and FGF production (p21) displayed moderate influence on model outcomes, suggesting that these could be interesting to investigate for their roles on spatial differentiation patterns.

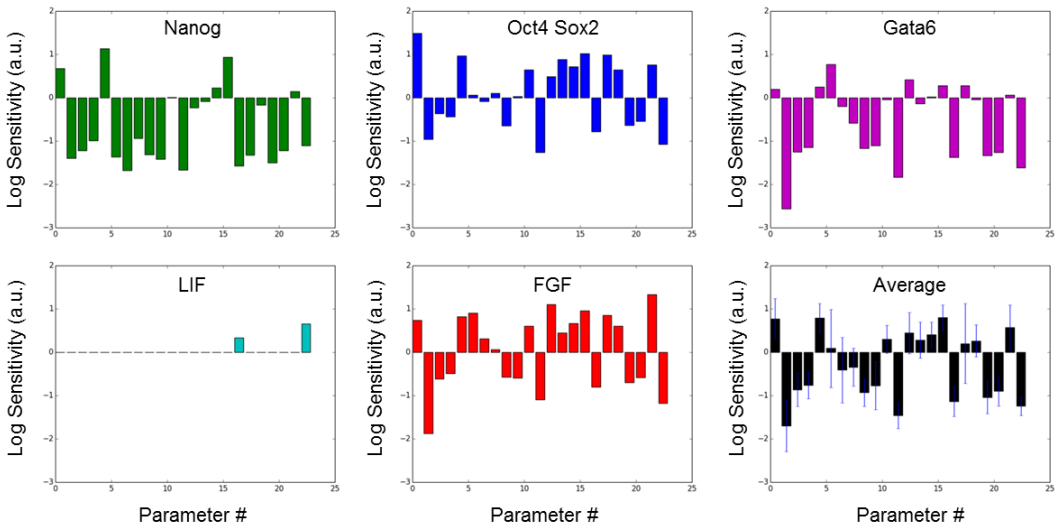

Figure 7-5: Senstivity Analysis of the ESC pluripotency network. For each species in the model sensitvity analysis was performed: Nanog (green), Oct4-Sox2 (blue), Gata6 (magenta), LIF (cyan), FGF (red), average (black). For the average, bars represent means and errorbars are +/- one standadrd deviation from the mean. All senstivites are graphed on a log10 scale.

Next to assess the global stability of this parameter set Monte Carlo simulations were run (Figure 7-6). Using the results for the sensitivity analysis, parameter bounds were determined for all relevant parameters in the model. Then several Monte Carlo parameter sets were generated to equally sample the parameter distributions, and the distance from

the resulting simulation trajectory to the baseline fitted trajectory was computed for the Oct4 and Nanog species as well as the Nanog high/low residency time (Figure 7-6 A-C). For Oct4 and Nanog the top 100 parameter sets leading to maximal down-regulation and up-regulation were computed (Figure 7-6 D-F). For Oct4 p5, p13, p15 and p21 increased expression of Oct4, while p14 substantially decreased expression. For Nanog p12, p13 and p21 led to increased levels, while p5 p13 and p15 led to decreased levels. Surprisingly, only a minimal effect was observed for Nanog stochasticity (p18) on the high/low Nanog residency time. The parameter p13 represented the scaling factor for the feedback response curves, while p15 and p15 represented the Oct4 and Nanog degradation constants respectively; however, p21 represented the production term for FGF4. These results confirmed the findings of the sensitivity analysis, and particularly highlighted the role of FGF4 production (p21). Collectively these results demonstrated that the final parameter set chosen for the model was appropriate for describing the experimental data. Furthermore, the combined Monte Carlo and sensitivity analysis revealed insight into which parameters modulated differentiation behavior. The parameters governing production of FGF4 and stochasticity for Nanog were particularly interesting and would be subsequently modulated in the following section to screen for effects on spatial pattern formation.

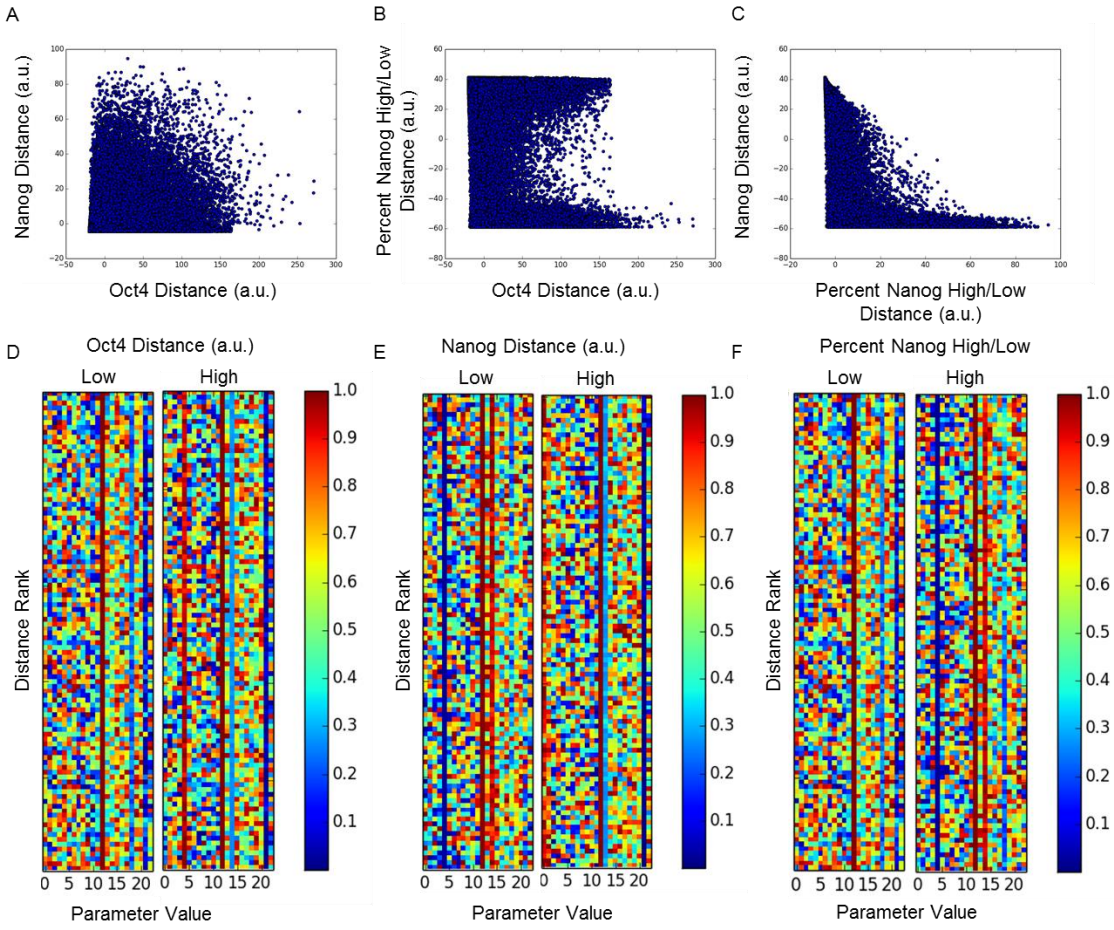

Figure 7-6: Monte Carlo examination of the parameter space. Monte Carlo simulations were run (n = 2E5) to examine the parameter space. The simualtions were quantified using three metrics, the average distance from the Oct4, Nanog and Percent time spent in Nanog High low states. Pairwise plots of parameter sets for Nanaog vs Oct4 (A), Oct4 vs. Percent Nanog High/Low (B) and Percent Nanog High/Low vs. Nanog (C). (D) The top 100 parameter sets for Oct4 showing parameters which led to decreased (left) or increased (right) signal. (E) The top 100 parameter sets for Oct4 showing parameters which led to decreased (left) or increased (right) signal. (F) The top 100 parameter sets for Nanog High/Low residence showing parameters which lead to decrease (left) or increase (right) in signal. A low parameter vlaue is indicated in blue, while a high value is indicated in red.

7.3.3 Modeling the Role of Nanog in ESC Spatial Differentiation

To assess spatial patterns associated with Nanog and Oct4, the stochastic model described was implemented in the agent based aggregate model. Based on the sensitivity analysis and Monte Carlo results, the parameter governing FGF4 production (p21) was a

definite candidate for modulating differentiation behavior. This was consistent with the known role of FGF4 in modulating Nanog levels in ESCs; increased FGF4 production leads to decreased Nanog levels, thus increasing differentiation, while decreasing FGF4 production leads to increased Nanog levels resulting in little differentiation. Interestingly, while the Monte Carlo results did not predict Nanog stochasticity would play a large role in ultimate differentiation decisions, the hypothesis based on the literature was that this parameter would be important in modulating spatial differentiation patterns. However, the Monte Carlo screen was designed to pick out parameters which maximally increased or decreased Oct4/Nanog expression, while the stochasticity parameter served as a modulator between the differentiation permissive Nanog low state and the differentiation resistant Nanog high state, meaning it would be difficult to pick up the effects of stochasticity via Monte Carlo approaches.

To assess the effects of intrinsic noise, versus the spatial heterogeneity induced by FGF4, the parameters governing the stochasticity and strength of FGF4 inhibition on Nanog were varied (Figure 7-7 A). With increasing FGF4 dependence, the pattern of Nanog differentiation was largely spatially homogenous, resulting in largely outside in differentiation (Figure 7-7 B). Increasing the patterns associated with Nanog stochasticity induced a more random looking spatial pattern (Figure 7-7 C) and, in contrast to the Monte Carlo results, Nanog stochasticity played a large role in modulating the systems spatial behavior. Increasing both the stochastic and the FGF4 production parameters resulted in a fast differentiation trajectory which had transient spatial patterns (Figure 7-7 A). These results suggested that modulating both parameters not only tuned the spatial patterns associated with the system, but also the differentiation kinetics.

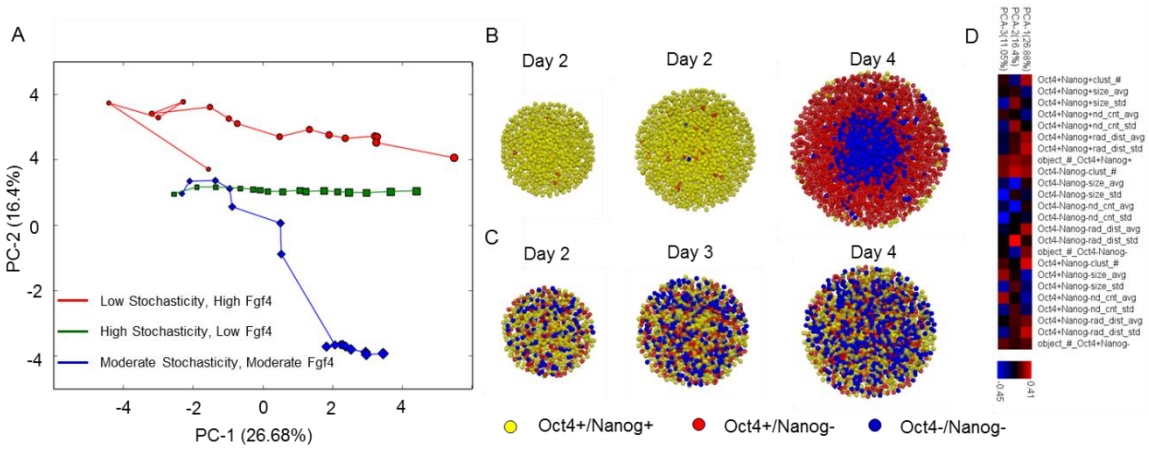

Figure 7-7: Multiscale modeling of Nanog fluctuations in ESC aggregates. (A) Representative simulation traces for low stochasticity high FGF4 production (red), high stochasticity low FGF4 production (green) and moderate stochasticity and moderate FGF4 production (blue). (B) Representative simulation images for low stochasticity, high FGF4 case. (C) Represenative simulation images for the high stochasticity low FGF4 case. (D) Heatmap denoting the network based metric contributions to each of the principal components. Red indicates a positive contribution, while blue indicates a negative contribution.

Under the parameters tested, no outside-in differentiation was observed, in contrast to the previous ABM model (section 5.3). In the ABM model, the outside-in component of differentiation was modulated by LIF signaling, which suggested that the parameter space tested here did not properly sample parameters which modulated LIF production.

7.3.4 Nanog Spatial Pattern Assessment during Differentiation

To assess the model predictions about the spatial distributions of Nanog, aggregates were cultured over a 10 day period, with samples taken at days 1, 3, 4, 5, 6, 7, 9, and 10 for confocal imaging and subsequent network analysis. First, Nanog spatial distributions were assessed during this process. From the previous digital flow results (Figure 7-2) it was expected that Nanog would transfer between two states. After network analysis, a principal component analysis (PCA) of the metric space resulted in three separate groups of Nanog expression (Figure 7-8). These groups correlated roughly with high relatively

homogenous expression of Nanog (Day 1, Figure 7-8 A bottom left), to more heterogeneous expression (day 4 – 7, Figure 7-8 A, middle right), and finally a population which displayed radially distinct Nanog populations (day 8 – 10, Figure 7-8 A, upper left). Oct4 patterns matched the same type of states measured previously, albeit with slightly different kinetic in this case (Figure 7-8 C). Also, there appeared to be more inside-out differentiation than observed previously in Figure 5-3, suggesting that perhaps a different mechanism of differentiation could be occurring.

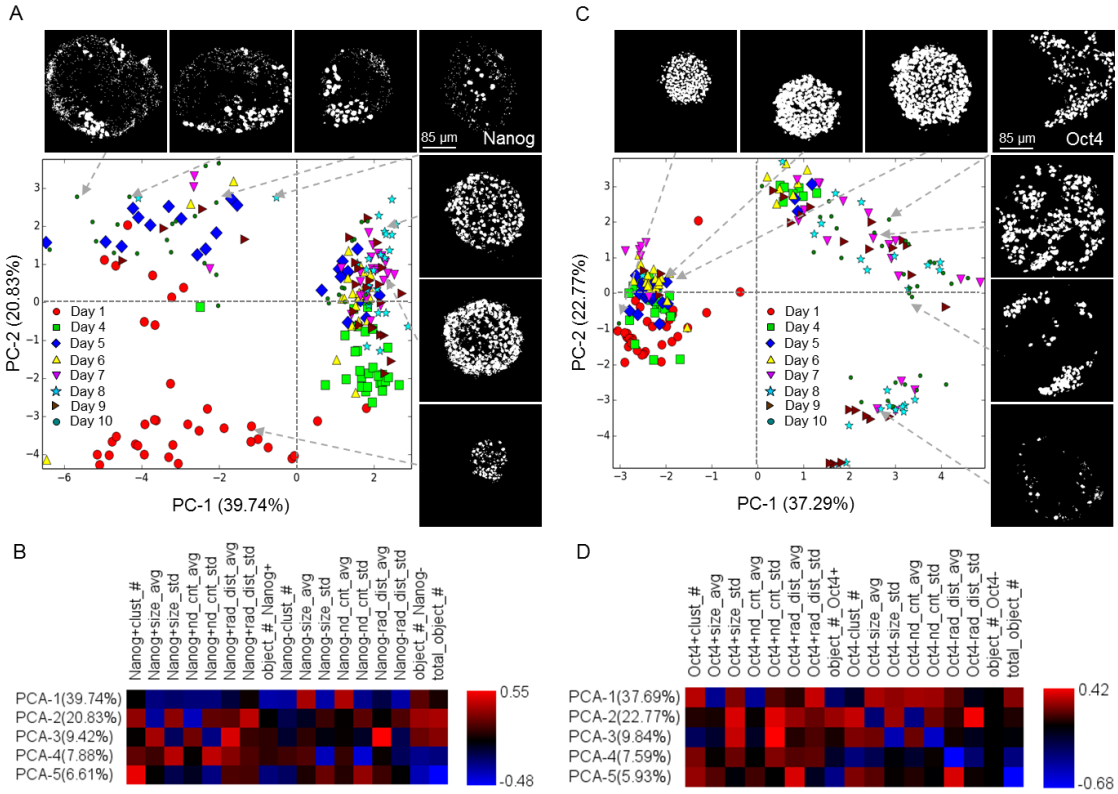

Figure 7-8: Oct4 and Nanog trajectories during the loss of pluripotency transition in ESC aggregates. (A) Principal component analysis (PCA) of patterns of Nanog expression with representative images. (B) Heatmap showing the metric weights for each different principal component, where blue represents a negative correlation, and red a positive correlation. (C) PCA of patterns of Oct4 expression with representative images. (D) Heatmap showing the metric weights for each different prinicpal component, where blue represents a negative correlation, and red a positive correlation. All images are

thresholded with the same value to convert to a black white scale. All scale bars are 85 μm.

Next the Oct4 and Nanog spatial networks were combined to create a set of three different cell types for analysis: Oct4+/Nanog+, Oct4+/Nanog-, and Oct4-/Nanog-. In this case a clear trajectory from early to late differentiation was visualized via PCA (Figure 7-9 A). Once again, though day 1 appeared to have a lot of heterogeneity along PC-1 and PC-2, it segregated nicely from the rest of the data set along PC-3 and correlated well with markers of Oct4+/Nanog+ size and count, suggesting it was a more enriched Nanog+ population than the other days (Figure 7-9 B). This is consistent with previously measured digital flow (Figure 7-2), and representative images (Figure 7-9 C). Analysis of the Oct4+/Nanog+ populations with respect to Oct4+/Nanog- populations showed that these cells were almost always surrounded Oct4+/Nanog- cell types, suggesting that a Oct4+/Nanog+ cell type precedes the Oct4+/Nanog- , and also suggested that these rare cells are often surrounded by more differentiated progeny. All in all this data suggested that network analysis captured Nanog/Oct4 transitions during ESC differentiation. Furthermore, it also showed that Nanog transitions occur before subsequent loss of Oct4, and that these two spatial patterns may be related.

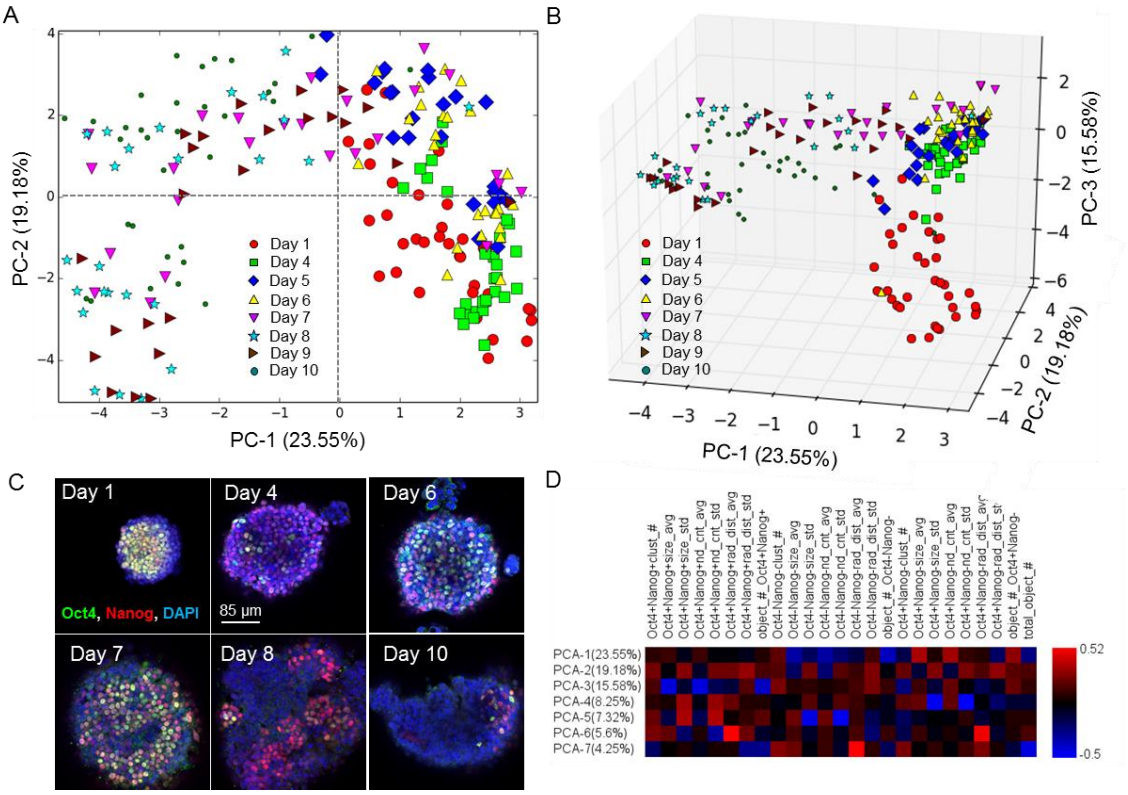

Figure 7-9: Combined Oct4 Nanog pattern trajecotory analysis. Principal component analysis (PCA) of the combined Oct4+Nanog+, Oct4+Nanog- and Oct4-Nanog- metrics spaces in 2D (A) and 3D (B) to show the seperation of day 1 from the rest of the data cloud. Representative images showing the evolution of Oct4 (red) and Nanog (green) spatial patterns over the 10 day differentiation period (C). Scale bar is 85 µm and DAPI is shown in blue. Overall the model needed six components to capture ~80% of the variance (D).

While the patterns of differentiation observed here occur in an inside out fashion this was expected based on the multiscale modeling predictions. To directly compare the models with experimental data, PCA was performed on the combined data set. Representative plots of the selected simulations (Figure 7-7) were depicted on the Oct4 spatial pattern axis (Figure 7-10). The extreme cases, low stochasticity high FGF4, and high stochasticity low FGF4, did not appear to explain the spatial pattern evolution observed experimentally. In fact both simulations appeared to truncate early in the Oct4 pattern space. Only the moderate stochasticity, moderate FGF4 system produced spatial patterns

which matched the pattern progression observed experimentally. This result suggested that both FGF4 signaling and stochasticity needed to be present in moderate amounts in the system to capture the biologically observed pattern evolution.

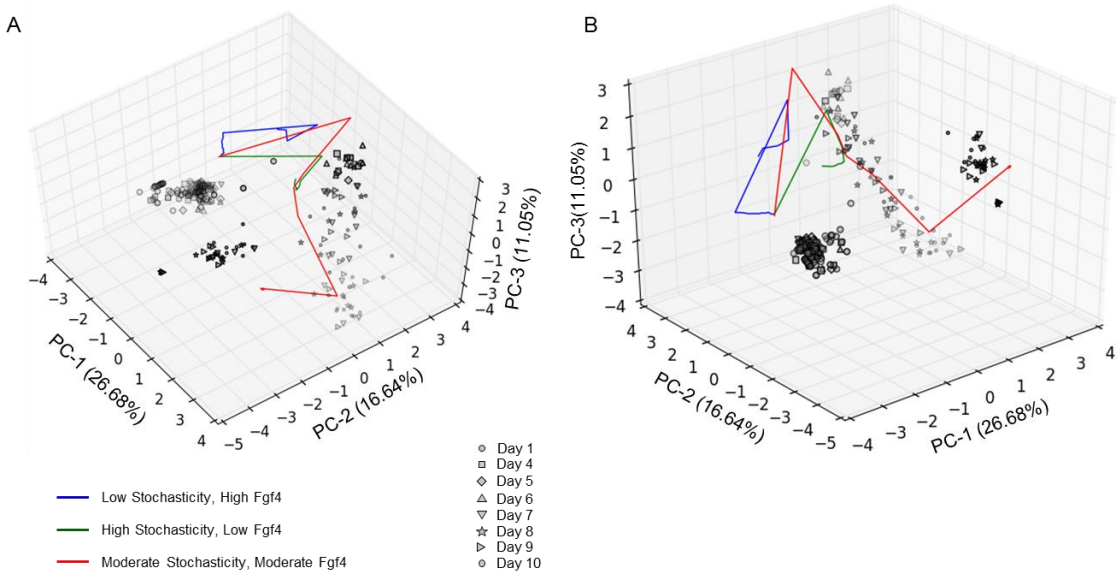

Figure 7-10: Comparisons of computational and experimental spatial Oct4 data. Principal component analysis (PCA) was performed on the combined computational and experimental data sets. The resulting PCA model captured 54.37% of the variance. (A,B) Representative views of the 3D PCA projection of the Oct4 experimental (shaded shapes) and computational (red, blue and green lines) data.

7.4 Discussion

Nanog displays stochastic fluctuations during embryonic stem cell (ESC) culture in a variety of different culture conditions [68, 102, 106]. This stochasticity is largely though to be inherent to the Nanog transcriptional machinery, however, it is also in part due to spatial heterogeneity in FGF4 signaling [106, 109, 113]. However, the extent to which Nanog stochasticity and spatial FGF4 gradients affect ESC differentiation has been largely unexplored. This work probed the roles of Nanog stochasticity and FGF4 signaling using a computational modeling approach.

To establish the spatial pattern associated with Nanog and Oct4 differentiation, network analysis was performed on ESC aggregates. This resulted in the quantification of Nanog spatial patterns which moved from an originally homogenous state, through a random internal state, to a final outside in state in which clusters of Nanog+ cells were seen primarily around the outside of the aggregate. The patterns of Oct4 differentiation observed in this study matched those observed in earlier work (Figure 5-3), although the kinetics and end points of differentiation appear to be slightly delayed. Furthermore, the day 10 samples displayed a striking morphology in which Oct4+ clusters are primarily located on the outside on the aggregates, which was in contrast to the earlier clusters of Oct4+ cells observed in the center of aggregates. The ABM model described in chapter 5 predicted that this could be due to an increased aggregate size (where FGF4 signaling would dominate differentiation) which is consistent with the stochastic multiscale model implemented here, but it is also possible that staining or imaging artifacts could cause such radial patterns to emerge. To validate this, staining on histological sections could be performed to indicate whether these patterns are truly radial in nature.

This work shows that a multiscale modeling approach was useful for explaining relevant experimentally observed spatial phenomena. Though this work was not the first multiscale aggregate model to be constructed, it highlighted nicely how synergisms between network based signaling models, and agent based models can be exploited to answer questions about biological pattern formation. Sensitivity analysis revealed a family of parameters which were expected to drastically modulate differentiation kinetics. Of the parameters tested it was shown that factors regulating FGF4 production as well as inherent Nanog stochasticity were able to modulate pattern formation.

Furthermore, our model predicts spontaneous fluctuations of Nanog will arise in ESC aggregates, which matches reports in the literature [68, 113, 294]. One way to validate this result would be to use live imaging of aggregates with a Nanog reporter line to examine if stochastic fluctuations occur between the Nanog high and Nanog low state during differentiation.

The model described here represents an amalgamation of models from many sources. Interestingly, incorporation of the "negative repressor gene" hypothesis was crucial to explain the observed Oct4 on-off transition kinetics. In this model the gene took the form of Gata6 (a well-known antagonist to Oct4, Sox2 and Nanog) which specifies early primitive endoderm in the early mammalian blastocyst [113, 296]. Nanog and Gata6 actively inhibit each other's expression, making this one of the first bi-stable toggle switches observed during development. The role of Gata6 was not explicitly explored in this study even though it was found to be crucial in modulating Oct4 dynamics, therefore a subsequent study examining Gata6 dynamics is necessary. Furthermore, this model explicitly included the autocrine effects of LIF, known to promote pluripotency in mouse embryonic stem cells via Jak/STAT signaling, whereas previous models modulated differentiation using an artificial input. It was also important to include cross talk between the FGF4 MEK/ERK pathways and the LIF JAK/STAT pathway in order to accurately capture differentiation kinetics.

While digital flow cytometry is a well validated technique, the author notes that it is not a direct replacement for flow cytometery. However, technical limitations were encountered isolating cells for flow cytometry from cellular aggregates. It was extremely difficult to get enough cells out of the aggregates to quantify, and the subsequent antibody staining and wash steps reduced the cell number substantially. Rather than try to optimize this part of the protocol, the established digital flow technique was examined as an alternative for quantifying individual cell morphologies. Furthermore, this protocol

was a natural extension of the network based framework, as annotated cell information in the form of normalized intensities was already generated during the network digitization process. This approach was also ultimately useful for identifying intensity threshold cutoffs for subsequent digitization of networks, which was a necessary step for spatial pattern evaluation.

Taken together this work significantly advanced the understanding of how Nanog regulates pluripotency in the context of 3D multicellular aggregates. This hypothesis of this study was that Nanog regulation could be responsible for the observed Oct4 spatial patterns. A multiscale modeling approach implicated that Nanog regulation via inherent stochasticity and FGF4 signaling was crucial for explaining the Oct4 patterns observed in ESC aggregates. These studies showed inherent noise in Nanog transcriptional machinery, as well as the production of FGF4, and FGF4 regulation on Nanog transcription levels modulated complex spatial pattern phenotypes. Interestingly, this suggested that if the stochasticity associated with Nanog fluctuations could be uncoupled from FGF4 signaling and independently modulated, radically different types of pattern formation could result. However, due to the intrinsic link between FGF4 and Nanog stochasticity, it would be difficult to design an experiment which would completely decouple these two processes. In summary this work outlines a novel strategy for multiscale modeling of complex genetic regulatory networks in conjunction with soluble cues to examine spatial patterns associated with ESC aggregate differentiation.

CHAPTER 8 CONCLUSIONS AND FUTURE DIRECTIONS

8.1 Network Analysis

Network analysis is typically utilized to examine large interaction networks, ranging from social networks, to biological interaction networks. Various types of fields make use of network analysis to extract features form networks, such as shortest paths, clichés, and motif analysis. In the case of biological networks, these methods are employed to identify important species within these interactomes, or to identify enriched regulatory subnetworks. In all cases these methods involve extracting some information form the network topology and using it to describe the network. In this work the basics of this network approach were extended to cover spatial dimensions. In this case a series of network descriptors were extracted from the network which describe spatial information and distributions of markers of various types.

The author borrowed the idea of identifying interesting subnetworks and their enrichment, instead opting to identify subnetworks from local cell phenotypes. Rather than simply reporting the numbers of such subnetworks, spatial information about the location of each of these subnetworks within the larger network, and also, their relationships to each other is extracted. The result is the creation of a novel modular spatial pattern recognition approach which allows extraction of a series of meaningful spatial pattern metrics as well as metrics related to the number of cells in a certain state across multiple different systems of interest. These metrics cluster broadly into two categories, those related solely to spatial information, and those related to relative number of cell/cell types present. Decoupling these two classes of metrics from each other in order to describe spatial pattern formation is a difficult task, and this work highlights that often substantial portion of spatial pattern heterogeneity is inherently tied to differences in cell types present.

This study tests the validity of this approach across a number of different model systems: ESC cellular aggregates, computational models, and Cichlid fish. A variety of different biological processes were tested: early loss of pluripotency, late mesenchymal like differentiation, neural differentiation, and gastrulation. The spatial scales assessed cover a wide variety of scales ranging from tens of microns (early ESC aggregates) up to hundreds of microns (late gastrulation). Due to restrictions with imaging, in the context of the networks measured this equates to anywhere from tens to thousands of network nodes (individual cells) with thousands to hundreds of thousands of network edges (or cellular interactions). Furthermore, in each biological system this work uncovered novel biological insight.

For loss of pluripotency in ESC aggregates this method uncovered a complex paracrine signaling mechanism which explained spatial patterns of differentiation as well as observed differences kinetics of ESC aggregates of differing sizes (Chapter 5 and 7). The two paracrine process proposed here explained these differences (one in a secreted factor is responsible for maintain pluripotency, and the other where more differentiated cells secrete a factor which induces differentiation), and mirrored the known properties of soluble LIF and FGF4 signaling respectively [102, 106, 110, 290]. Surprisingly but interestingly, the lack of local neighbor-to-neighbor regulation of phenotypic state, as analyzed by this methodology, suggested that transmission of cell state information by intercellular cues, such as Notch, may impact later stages of differentiation than the time period examined here. Though one other report of comparing image derived metrics to computational metrics does exist (see section 5.4), this work represented the first direct quantitative comparison between computational modeling and complex emergent spatiotemporal patterns during multicellular lineage commitment in 3D ESC aggregates.

Analysis in cichlid was quite a difficult challenge (Section 6.2). The images to be analyzed were 3D confocal images, and much larger than anything previously analyzed. Furthermore, the readout of BMP signaling via measurement of pSmad, represented a

gradient of staining in response to the soluble gradient, which made creating a binary annotation out of this signal challenging. Furthermore the spatial pattern domains of interest were often one single subnetwork which changed shape over time. This represented an extreme from the ESC cases discussed previously, were many small subnetworks could be measured in a single aggregate. Nevertheless this approach proved sensitive enough to separate different cichlid fish embryos based on their staging during gastrulation, and identify key the key temporal changes in dlx3b gene expression and BMP signaling responsible for driving this event.

In the context of neural differentiation this approach was able to shed some light on the glial switch (Section 6.3). Here again ESC aggregates were used as a model system in order to try and understand the regulation of the switch form production of motor neurons to oligodendrocytes (glial cells). Using a hybrid approach of computational models with digital flow cytometry methods the study showed that there was no way a rigid differentiation schema was capable of explaining the robust motor neuron induction observed in this system, or the fact that these motor neurons appeared before the establishment of a supposed neural progenitor population. The model postulated that during early differentiation, this progenitor population is not stable, and it was not until a substantial portion of motor neurons are established in the culture that these neural progenitors can establish themselves in this culture system. Furthermore, the models predicted that this signal was likely a soluble factor, providing evidence that this method was capable of distinguishing between soluble and local cell-cell modes of regulation. This highlighted the scalability of this approach for analyzing multiple different cell types of interest and the spatial pattern associated with each. The predictions of this model are being validated in collaboration with Raymond Swetenburg in Dr. Stice's lab at the University of Georgia.

Analysis of histological data samples also provided a variety of new challenges (Section 6.4). First off, single cell markers are often not available in histological samples.

In this work network derived metrics were used to inform single cell classification efforts. This study showed that network based classification improved the performance of shape based classification methods substantially. Furthermore, this study highlighted a comparison between two entirely different biological systems which both induced epithelial to mesenchymal like transitions (EMT): treatment of ESC aggregate with BMP4, and gelatin methacrylate microparticles. The results showed that the microparticles appeared to induce slightly more pronounced EMT than the BMP4, but that the overall distributions and spatial patterns of this process were strikingly similar. This underscored the portability of this approach and its use in comparing experiments across different systems in a meaningful straightforward manner. The method also identified other structures in ESC aggregates: cysts and super-dense epithelial areas. Interestingly the method was unable to distinguish rosette like structures robustly, suggesting that further work is needed in validating and training those classifiers.

In the future this approach could be applied to various single cell live imaging systems. A variety of such systems exist, and data from whole embryos undergoing development are already starting to become available form light sheet experiments. Many of these systems report robust systems for tracking, and identification of single cells during the process, however, as of yet, methods for quantifying the spatial evolution of these processes have not been explored. Our method synergizes well for these applications, as all of the data is inherently single cell and can be expressed in a network format. This allows extraction of similar spatial network metrics described here, but with the added advantage of following a single aggregate through time, which abolishes many of the issues with the subsequent data analysis techniques described in Section 7.3.

The multiple examples illustrated herein highlight the broad utility of network-based analysis for identification of spatial biological patterns via the formulation of novel metrics. This study reports the derivation of pattern trajectories associated with several systems: experimentally observed loss of Oct4 in ESC aggregates, computational models

of Oct4 loss in ESC aggregates, and gastrulation in cichlids. Novel biological insights gained using our network analysis approach included: 1) differences in ESC aggregate spatiotemporal pattern kinetics can be explained by a combined paracrine signaling methodology, 2) gastrulation in cichlid fishes can be split into a set of discrete stages and 3) neural progenitors undergo a spontaneous differentiation towards motor neurons, and motor neurons must actively inhibit this process via a soluble factor to drive glial differentiation. In the case of ESCs, a large body of literature exists which suggests that differentiation is heavily modulated by ESC size [112, 142, 151-153]. In summation, this novel pattern classification pipeline permits entirely new forms of quantitative analysis based upon the fundamental interconnectivity of multicellular networks, which could revolutionize the characterization of biologically complex spatiotemporal phenomena.

8.2 Computational Multiscale Modeling of Tissues

Biological systems present many interesting processes on multiple different spatial and temporal scales. These range from protein-protein molecular interactions, to cell-cell interactions, all the way up to tissue and organism level behaviors. Depending on the level at which the process of interest resides, various different methods can be used to simulate it. In the context of simulating dynamics in cellular aggregates or early developmental systems, a variety of biological processes need to be taken into account. Cells must be allowed to proliferate, move, and interact with their environment. During development, cells are heavily influenced by soluble queues, which need to be simulated over complex spatial domains. This can be accomplished utilizing forward Euler integration schemas to solve complex reaction diffusion processes, though the method is computationally intensive as it requires high spatial resolution (10-15 μm ~ one cell diameter) and complete integration through time. Even though these spatial soluble gradients can be simulated, cell responses to these cues are often not linear, and therefore the intracellular signaling dynamics are also important to consider. Furthermore,

depending on the stage or process simulated, it is necessary to consider single cell responses to account for stochastic fluctuations and spatial heterogeneity[297-299]. Accounting for all of these properties leads to inherent multiscale models in which the specific signaling pathways governing cell responses to these signals, the location and behavior of single cells and the spatial simulation of the soluble signals is crucial.

Agent based modeling (ABM) is one such technique which allows simulation of each individual cell as its own entity. Classically agent based models have been used to study population dynamics, ecological niches, dynamic markets. The flexible framework for allowing entities to interact based on simple sets of rules is useful in a variety of contexts. In this work, ABM approaches are utilized to model cells in multicellular aggregates. The individual cells (or agents) are allowed to make decisions about movement, differentiation, and vision based on a simple set of rules. Differentiation was assumed to be governed by two sets of queues: soluble, and local cell-cell interactions. Examples of biologically relevant signals at play during early ESC loss of pluripotency are LIF, FGF4 and BMP, while relevant local signaling factors include Notch. While soluble factors have been implicated during early differentiation processes (FGF4, BMP, and Wnt), local cell-cell interactions such as Notch are typically relegated to maintenance of adult stem cell populations.

Division and other physical interactions were accounted for by simulating the group of cells as a mass spring network. This offered a variety of advantages over FE or continuum based approaches. First, the cells were maintained in a network structure, which made finding cell-cell interactions an inherently easy task. Second, it enabled rapid extraction of network based metrics from these physical networks. Thus the network structures used in this model serve largely two functions, the ability to maintain physical interactions, and keeping track of local cell-cell interactions.

ABM approaches generally suffer from a lack of quantitation. Even though simulation parameters can be tweaked until they match some experimental system,

quantitative extraction of metrics governing the evolution of these systems is quite difficult. Furthermore, due to the increased complexity in these systems many of the tools that were useful at the network level become much more difficult to implement. Sensitivity analysis and parameter fitting via optimization strategies become increasingly more computationally intensive as the time to run each simulation increases. Monte Carlo based approaches are out of the question due to the parameter space necessary and the time to run each simulation. Therefore to optimize parameters, a grid search method was used with only a sparse sampling of each parameter. This approach is not guaranteed to find an optimal solution, but rather it provides an idea of the landscape of behaviors the ABM models can provide. Using such approaches relevant ranges of behaviors for cells communicating to each other via both local and soluble paradigms were defined.

Interestingly, work in Chapter 3 showed that local cell–cell interactions could capture spatial pattern associated with differentiation. In later diffusion based simulations, depending on the ratio of parameters governing consumption, production and diffusion of the factor, gradients of various spatial dimensions could be achieved. Thus, it was possible to have both local paracrine like factor secretion, and also more global factor effects all stemming from secretion of soluble factors. It is interesting to note that the network based approaches could distinguish these different signaling regimes, particularly in the context of neural differentiation (Section 6.3). However, in the neural work, the soluble factors were functioning on am more global scale, making the distinction between local interactions and this mode of signaling quite stark. In the case of the early ESC pluripotency, soluble factors were only implicated because those simulations lead to differences in differentiation between sizes. When the factors were functioning on a local paracrine scale it was difficult to tell them apart from pattern induced via local cell-cell interactions; however this is largely due to the fact that the two regimes simply lead to similar spatial pattern formation.

While rules based modeling approaches can be useful for elucidating the general rules governing a systems behavior, extrapolating these rules into a set of relevant molecular mechanisms is quite difficult. Therefore, to increase the biological insight of these models, an underlying set of stochastic ODEs was used to drive cell behavior. After fitting this system to experimental data and validating the local parameter sets, sensitivity analysis revealed a set of parameters which were maximally influenced the system. Once this system was implemented in our ABM model, a similar approach was performed to see which of these parameters lead to. It was observed that the parameter governing Nanog stochasticity and FGF4 secretion had the most effect on the spatial pattern associated with this differentiation process. Thus this study could link the abstract parameters used in the rules based simulations back to particular molecular mechanisms which could be responsible for regulating spatial pattern development and evolution in ESC aggregates. In the future, it would be extremely interesting the modulate parameters governing FGF4 signaling possibly via application of inhibitors or via gene editing approaches. Unfortunately it is difficult at the current time to modulate Nanog stochasticity directly to test its effects on ESC differentiation. Some work has been done showing that through titration of different small molecules the noise associated with Nanog transcription can be increased, however, many of these are directly tied to FGF4 signaling, so decoupling these two processes would be extremely difficult.

It is also important to note that the influence of the extracellular matrix environment is not taken into account in these simulations. While for early loss of pluripotency transitions, there is some evidence that matrix plays an important role in cell fate decisions[13, 18, 230], it is certainly an important determine factor in tissue maturation, EMT, and other extremely relevant processes to tissue development [12, 13, 16-18, 56, 230]. In the future such interactions could be incorporated into the modeling framework in a variety of ways. Small agents could be secreted form the cells representing matrix molecules. This would allow for propagation of physical interactions

in a similar fashion to those described above for cells. However, though extremely accurate, this approach would likely lead to a dramatic increase in computational time, making it only practical if this resolution is absolutely required. An alternative approach which has enjoyed some popularity in the simulation community is to treat the ECM as "gradient" which though it is not subject to diffusive conditions, can be degraded and deposited in a similar manner to a reaction diffusion system. This results in a robust modeling approach for simulating deposition and remodeling of matrix associated species, which would also work well in the context of the current diffusive framework.

8.3 Comparisons of Biological and Computational Data

In the course of creating mathematical models of systems, access to biological data for initial fitting and subsequent testing of the model is crucial. In the case of classical computational modeling of signaling systems utilizing ordinary differential equations (ODEs), this involves capturing active signaling dynamics over time typically through the use of antibodies recognizing active signaling proteins via either flow cytometry or western blot. Or in the case where reporter systems are present, live real-time readouts of the system in response to various inputs can be acquired. In either case subsequent comparisons between the model and experimental data can be performed.

A variety of methods exist for optimizing parameters in mathematical models to fit to experimental data. An entire field of optimization theory has been utilized to derive a whole set of optimization algorithms including but not limited to: conjugate gradient descent, Powell, Broyden Fletcher GoldFarb Shanno (BFGS), Newton conjugate gradient, and downhill simplex algorithms. Each of these algorithms has strengths and weakness depending on the size of the optimization problem. Most of these methods work reasonably well assuming the problem is formulated in a linear fashion however, in

the case where stochastic noise is added to the system, this can be a bit more difficult to solve. Thus, for the case of the stochastic ODE simulations used in this work, a Monte Carlo based approach was used to sample parameter space to judge the viability of the current parameter set. This worked reasonably well as a substitute for the optimization algorithms, and gave a passable picture of the shape/size of the parameter space.

In general, optimization approaches works quite well for well-defined systems of parameters where the output and objective function for comparison is well known; however, methods for comparing biological images to experimental simulations are still lacking. This work derives a common set of mathematical descriptors via network analysis which can be used in this fitting. Once these metrics are described they can be compared between any set of systems. However, this type of data presents several challenges. Many metrics are created for each pattern quantified leading to a high dimensional data set, and many different types of patterns are quantified at each time point which further complicates the process by creating multi-dimensional data clouds. For any large dimensional data set extraction of important features is crucial to understanding the data dynamics and accurately comparing the results. Simple Euclidean distance metric comparisons between high dimensional vectors fail to capture subtle differences in the data due to assigning even weights to all data for comparison. This study addressed this problem by using feature extraction method to identify important metrics in the data set, or in the case of principal component analysis (PCA) combinations of metrics in the data set which capture the most variance. Once the data was transformed into the space emphasizing important features, Euclidean distance metrics can be used to compare the data clouds.

However comparison between point n-dimensional clouds is still a complicated problem. A simple point by point comparison method works well for establishing distances between point clouds assuming that the clouds display some sort of uniform distribution. Depending on the system of interest, the point clouds may have separate subpopulations of behaviors with in each group for comparison. This work addresses this problem utilizing a novel network based approach which utilizes a KDTree combined with K-means clustering to split irregularly shaped data clouds up. One weakness of this approach is that it currently does not implement an automatic mode for selecting or testing if multiple clusters exist, the difference threshold between clusters must be specified. This threshold factor is equivalent in some sense to the p-value used in statistical tests. Incorporation of an unsupervised K-means approach for testing clusters would allow a more automated approach to cluster identification. It is important to note that this method is similar to Gaussian mixture models (GMM), however, fitting GMM models requires many points in the data cloud, where the approach outlined here works well for sparse data clouds. This approach works reasonably well for the data tested, and really only proved necessary in the cases where time points displayed strong bimodal or trimodal segregation into different states. However, further testing is needed to perfect this algorithm and rigorously test its ability to accurately estimate distances between point clouds in a reliable fashion. Additionally, a method for assessing whether these extracted differences are statistically different form each other is crucial for assessing the validity of this method.

Once the pattern data is processed, the comparison between biological image data and simulation data can be achieved. Subsequent comparisons of the resultant data clouds

moving through time can be accomplished through a variety of ways. Time can be non-dimensionalized, which results in computing the closest distance in a pairwise method for all clouds of data belonging to each process. Or, time can be enforced, and comparisons between data clouds are only enforced for given time points. In most of the analysis performed in this study time was enforced for such comparisons, but for the purposes of establishing if a simulation follows a different trend, simply shifted in the time, these methods were not sufficient. To answer this problem, the non-dimensionalized time comparisons were performed. All in all, this works highlights a strategy for comparing many sets of biological images to computational simulations in a tractable way. The strategies outlined herein represent a simple solution for comparing multidimensional data and could be applied to the comparisons of any such data sets.

8.4 Regulation of Pluripotency in Embryonic Stem Cells

The roles of Oct, Sox2 and Nanog in maintain pluripotency are well established. Oct4 and Sox2 form a transcriptionally activated complex which in turn regulates a myriad of downstream genes, including Oct4 and Sox2 [72, 77, 78, 125]. Nanog, has also been shown to maintain pluripotency by repressing a series of genes associated with differentiation, most notably Gata6 [113]. Nanog activates itself in a positive feedback loop in addition to activating expression of Oct4 and Sox2, while the Oct4-Sox2 complex activates expression of Nanog. These redundant positive feedback loops maintains ESCs in a pluripotent state. Classically in conditions favoring differentiation it is observed that Nanog is lost first, and Sox2 and Oct34 follow soon after; thus Nanog is often thought of as a master regulator of differentiation. However, in serum containing culture with LIF, stochastic fluctuations in Nanog levels have been observed [78, 79, 300]. Indeed many have observed that these fluctuations appear stochastic in 2D culture, where Nanog appears to switch back and forth between two cell fates. Until recently this process was

thought to be largely random, and either a product of intrinsic cellular stochasticity associated with Nanog, or perhaps diverse epigenetic regulation. A recent paper put forth by Ochiai *et al.* showed via a CRSIPR reporter system that the intrinsic and extrinsic noise have relatively equal contributions to Nanog heterogeneity in culture [109]. It is also well known that Nanog can be regulated by FGF signaling thorough MAPK pathways. This regulation is abolished in 2i culture, and leads to homogeneous Nanog expression. There are also scatter reports that ESC aggregates cultured in LIF and 2i appear to have a spatially random differentiation patterns, although this has yet to be confirmed with any quantitative methods.

This work highlights the existence of spatial patterns in Oct4 during the loss of pluripotency in ESC aggregates. These patterns co-vary with Sox2 expression, which agrees with the known roles of Oct4 and Sox2 in early differentiation steps. Agent based modeling revealed that loss of Oct4 was likely accomplished with a system containing two soluble factors, one which induces differentiation, and one which prevents it. Based on the literature it was concluded that these species are likely FGF4, which as discussed above is implicated in inducing differentiation via down-regulation of Nanog, and LIF which activates Oct4 and Sox2 directly via activation of Stat3 (via phosphorylation). In order for this model to accurately explain the results, particularly differentiation induced across different sizes, FGF4 must be a potent inducer of differentiation, and be capable of inducing resistance to LIF secreted by mouse embryonic stem cells. Reports in the literature indicate that ERK/MAK signaling is capable of inhibiting Stat3 phosphorylation, providing a mechanism by which ESCs can ignore LIF induced pluripotency in the presence of enough FGF4. Our model further indicates that early spatial distributions in Nanog are responsible for subsequent patterns of Oct4 in ESC aggregate differentiation. A computational model constructed of this process indicated that the contributions of intrinsic and extrinsic noise in Nanog regulation are approximately equal. While the exact mechanism governing intrinsic noise at the Nanog

allele is unknown, our model predicts that the extrinsic noise can largely be captured by a spatially heterogeneous gradient of FGF. One possible way to test such claims would be to culture ESCs in 2i conditions, dissociate them to make 3D aggregates, and then expose them to the same differentiation conditions. The current model predicts that culture of ESCS in 2i containing conditions prior to differentiation would lead to more spatially random loss of Nanog and subsequent Oct4 patterns.

The model also suggests that Nanog fluctuations between high and low states do occur in ESC aggregates. Though this conclusion is only drawn from results based on the computational model, examining expression of Nanog in ESC aggregates in real time via live imaging and reporter system would address many of these questions. Real time imaging would allow comparisons of live single cell traces with spatial information to our computational models about Nanog fluctuations, while simultaneously allowing examination of a whole host of questions related to heterogeneity and perturbations in the microenvironment via. Ideally, differentiation trajectories would be studied using both a Nanog and a Gata6 reporter line. Gata6 and Nanog inhibit each other's expression, with Gata6 typically being associated with early epiblast cell formation, making the Nanog-Gata6 switch one of the first fate transition in development. Furthermore, this would be an excellent platform to probe Gata6 spatial expression in real time to see if the predicted Gata6 spatial patterns match those predicted by the computational model. This represents an excellent application and extension of our computational modeling and network analysis platforms, and would allow us to further address outstanding questions in the field regarding Nanog expression.

APPENDIX

A.1. Supplementary Methods

A.1.1. Network Based Metrics

A.1.1.1. Loss of Pluripotency Binary Metrics (Oct4+, Oct4-)

Oct4+_clust_# - the number of Oct4 + clusters (a cluster is defined as more than a single node)
Oct4+_size_avg – the average radius of the Oct4+ clusters
Oct4+_size_std – the standard deviation in the radius of the Oct4+ clusters
Oct4+_nd_cnt_avg – the average number of nodes of the Oct4+ clusters
Oct4+_nd_cnt_std – the standard deviation in the number of nodes of the Oct4+ clusters
Oct4+_rad_dist_avg – the average radial distance of the Oct4+ clusters
Oct4+_rad_dist_std – the standard deviation of the radial distances of the Oct4+ clusters
Oct4-_clust_# - the number of Oct4- clusters (a cluster is defined as more than a single node)
Oct4-_size_avg – the average radius of the Oct4- clusters
Oct4-_size_std – the standard deviation in the radius of the Oct4- clusters
Oct4-_nd_cnt_avg – the average number of nodes of the Oct4- clusters
Oct4-_nd_cnt_std – the standard deviation in the number of nodes of the Oct4- clusters
Oct4-_rad_dist_avg – the average radial distance of the Oct4- clusters
Oct4-_rad_dist_std – the standard deviation of the radial distances of the Oct4- clusters
Total_obj_# - the total number of cells in the system
Object_#_ Oct4+ – the total number of Oct4+ cells in the system
Object_#_ Oct4- – the total number of Oct4- cells in the system
Percent_diff – the total number of Oct4- cells / the total number of cells
Agg_radius – the maximal radius (size) of the aggregate as measured from the center

A.1.1.2. Cichlid Gastrulation Multiclass Metrics (dlx3b+, pSmad+, pSmad/dlx3b+)

dlx3b+_clust_# - the number of dlx3b+ clusters
dlx3b+_size_avg – the average radius of the dlx3b+ clusters
dlx3b+_size_std – the standard deviation in the radius of the dlx3b+ clusters
dlx3b+_nd_cnt_avg – the average number of nodes of the dlx3b+ clusters
dlx3b+_nd_cnt_std – the standard deviation in the number of nodes of the dlx3b+ clusters
dlx3b+_rad_dist_avg – the average radial distance of the dlx3b+ clusters
dlx3b+_rad_dist_std – the standard deviation of the radial distances of the dlx3b+ clusters
dlx3b+_clust_circ_avg – the average circularity of the dlx3b+ clusters
dlx3b+_clust_circ_std – the standard deviation of the circularities of the dlx3b+ clusters
dlx3b+_ecc_avg – the average eccentricity of the dlx3b+ clusters
dlx3b+_ecc_std – the standard deviation of the eccentricities of the dlx3b+ clusters

pSmad+_clust_# - the number of pSmad+ clusters
pSmad+_size_avg – the average radius of the pSmad+ clusters
pSmad+_size_std – the standard deviation in the radius of the pSmad+ clusters
pSmad+_nd_cnt_avg – the average number of nodes of the pSmad+ clusters
pSmad+_nd_cnt_std – the standard deviation in the number of nodes of the pSmad+ clusters
pSmad+_rad_dist_avg – the average radial distance of the pSmad+ clusters
pSmad+_rad_dist_std – the standard deviation of the radial distances of the pSmad+ clusters
pSmad+_clust_circ_avg – the average circularity of the pSmad+ clusters
pSmad+_clust_circ_std – the standard deviation of the circularities of the pSmad+ clusters
pSmad+_ecc_avg – the average eccentricity of the pSmad+ clusters
pSmad+_ecc_std – the standard deviation of the eccentricities of the pSmad+ clusters
pSmad+/dlx3b+_clust_# - the number of pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_size_avg – the average radius of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_size_std – the standard deviation in the radius of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_nd_cnt_avg – the average number of nodes of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_nd_cnt_std – the standard deviation in the number of nodes of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_rad_dist_avg – the average radial distance of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_rad_dist_std – the standard deviation of the radial distances of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_clust_circ_avg – the average circularity of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_clust_circ_std – the standard deviation of the circularities of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_ecc_avg – the average eccentricity of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_ecc_std – the standard deviation of the eccentricities of the pSmad+/ dlx3b+ clusters
pSmad+/dlx3b+_r_ratio – r_clust_nd_count_avg / y_clust_nd_count_avg
pSmad+/dlx3b+_g_ratio - r_clust_nd_count_avg / y_clust_nd_count_avg
Total_obj_# - the total number of cells in the system
Object_#_pSmad+/dlx3b+ – the total number of pSmad+/dlx3b+ cells in the system
Object_#_dlx3b+ – the total number of dlx3b+ cells in the system
Object_#_ pSmad+ – the total number of pSmad+ cells in the system
Agg_radius – the maximal radius (size) of the aggregate as measured from the center

A.1.1.3. Mesenchymal Morphogenesis Binary Metrics (Mesenchymal, Epithelial)

M_clust_# - the number of mesenchymal clusters
M_size_avg – the average radius of the mesenchymal clusters
M_size_std – the standard deviation in the radius of the mesenchymal clusters

M_nd_cnt_avg – the average number of nodes of the mesenchymal clusters
M_nd_cnt_std – the standard deviation in the number of nodes of the mesenchymal clusters
M_rad_dist_avg – the average radial distance of the mesenchymal clusters
M_rad_dist_std – the standard deviation of the radial distances of the mesenchymal clusters
E_clust_# - the number of epithelial clusters
E_size_avg – the average radius of the epithelial clusters
E_size_std – the standard deviation in the radius of the epithelial clusters
E_nd_cnt_avg – the average number of nodes of the epithelial clusters
E_nd_cnt_std – the standard deviation in the number of nodes of the epithelial clusters
E_rad_dist_avg – the average radial distance of the epithelial clusters
E_rad_dist_std – the standard deviation of the radial distances of the epithelial clusters
Total_obj_# - the total number of cells in the system
Object_#_E – the total number of mesenchymal cells in the system
Object_#_M – the total number of epithelial cells in the system
Percent_diff – the total number of epithelial cells / the total number of cells
Agg_radius – the maximal radius (size) of the aggregate as measured from the center

A.1.1.4. Neuronal Multiclass Metrics (Stem Cell, Olig2+, HB9+, Olig2+/Nkx2.2+)

U_# - the number of unmarked clusters
U_size_avg – the average radius of the unmarked clusters
U_size_std – the standard deviation in the radius of the unmarked clusters
U_nd_cnt_avg – the average number of nodes of the unmarked clusters
U_nd_cnt_std – the standard deviation in the number of nodes of the unmarked clusters
U_rad_dist_avg – the average radial distance of the unmarked clusters
U_rad_dist_std – the standard deviation of the radial distances of the unmarked clusters
NP_# - the number of progenitor clusters
NP_size_avg – the average radius of the progenitor clusters
NP_size_std – the standard deviation in the radius of the progenitor clusters
NP_nd_cnt_avg – the average number of nodes of the progenitor clusters
NP_nd_cnt_std – the standard deviation in the number of nodes of the progenitor clusters
NP_rad_dist_avg – the average radial distance of the progenitor clusters
NP_rad_dist_std – the standard deviation of the radial distances of the progenitor clusters
MN_clust_# - the number of motor neuron clusters
MN_size_avg – the average radius of the motor neuron clusters
MN_size_std – the standard deviation in the radius of the motor neuron clusters
MN_nd_cnt_avg – the average number of nodes of the motor neuron clusters
MN_nd_cnt_std – the standard deviation in the number of nodes of the motor neuron clusters
MN_rad_dist_avg – the average radial distance of the motor neuron clusters
MN_rad_dist_std – the standard deviation of the radial distances of the motor neuron clusters
G_clust_# - the number of glial clusters

G_size_avg – the average radius of the glial clusters
G_size_std – the standard deviation in the radius of the glial clusters
G_nd_cnt_avg – the average number of nodes of the glial clusters
G_nd_cnt_std – the standard deviation in the number of nodes of the glial clusters
G_rad_dist_avg – the average radial distance of the glial clusters
G_rad_dist_std – the standard deviation of the radial distances of the glial clusters
Total_obj_# - the total number of cells in the system
Object_#_ MN – the total number of motor neurons in the system
Object_#_NP – the total number of progenitor cells in the system
Object_#_G – the total number of glial cells in the system
Object_#_U – the total number of unmarked cells in the system
Agg_radius – the maximal radius (size) of the aggregate as measured from the center

A.1.2. Classification - Grid Search Parameters (Python Code)

A.1.2.1. NUSVC

```
nus = [1E-7, 1E-6, 1E-5, 1E-4, 1E-3, 1E-2, .1],
params = [{'nu':nus, 'kernel': ['linear']},
            {'nu':nus, 'gamma':np.logspace(-5,0,num=6), 'kernel':['rbf']},
            {'nu':nus, 'gamma':np.logspace(-5,0,num=6), 'degree':np.arange(2,5), 'kernel':['poly']},
            {'nu':nus, 'gamma':np.logspace(-5,0,num=6), 'kernel':['sigmoid']}]
 classifier = NuSVC(probability = True)
 gs = GridSearchCV(classifier, params)
```

A.1.2.2. SVC

```
params = [{'C':np.logspace(-5,5,num=11), 'kernel': ['linear']},
            {'C':np.logspace(-5,5,num=11), 'gamma':np.logspace(-5,0,num=6), 'kernel':['rbf']},
            {'C':np.logspace(-5,5,num=11), 'gamma':np.logspace(-5,0,num=6), 'degree':np.arange(2,5), 'kernel':['poly']},
           {'C':np.logspace(-5,5,num=11), 'gamma':np.logspace(-5,0,num=6), 'kernel':['sigmoid']}]
classifier = SVC(probability = True)
gs = GridSearchCV(classifier, params)
```

A.1.2.3. Stochastic Gradient Descent (SGD)

```
params = [{'alpha':np.logspace(-5,5,num=11), 'loss':['log']},
            {'alpha':np.logspace(-5,5,num=11), 'loss':['modified_huber']},
            {'alpha':np.logspace(-5,5,num=11), 'loss':['perceptron']},
            {'alpha':np.logspace(-5,5,num=11), 'loss':['squared_hinge']}]
classifier = SGDClassifier(shuffle = True)
 gs = GridSearchCV(classifier, params)
```

A.1.2.4. Linear SVC

```
params = [{'C':np.logspace(-5,5,num=22)}]
classifier = LinearSVC()
gs = GridSearchCV(classifier, params)
```

A.1.2.5. Decision Tree

```
params = [{'criterion':['entropy'], 'max_features':np.arange(1, len(ml))},
```

```
            {'criterion':['gini'], 'max_features':np.arange(1, len(ml))}]
classifier = DecisionTreeClassifier()
gs = GridSearchCV(classifier, params)
```

A.1.2.6. K Nearest Neighbors

```
params = [{'n_neighbors':np.arange(1,20)}]
classifier = KNeighborsClassifier()
gs = GridSearchCV(classifier, params)
```

REFERENCES

1. Bratt-Leal, A.M., R.L. Carpenedo, and T.C. McDevitt, *Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation.* Biotechnology progress, 2009. **25**(1): p. 43-51.

2. Kurosawa, H., *Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells.* Journal of bioscience and bioengineering, 2007. **103**(5): p. 389-398.

3. Doevendans, P.A., et al., *Differentiation of cardiomyocytes in floating embryoid bodies is comparable to fetal cardiomyocytes.* Journal of molecular and cellular cardiology, 2000. **32**(5): p. 839-851.

4. Itskovitz-Eldor, J., et al., *Differentiation of human embryonic stem cells into embryoid bodies compromising the three embryonic germ layers.* Molecular medicine (Cambridge, Mass.), 2000. **6**(2): p. 88-95.

5. Desbaillets, I., et al., *Embryoid bodies: an in vitro model of mouse embryogenesis.* Experimental physiology, 2000. **85**(6): p. 645-651.

6. Chinzei, R., et al., *Embryoid-body cells derived from a mouse embryonic stem cell line show differentiation into functional hepatocytes.* Hepatology (Baltimore, Md.), 2002. **36**(1): p. 22-29.

7. Kramer, J., et al., *Embryonic stem cell-derived chondrogenic differentiation in vitro: activation by BMP-2 and BMP-4.* Mechanisms of development, 2000. **92**(2): p. 193-205.

8. Farina, A., et al., *Temporal proteomic profiling of embryonic stem cell secretome during cardiac and neural differentiation.* Proteomics, 2011. **11**(20): p. 3972-3982.

9. Ker, E.D.F., et al., *Engineering spatial control of multiple differentiation fates within a stem cell population.* Biomaterials, 2011. **32**(13): p. 3413-3422.

10. Galvin, K.E., et al., *Nodal signaling regulates the bone morphogenic protein pluripotency pathway in mouse embryonic stem cells.* The Journal of biological chemistry, 2010. **285**(26): p. 19747-19756.

11. ten Berge, D., et al., *Embryonic stem cells require Wnt proteins to prevent differentiation to epiblast stem cells.* Nature cell biology, 2011. **13**(9): p. 1070-1075.

12. Haque, A., et al., *Characterization and neural differentiation of mouse embryonic and induced pluripotent stem cells on cadherin-based substrata.* Biomaterials, 2012. **33**(20): p. 5094-5106.

13. Hayashi, Y., et al., *Integrins regulate mouse embryonic stem cell self-renewal.* Stem cells (Dayton, Ohio), 2007. **25**(12): p. 3005-3015.

14. Smith, L.A., et al., *Enhancing osteogenic differentiation of mouse embryonic stem cells by nanofibers.* Tissue engineering. Part A, 2009. **15**(7): p. 1855-1864.

15. Sala, A., et al., *Engineering 3D cell instructive microenvironments by rational assembly of artificial extracellular matrices and cell patterning.* Integrative biology: quantitative biosciences from nano to macro, 2011. **3**(11): p. 1102-1111.

16. Evans, N.D., et al., *Extracellular matrix-mediated osteogenic differentiation of murine embryonic stem cells.* Biomaterials, 2010. **31**(12): p. 3244-3252.

17. Peng, S., et al., *Gelatin induces trophectoderm differentiation of mouse embryonic stem cells.* Cell biology international, 2011. **35**(6): p. 587-591.

18. Przybyla, L.M. and J. Voldman, *Attenuation of extrinsic signaling reveals the importance of matrix remodeling on maintenance of embryonic stem cell self-renewal.* Proceedings of the National Academy of Sciences of the United States of America, 2012. **109**(3): p. 835-840.

19. Evans, N.D., et al., *Substrate stiffness affects early differentiation events in embryonic stem cells.* European cells & materials, 2009. **18**: p. 1-13; discussion 13-14.

20. Asthagiri, A.R. and D.A. Lauffenburger, *Bioengineering models of cell signaling.* Annual review of biomedical engineering, 2000. **2**: p. 31-53.

21. Lutter, D., P. Bruns, and F.J. Theis, *An ensemble approach for inferring semi-quantitative regulatory dynamics for the differentiation of mouse embryonic stem cells using prior knowledge.* Advances in experimental medicine and biology, 2012. **736**: p. 247-260.

22. Tasseff, R., et al., *Modeling and analysis of retinoic acid induced differentiation of uncommitted precursor cells.* Integrative biology: quantitative biosciences from nano to macro, 2011. **3**(5): p. 578-591.

23. Ng, P.M.-L. and T. Lufkin, *Embryonic stem cells: protein interaction networks.* Biomolecular concepts, 2011. **2**(1-2): p. 13-25.

24. D'Haeseleer, P., S. Liang, and R. Somogyi, *Genetic network inference: from co-expression clustering to reverse engineering.* Bioinformatics (Oxford, England), 2000. **16**(8): p. 707-726.

25. Edelman, L.B., J.A. Eddy, and N.D. Price, *In silico models of cancer.* Wiley interdisciplinary reviews. Systems biology and medicine, 2010. **2**(4): p. 438-459.

26. Narang, V., et al., *Selection of Mesenchymal-Like Metastatic Cells in Primary Tumors - An in silico Investigation.* Frontiers in immunology, 2012. **3**.

27. Azimi, M., Y. Jamali, and M.R.K. Mofrad, *Accounting for diffusion in agent based models of reaction-diffusion systems with application to cytoskeletal diffusion.* PloS one, 2011. **6**(9).

28. Buske, P., et al., *On the biomechanics of stem cell niche formation in the gut--modelling growing organoids.* The FEBS journal, 2012. **279**(18): p. 3475-3487.

29. Chen, W., et al., *Agent based modeling of blood coagulation system: implementation using a GPU based high speed framework.* Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2011. **2011**: p. 145-148.

30. Qutub, A.A. and A.S. Popel, *Elongation, proliferation & migration differentiate endothelial cell phenotypes and determine capillary sprouting.* BMC systems biology, 2009. **3**.

31. Setty, Y., et al., *Four-dimensional realistic modeling of pancreatic organogenesis.* Proceedings of the National Academy of Sciences of the United States of America, 2008. **105**(51): p. 20374-20379.

32. Thorne, B.C., A.M. Bailey, and S.M. Peirce, *Combining experiments with multi-cell agent-based modeling to study biological tissue patterning.* Briefings in bioinformatics, 2007. **8**(4): p. 245-257.

33. Zhang, L., et al., *Developing a multiscale, multi-resolution agent-based brain tumor model by graphics processing units.* Theoretical biology & medical modelling, 2011. **8**.

34. Thorne, B.C., et al., *Agent-based modeling of multicell morphogenic processes during development.* Birth defects research. Part C, Embryo today: reviews, 2007. **81**(4): p. 344-353.

35. Beyer, T. and M. Meyer-Hermann, *Multiscale modeling of cell mechanics and tissue organization.* IEEE engineering in medicine and biology magazine: the

quarterly magazine of the Engineering in Medicine & Biology Society, 2009. **28**(2): p. 38-45.

36. Dietrich, J.-E. and T. Hiiragi, *Stochastic patterning in the mouse pre-implantation embryo.* Development (Cambridge, England), 2007. **134**(23): p. 4219-4231.

37. Sandersius, S.A., C.J. Weijer, and T.J. Newman, *Emergent cell and tissue dynamics from subcellular modeling of active biomechanical processes.* Physical biology, 2011. **8**(4): p. 045007.

38. Martin, G.R., *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* Proc Natl Acad Sci U S A, 1981. **78**(12): p. 7634-8.

39. Moreadith, R.W. and K.H. Graves, *Derivation of pluripotential embryonic stem cells from the rabbit.* Trans Assoc Am Physicians, 1992. **105**: p. 197-203.

40. Thomson, J.A., et al., *Isolation of a primate embryonic stem cell line.* Proc Natl Acad Sci U S A, 1995. **92**(17): p. 7844-8.

41. Thomson, J.A., et al., *Embryonic stem cell lines derived from human blastocysts.* Science, 1998. **282**(5391): p. 1145-7.

42. Robertson, E.J., *Derivation and maintenance of embryonic stem cell cultures.* Methods Mol Biol, 1990. **5**: p. 223-36.

43. Metzger, J.M., et al., *Embryonic stem cell cardiogenesis applications for cardiovascular research.* Trends Cardiovasc Med, 1997. **7**(2): p. 63-8.

44. Rust, E.M., et al., *Gene transfer into mouse embryonic stem cell-derived cardiac myocytes mediated by recombinant adenovirus.* In Vitro Cell Dev Biol Anim, 1997. **33**(4): p. 270-6.

45. Mignone, J.L., et al., *Cardiogenesis from human embryonic stem cells.* Circ J, 2010. **74**(12): p. 2517-26.

46. Bain, G., et al., *Retinoic acid promotes neural and represses mesodermal gene expression in mouse embryonic stem cells in culture.* Biochem Biophys Res Commun, 1996. **223**(3): p. 691-4.

47. Lee, C.S., et al., *Neurotrophin and GDNF family ligands promote survival and alter excitotoxic vulnerability of neurons derived from murine embryonic stem cells.* Exp Neurol, 2005. **191**(1): p. 65-76.

48. Krug, A.K., et al., *Human embryonic stem cell-derived test systems for developmental neurotoxicity: a transcriptomics approach.* Arch Toxicol, 2013. **87**(1): p. 123-43.

49. Vojnits, K., et al., *A tanscriptomics study to elucidate the toxicological mechanism of methylmercury chloride in a human stem cell based in vitro test.* Curr Med Chem, 2012. **19**(36): p. 6224-32.

50. Kayama, M., et al., *Recent advances in corneal regeneration and possible application of embryonic stem cell-derived corneal epithelial cells.* Clin Ophthalmol, 2007. **1**(4): p. 373-82.

51. Laflamme, M.A., et al., *Cardiomyocytes derived from human embryonic stem cells in pro-survival factors enhance function of infarcted rat hearts.* Nat Biotechnol, 2007. **25**(9): p. 1015-24.

52. Ueno, H., et al., *Experimental transplantation of corneal epithelium-like cells induced by Pax6 gene transfection of mouse embryonic stem cells.* Cornea, 2007. **26**(10): p. 1220-7.

53. Ueno, H. and I.L. Weissman, *Stem cells: blood lines from embryo to adult.* Nature, 2007. **446**(7139): p. 996-7.

54. Watanabe, K., et al., *A ROCK inhibitor permits survival of dissociated human embryonic stem cells.* Nat Biotechnol, 2007. **25**(6): p. 681-6.

55. Psathaki, O.E., et al., *Ultrastructural characterization of mouse embryonic stem cell-derived oocytes and granulosa cells.* Stem Cells Dev, 2011. **20**(12): p. 2205-15.

56. Ghaedi, M., et al., *Hepatic differentiation of human embryonic stem cells on growth factor-containing surfaces.* J Tissue Eng Regen Med, 2014. **8**(11): p. 886-95.

57. Ghaedi, M., et al., *Hepatic differentiation from human mesenchymal stem cells on a novel nanofiber scaffold.* Cell Mol Biol Lett, 2012. **17**(1): p. 89-106.

58. De Angelis, M.T., et al., *Novel pancreas organogenesis markers refine the pancreatic differentiation roadmap of embryonic stem cells.* Stem Cell Rev, 2014. **10**(2): p. 269-79.

59. Sui, J., et al., *Directed differentiation of embryonic stem cells allows exploration of novel transcription factor genes for pancreas development.* Stem Cell Rev, 2012. **8**(3): p. 803-12.

60. Terada, T., *Huge clusters of embryonic stem cells in human embryos: a morphologic study.* Microsc Res Tech, 2014. **77**(10): p. 825-31.

61. Rolletschek, A., P. Blyszczuk, and A.M. Wobus, *Embryonic stem cell-derived cardiac, neuronal and pancreatic cells as model systems to study toxicological effects.* Toxicol Lett, 2004. **149**(1-3): p. 361-9.

62. Gong, S.P., et al., *The co-injection of somatic cells with embryonic stem cells affects teratoma formation and the properties of teratoma-derived stem cell-like cells.* PloS one, 2014. **9**(9): p. e105975.

63. Qin, Y., et al., *Generation of Embryonic Stem Cells from Mouse Adipose-tissue Derived Cells via Somatic Cell Nuclear Transfer.* Cell Cycle, 2015: p. 0.

64. Zhao, W., et al., *Embryonic stem cell markers.* Molecules, 2012. **17**(6): p. 6196-236.

65. Incitti, T., et al., *Sorting of Sox1-GFP Mouse Embryonic Stem Cells Enhances Neuronal Identity Acquisition upon Factor-Free Monolayer Differentiation.* Biores Open Access, 2014. **3**(3): p. 127-35.

66. Mekala, S.R., et al., *Derivation, characterization and retinal differentiation of induced pluripotent stem cells.* J Biosci, 2013. **38**(1): p. 123-34.

67. Dowell, K.G., et al., *Novel insights into embryonic stem cell self-renewal revealed through comparative human and mouse systems biology networks.* Stem cells (Dayton, Ohio), 2014. **32**(5): p. 1161-72.

68. Faddah, D.A., et al., *Single-cell analysis reveals that expression of nanog is biallelic and equally variable as that of other pluripotency factors in mouse ESCs.* Cell Stem Cell, 2013. **13**(1): p. 23-9.

69. Huang, G., et al., *Molecular basis of embryonic stem cell self-renewal: from signaling pathways to pluripotency network.* Cell Mol Life Sci, 2015.

70. Kalkan, T. and A. Smith, *Mapping the route from naive pluripotency to lineage specification.* Philos Trans R Soc Lond B Biol Sci, 2014. **369**(1657).

71. Ormsbee Golden, B.D., E.L. Wuebben, and A. Rizzino, *Sox2 expression is regulated by a negative feedback loop in embryonic stem cells that involves AKT signaling and FoxO1.* PloS one, 2013. **8**(10): p. e76345.

72. Rizzino, A., *Concise review: The Sox2-Oct4 connection: critical players in a much larger interdependent network integrated at multiple levels.* Stem cells (Dayton, Ohio), 2013. **31**(6): p. 1033-9.

73. Takashima, Y., et al., *Resetting transcription factor control circuitry toward ground-state pluripotency in human.* Cell, 2014. **158**(6): p. 1254-69.

74. Zhang, X., et al., *Gene regulatory networks mediating canonical Wnt signal-directed control of pluripotency and differentiation in embryo stem cells.* Stem cells (Dayton, Ohio), 2013. **31**(12): p. 2667-79.

75. Li, P., et al., *A tight control of Rif1 by Oct4 and Smad3 is critical for mouse embryonic stem cell stability.* Cell Death Dis, 2015. **6**: p. e1588.

76. Sun, L.T., et al., *Nanog co-regulated by Nodal/Smad2 and Oct4 is required for pluripotency in developing mouse epiblast.* Dev Biol, 2014. **392**(2): p. 182-92.

77. Wu, G. and H.R. Scholer, *Role of Oct4 in the early embryo development.* Cell Regen (Lond), 2014. **3**(1): p. 7.

78. Munoz Descalzo, S., et al., *Correlations between the levels of Oct4 and Nanog as a signature for naive pluripotency in mouse embryonic stem cells.* Stem cells (Dayton, Ohio), 2012. **30**(12): p. 2683-91.

79. MacArthur, B.D., et al., *Nanog-dependent feedback loops regulate murine embryonic stem cell heterogeneity.* Nat Cell Biol, 2012. **14**(11): p. 1139-47.

80. Collu, G.M., A. Hidalgo-Sastre, and K. Brennan, *Wnt-Notch signalling crosstalk in development and disease.* Cell Mol Life Sci, 2014. **71**(18): p. 3553-67.

81. Aijian, A.P. and R.L. Garrell, *Digital Microfluidics for Automated Hanging Drop Cell Spheroid Culture.* J Lab Autom, 2014.

82. Bartosh, T.J. and J.H. Ylostalo, *Preparation of anti-inflammatory mesenchymal stem/precursor cells (MSCs) through sphere formation using hanging-drop culture technique.* Curr Protoc Stem Cell Biol, 2014. **28**: p. Unit 2B 6.

83. Sargent, C.Y., et al., *Hydrodynamic modulation of embryonic stem cell differentiation by rotary orbital suspension culture.* Biotechnol Bioeng, 2010. **105**(3): p. 611-26.

84. Magyar, J.P., et al., *Mass production of embryoid bodies in microbeads.* Ann N Y Acad Sci, 2001. **944**: p. 135-43.

85. Ungrin, M.D., et al., *Reproducible, ultra high-throughput formation of multicellular organization from single cell suspension-derived human embryonic stem cell aggregates.* PloS one, 2008. **3**(2): p. e1565.

86. Kinney, M.A., R. Saeed, and T.C. McDevitt, *Systematic analysis of embryonic stem cell differentiation in hydrodynamic environments with controlled embryoid*

body size. Integrative biology: quantitative biosciences from nano to macro, 2012. **4**(6): p. 641-50.

87. Home, P., et al., *Altered subcellular localization of transcription factor TEAD4 regulates first mammalian cell lineage commitment.* Proc Natl Acad Sci U S A, 2012. **109**(19): p. 7362-7.

88. Bielby, R.C., et al., *In vitro differentiation and in vivo mineralization of osteogenic cells derived from human embryonic stem cells.* Tissue Eng, 2004. **10**(9-10): p. 1518-25.

89. Rong, L., et al., *GATA-6 promotes cell survival by up-regulating BMP-2 expression during embryonic stem cell differentiation.* Mol Biol Cell, 2012. **23**(18): p. 3754-63.

90. Shevde, N.K. and A.A. Mael, *Techniques in embryoid body formation from human pluripotent stem cells.* Methods Mol Biol, 2013. **946**: p. 535-46.

91. Sun, C., et al., *Epigenetic regulation of histone modifications and Gata6 gene expression induced by maternal diet in mouse embryoid bodies in a model of developmental programming.* BMC Dev Biol, 2015. **15**(1): p. 3.

92. Vassilieva, S., et al., *A system to enrich for primitive streak-derivatives, definitive endoderm and mesoderm, from pluripotent cells in culture.* PloS one, 2012. **7**(6): p. e38645.

93. Teramoto, K., et al., *Hepatocyte differentiation from embryonic stem cells and umbilical cord blood cells.* J Hepatobiliary Pancreat Surg, 2005. **12**(3): p. 196-202.

94. Teramoto, K., et al., *Teratoma formation and hepatocyte differentiation in mouse liver transplanted with mouse embryonic stem cell-derived embryoid bodies.* Transplant Proc, 2005. **37**(1): p. 285-6.

95. Kumashiro, Y., et al., *Enrichment of hepatocytes differentiated from mouse embryonic stem cells as a transplantable source.* Transplantation, 2005. **79**(5): p. 550-7.

96. Kumashiro, Y., et al., *Isolation of hepatocyte-like cells from mouse embryoid body cells.* Transplant Proc, 2005. **37**(1): p. 299-300.

97. Lian, X., et al., *Directed cardiomyocyte differentiation from human pluripotent stem cells by modulating Wnt/beta-catenin signaling under fully defined conditions.* Nat Protoc, 2013. **8**(1): p. 162-75.

98. Lian, X., et al., *Insulin inhibits cardiac mesoderm, not mesendoderm, formation during cardiac differentiation of human pluripotent stem cells and modulation of canonical Wnt signaling can rescue this inhibition.* Stem cells (Dayton, Ohio), 2013. **31**(3): p. 447-57.

99. Zhang, J., et al., *Functional cardiomyocytes derived from human induced pluripotent stem cells.* Circ Res, 2009. **104**(4): p. e30-41.

100. Crawford, T.Q. and H. Roelink, *The notch response inhibitor DAPT enhances neuronal differentiation in embryonic stem cell-derived embryoid bodies independently of sonic hedgehog signaling.* Dev Dyn, 2007. **236**(3): p. 886-92.

101. Carpenedo, R.L., et al., *Homogeneous and organized differentiation within embryoid bodies induced by microsphere-mediated delivery of small molecules.* Biomaterials, 2009. **30**(13): p. 2507-15.

102. Munoz Descalzo, S., et al., *Correlations between the levels of Oct4 and Nanog as a signature for naive pluripotency in mouse embryonic stem cells.* Stem Cells, 2012. **30**(12): p. 2683-91.

103. Wang, Z., et al., *Distinct lineage specification roles for NANOG, OCT4, and SOX2 in human embryonic stem cells.* Cell Stem Cell, 2012. **10**(4): p. 440-54.

104. Radzisheuskaya, A., et al., *A defined Oct4 level governs cell state transitions of pluripotency entry and differentiation into all embryonic lineages.* Nat Cell Biol, 2013. **15**(6): p. 579-90.

105. Faunes, F., et al., *A membrane-associated beta-catenin/Oct4 complex correlates with ground-state pluripotency in mouse embryonic stem cells.* Development, 2013. **140**(6): p. 1171-83.

106. Herberg, M., et al., *A model-based analysis of culture-dependent phenotypes of mESCs.* PLoS One, 2014. **9**(3): p. e92496.

107. Gagliardi, A., et al., *A direct physical interaction between Nanog and Sox2 regulates embryonic stem cell self-renewal.* EMBO J, 2013. **32**(16): p. 2231-47.

108. Navarro, P., et al., *OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells.* EMBO J, 2012. **31**(24): p. 4547-62.

109. Ochiai, H., et al., *Stochastic promoter activation affects Nanog expression variability in mouse embryonic stem cells.* Sci Rep, 2014. **4**: p. 7125.

110. Ohnishi, Y., et al., *Cell-to-cell expression variability followed by signal reinforcement progressively segregates early mouse lineages.* Nat Cell Biol, 2014. **16**(1): p. 27-37.

111. Yasui, N., et al., *Directed network wiring identifies a key protein interaction in embryonic stem cell differentiation.* Mol Cell, 2014. **54**(6): p. 1034-41.

112. Joo, J.Y., et al., *Establishment of a primed pluripotent epiblast stem cell in FGF4-based conditions.* Sci Rep, 2014. **4**: p. 7477.

113. Bessonnard, S., et al., *Gata6, Nanog and Erk signaling control cell fate in the inner cell mass through a tristable regulatory network.* Development, 2014. **141**(19): p. 3637-48.

114. Xu, H., et al., *Construction and validation of a regulatory network for pluripotency and self-renewal of mouse embryonic stem cells.* PLoS Comput Biol, 2014. **10**(8): p. e1003777.

115. Hall, J., et al., *Oct4 and LIF/Stat3 additively induce Kruppel factors to sustain embryonic stem cell self-renewal.* Cell Stem Cell, 2009. **5**(6): p. 597-609.

116. Do, D.V., et al., *A genetic and developmental pathway from STAT3 to the OCT4-NANOG circuit is essential for maintenance of ICM lineages in vivo.* Genes Dev, 2013. **27**(12): p. 1378-90.

117. Morgenthal, K., W. Weckwerth, and R. Steuer, *Metabolomic networks in plants: Transitions from pattern recognition to biological interpretation.* Biosystems, 2006. **83**(2-3): p. 108-17.

118. Steuer, R., *Review: on the analysis and interpretation of correlations in metabolomic data.* Brief Bioinform, 2006. **7**(2): p. 151-8.

119. Yeang, C.H. and M. Vingron, *A joint model of regulatory and metabolic networks.* BMC Bioinformatics, 2006. **7**: p. 332.

120. Schoeberl, B., et al., *Computational modeling of the dynamics of the MAP kinase cascade activated by surface and internalized EGF receptors.* Nat Biotechnol, 2002. **20**(4): p. 370-5.

121. Stelling, J. and E.D. Gilles, *Mathematical modeling of complex regulatory networks.* IEEE Trans Nanobioscience, 2004. **3**(3): p. 172-9.

122. Stelling, J., E.D. Gilles, and F.J. Doyle, 3rd, *Robustness properties of circadian clock architectures.* Proc Natl Acad Sci U S A, 2004. **101**(36): p. 13210-5.

123. Artyomov, M.N., A. Meissner, and A.K. Chakraborty, *A model for genetic and epigenetic regulatory networks identifies rare pathways for transcription factor induced pluripotency.* PLoS Comput Biol, 2010. **6**(5): p. e1000785.

124. Hu, Z., M. Qian, and M.Q. Zhang, *Novel Markov model of induced pluripotency predicts gene expression changes in reprogramming.* BMC systems biology, 2011. **5 Suppl 2**: p. S8.

125. Chickarmane, V., et al., *Transcriptional dynamics of the embryonic stem cell switch.* PLoS Comput Biol, 2006. **2**(9): p. e123.

126. Villani, M., A. Barbieri, and R. Serra, *A dynamical model of genetic networks for cell differentiation.* PloS one, 2011. **6**(3): p. e17703.

127. Setty, Y., *Multi-scale computational modeling of developmental biology.* Bioinformatics (Oxford, England), 2012. **28**(15): p. 2022-8.

128. Setty, Y., et al., *A model of stem cell population dynamics: in silico analysis and in vivo validation.* Development (Cambridge, England), 2012. **139**(1): p. 47-56.

129. Meng, T.C., S. Somani, and P. Dhar, *Modeling and simulation of biological systems with stochasticity.* In Silico Biol, 2004. **4**(3): p. 293-309.

130. Chickarmane, V., V. Olariu, and C. Peterson, *Probing the role of stochasticity in a model of the embryonic stem cell: heterogeneous gene expression and reprogramming efficiency.* BMC systems biology, 2012. **6**: p. 98.

131. Krupinski, P., V. Chickarmane, and C. Peterson, *Computational multiscale modeling of embryo development.* Curr Opin Genet Dev, 2012. **22**(6): p. 613-8.

132. Wu, J. and E.S. Tzanakakis, *Contribution of stochastic partitioning at human embryonic stem cell division to NANOG heterogeneity.* PloS one, 2012. **7**(11): p. e50715.

133. Van Winkle, A.P., I.D. Gates, and M.S. Kallos, *Mass transfer limitations in embryoid bodies during human embryonic stem cell differentiation.* Cells, tissues, organs, 2012. **196**(1): p. 34-47.

134. Byrne, M.B., et al., *Microfluidic platform for the study of intercellular communication via soluble factor-cell and cell-cell paracrine signaling.* Biomicrofluidics, 2014. **8**(4): p. 044104.

135. Ellison, D., A. Munden, and A. Levchenko, *Computational model and microfluidic platform for the investigation of paracrine and autocrine signaling in mouse embryonic stem cells.* Mol Biosyst, 2009. **5**(9): p. 1004-12.

136. Cowan, A.E., et al., *Spatial modeling of cell signaling networks.* Methods Cell Biol, 2012. **110**: p. 195-221.

137. Grundl, D., et al., *Reaction-diffusion modelling for microphysiometry on cellular specimens.* Med Biol Eng Comput, 2013. **51**(4): p. 387-95.

138. Ayodele, S.G., F. Varnik, and D. Raabe, *Lattice Boltzmann study of pattern formation in reaction-diffusion systems.* Phys Rev E Stat Nonlin Soft Matter Phys, 2011. **83**(1 Pt 2): p. 016702.

139. Dziekan, P., A. Lemarchand, and B. Nowakowski, *Particle dynamics simulations of Turing patterns.* J Chem Phys, 2012. **137**(7): p. 074107.

140. Chen, J.X. and R. Kapral, *Mesoscopic dynamics of diffusion-influenced enzyme kinetics.* J Chem Phys, 2011. **134**(4): p. 044503.

141. Ansumali, S. and I.V. Karlin, *Consistent lattice Boltzmann method.* Phys Rev Lett, 2005. **95**(26): p. 260605.

142. Bonabeau, E., *Agent-based modeling: methods and techniques for simulating human systems.* Proc Natl Acad Sci U S A, 2002. **99 Suppl 3**: p. 7280-7.

143. Grimm, V., et al., *Pattern-oriented modeling of agent-based complex systems: lessons from ecology.* Science, 2005. **310**(5750): p. 987-91.

144. Knoeri, C., et al., *Towards a dynamic assessment of raw materials criticality: linking agent-based demand--with material flow supply modelling approaches.* Sci Total Environ, 2013. **461-462**: p. 808-12.

145. Smith, E.R. and F.R. Conrey, *Agent-based modeling: a new approach for theory building in social psychology.* Pers Soc Psychol Rev, 2007. **11**(1): p. 87-104.

146. Wynn, M.L., P.M. Kulesa, and S. Schnell, *Computational modelling of cell chain migration reveals mechanisms that sustain follow-the-leader behaviour.* J R Soc Interface, 2012. **9**(72): p. 1576-88.

147. Bentley, K., H. Gerhardt, and P.A. Bates, *Agent-based simulation of notch-mediated tip cell selection in angiogenic sprout initialisation.* J Theor Biol, 2008. **250**(1): p. 25-36.

148. Adra, S., et al., *Development of a three dimensional multiscale computational model of the human epidermis.* PloS one, 2010. **5**(1): p. e8511.

149. Figueredo, G.P., et al., *Comparing stochastic differential equations and agent-based modelling and simulation for early-stage cancer.* PloS one, 2014. **9**(4): p. e95150.

150. Norton, K.A. and A.S. Popel, *An agent-based model of cancer stem cell initiated avascular tumour growth and metastasis: the effect of seeding frequency and location.* J R Soc Interface, 2014. **11**(100): p. 20140640.

151. Zhang, L., et al., *Multiscale agent-based modelling of ovarian cancer progression under the stimulation of the STAT 3 pathway.* Int J Data Min Bioinform, 2014. **9**(3): p. 235-53.

152. Setty, Y., et al., *Four-dimensional realistic modeling of pancreatic organogenesis.* Proc Natl Acad Sci U S A, 2008. **105**(51): p. 20374-9.

153. Duan, G. and D. Walther, *The Roles of Post-translational Modifications in the Context of Protein Interaction Networks.* PLoS Comput Biol, 2015. **11**(2): p. e1004049.

154. Helou, Y.A. and A.R. Salomon, *Protein networks and activation of lymphocytes.* Curr Opin Immunol, 2015. **33C**: p. 78-85.

155. Kazemi-Pour, A., B. Goliaei, and H. Pezeshk, *Protein complex discovery by interaction filtering from protein interaction networks using mutual rank coexpression and sequence similarity.* Biomed Res Int, 2015. **2015**: p. 165186.

156. Hartman, J.L., et al., *Yeast Phenomics: An Experimental Approach for Modeling Gene Interaction Networks that Buffer Disease.* Genes (Basel), 2015. **6**(1): p. 24-45.

157. Mas, S., et al., *Network analysis of gene expression in peripheral blood identifies mTOR and NF-kappaB pathways involved in antipsychotic-induced extrapyramidal symptoms.* Pharmacogenomics J, 2015.

158. Wu, S., et al., *Network propagation with dual flow for gene prioritization.* PloS one, 2015. **10**(2): p. e0116505.

159. Mao, L. and H. Resat, *Probabilistic representation of gene regulatory networks.* Bioinformatics (Oxford, England), 2004. **20**(14): p. 2258-69.

160. Woolf, P.J., et al., *Bayesian analysis of signaling networks governing embryonic stem cell fate decisions.* Bioinformatics (Oxford, England), 2005. **21**(6): p. 741-53.

161. Cohen, M., et al., *A theoretical framework for the regulation of Shh morphogen-controlled gene expression.* Development (Cambridge, England), 2014. **141**(20): p. 3868-78.

162. Huang, X. and Z. Zi, *Inferring cellular regulatory networks with Bayesian model averaging for linear regression (BMALR).* Mol Biosyst, 2014. **10**(8): p. 2023-30.

163. Lutter, D., P. Bruns, and F.J. Theis, *An ensemble approach for inferring semi-quantitative regulatory dynamics for the differentiation of mouse embryonic stem cells using prior knowledge.* Adv Exp Med Biol, 2012. **736**: p. 247-60.

164. Mateo, J.L., et al., *Characterization of the neural stem cell gene regulatory network identifies OLIG2 as a multifunctional regulator of self-renewal.* Genome Res, 2015. **25**(1): p. 41-56.

165. Shellman, E.R., et al., *Metabolic network motifs can provide novel insights into evolution: The evolutionary origin of Eukaryotic organelles as a case study.* Comput Biol Chem, 2014. **53PB**: p. 242-250.

166. Wang, P., J. Lu, and X. Yu, *Identification of important nodes in directed biological networks: a network motif approach.* PloS one, 2014. **9**(8): p. e106132.

167. Ruymgaart, A.P. and R. Elber, *Revisiting Molecular Dynamics on a CPU/GPU system: Water Kernel and SHAKE Parallelization.* J Chem Theory Comput, 2012. **8**(11): p. 4624-4636.

168. Glowacki, D.R., et al., *A GPU-accelerated immersive audio-visual framework for interaction with molecular dynamics using consumer depth sensors.* Faraday Discuss, 2014. **169**: p. 63-87.

169. Stone, J.E., et al., *GPU-accelerated analysis and visualization of large structures solved by molecular dynamics flexible fitting.* Faraday Discuss, 2014. **169**: p. 265-83.

170. Bidhendi, A.J. and R.K. Korhonen, *A finite element study of micropipette aspiration of single cells: effect of compressibility.* Comput Math Methods Med, 2012. **2012**: p. 192618.

171. Mendizabal, A., I. Aguinaga, and E. Sanchez, *Characterisation and modelling of brain tissue for surgical simulation.* J Mech Behav Biomed Mater, 2015. **45C**: p. 1-10.

172. Garcia, J.J. and D.H. Cortes, *A nonlinear biphasic viscohyperelastic model for articular cartilage.* J Biomech, 2006. **39**(16): p. 2991-8.

173. Payne, T., et al., *The evaluation of new multi-material human soft tissue simulants for sports impact surrogates.* J Mech Behav Biomed Mater, 2015. **41**: p. 336-56.

174. Albert, P.J. and U.S. Schwarz, *Dynamics of cell shape and forces on micropatterned substrates predicted by a cellular Potts model.* Biophys J, 2014. **106**(11): p. 2340-52.

175. Daub, J.T. and R.M. Merks, *Cell-based computational modeling of vascular morphogenesis using Tissue Simulation Toolkit.* Methods Mol Biol, 2015. **1214**: p. 67-127.

176. Li, J.F. and J. Lowengrub, *The effects of cell compressibility, motility and contact inhibition on the growth of tumor cell clusters using the Cellular Potts Model.* J Theor Biol, 2014. **343**: p. 79-91.

177. Andasari, V., et al., *Integrating intracellular dynamics using CompuCell3D and Bionetsolver: applications to multiscale modelling of cancer cell growth and invasion.* PloS one, 2012. **7**(3): p. e33726.

178. Yim, H. and Y. Sohn, *Numerical simulation and visualization of elastic waves using mass-spring lattice model.* IEEE Trans Ultrason Ferroelectr Freq Control, 2000. **47**(3): p. 549-58.

179. Mohammadi, H., *A numerical method to enhance the accuracy of mass-spring systems for modeling soft tissue deformations.* J Appl Biomech, 2009. **25**(3): p. 271-8.

180. Mosegaard, J., P. Herborg, and T.S. Sorensen, *A GPU accelerated spring mass system for surgical simulation.* Stud Health Technol Inform, 2005. **111**: p. 342-8.

181. Neylon, J., et al., *A GPU based high-resolution multilevel biomechanical head and neck model for validating deformable image registration.* Medical physics, 2015. **42**(1): p. 232.

182. Risholm, P. and E. Samset, *Haptic guided 3-D deformable image registration.* Int J Comput Assist Radiol Surg, 2009. **4**(3): p. 215-23.

183. Apolloni, B., et al., *A general framework for learning rules from data.* IEEE Trans Neural Netw, 2004. **15**(6): p. 1333-49.

184. Gallant, S.I., *Perceptron-based learning algorithms.* IEEE Trans Neural Netw, 1990. **1**(2): p. 179-91.

185. Jelinek, H.F., et al., *Machine learning and pattern classification in identification of indigenous retinal pathology.* Conference proceedings: ... Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Conference, 2011. **2011**: p. 5951-4.

186. Stallkamp, J., et al., *Man vs. computer: benchmarking machine learning algorithms for traffic sign recognition.* Neural Netw, 2012. **32**: p. 323-32.

187. Mumenthaler, C. and D. Sander, *Automatic Integration of Social Information in Emotion Recognition.* J Exp Psychol Gen, 2015.

188. Rice, L.M., et al., *Computer-Assisted Face Processing Instruction Improves Emotion Recognition, Mentalizing, and Social Skills in Students with ASD.* J Autism Dev Disord, 2015.

189. Wu, L., and Weng, *Probability estimates for multi-class classification by pariwise coupling.* JMLR, 2004. **5**: p. 975-1005.

190. A. McCallum, K.N., *A comparison of event models for Naive Bayes text classification.* Proc/ AAAI/ICM, 1998. **98**: p. 41-48.

191. Y. Tsuruoka, J.T., S. Ananiadou, *Stochastic Gradient Descent.* AFNL/ACL, 2009.

192. Friedman, J., *Greedy Function Approximation: A Gradietn Boosting Machine.* The Annals of Satistics 2001. **25**(5).

193. Breiman, L., *Arcing classifiers.* The Annals of Statistics, 1998. **26**(3).

194. Kim, J.J., S.L. Vega, and P.V. Moghe, *A high content imaging-based approach for classifying cellular phenotypes.* Methods Mol Biol, 2013. **1052**: p. 41-8.

195. Vidi, P.A., et al., *Interconnected contribution of tissue morphogenesis and the nuclear protein NuMA to the DNA damage response.* J Cell Sci, 2012. **125**(Pt 2): p. 350-61.

196. Vega, S.L., et al., *High-content imaging-based screening of microenvironment-induced changes to stem cells.* J Biomol Screen, 2012. **17**(9): p. 1151-62.

197. Nandy, K., et al., *Automatic segmentation and supervised learning-based selection of nuclei in cancer tissue images.* Cytometry A, 2012. **81**(9): p. 743-54.

198. Bergmeir, C., M. Garcia Silvente, and J.M. Benitez, *Segmentation of cervical cell nuclei in high-resolution microscopic images: A new algorithm and a web-based software framework.* Comput Methods Programs Biomed, 2012. **107**(3): p. 497-512.

199. Bergeest, J.P. and K. Rohr, *Efficient globally optimal segmentation of cells in fluorescence microscopy images using level sets and convex energy functionals.* Med Image Anal, 2012. **16**(7): p. 1436-44.

200. Du, T.H., W.C. Puah, and M. Wasser, *Cell cycle phase classification in 3D in vivo microscopy of Drosophila embryogenesis.* BMC Bioinformatics, 2011. **12 Suppl 13**: p. S18.

201. Bunyak, F., A. Hafiane, and K. Palaniappan, *Histopathology tissue segmentation by combining fuzzy clustering with multiphase vector level sets.* Adv Exp Med Biol, 2011. **696**: p. 413-24.

202. Ali, S., et al., *Adaptive energy selective active contour with shape priors for nuclear segmentation and gleason grading of prostate cancer.* Med Image Comput Comput Assist Interv, 2011. **14**(Pt 1): p. 661-9.

203. Langer, L., et al., *Computer-aided diagnostics in digital pathology: automated evaluation of early-phase pancreatic cancer in mice.* Int J Comput Assist Radiol Surg, 2014.

204. Davidson, R.M., et al., *Comparative transcriptomics of three Poaceae species reveals patterns of gene expression evolution.* Plant J, 2012. **71**(3): p. 492-502.

205. Treiser, M.D., et al., *Cytoskeleton-based forecasting of stem cell lineage fates.* Proc Natl Acad Sci U S A, 2010. **107**(2): p. 610-5.

206. Liu, E., et al., *Parsing the early cytoskeletal and nuclear organizational cues that demarcate stem cell lineages.* Cell Cycle, 2010. **9**(11): p. 2108-17.

207. Liu, E., et al., *High-content profiling of cell responsiveness to graded substrates based on combinyatorially variant polymers.* Comb Chem High Throughput Screen, 2009. **12**(7): p. 646-55.

208. Treiser, M.D., et al., *Profiling cell-biomaterial interactions via cell-based fluororeporter imaging.* Biotechniques, 2007. **43**(3): p. 361-6, 368.

209. Mukherjee, R., et al., *Automated tissue classification framework for reproducible chronic wound assessment.* Biomed Res Int, 2014. **2014**: p. 851582.

210. Ye, D.H., et al., *Regional manifold learning for disease classification.* IEEE Trans Med Imaging, 2014. **33**(6): p. 1236-47.

211. Cruz-Roa, A.A., et al., *A deep learning architecture for image representation, visual interpretability and automated basal-cell carcinoma cancer detection.* Med Image Comput Comput Assist Interv, 2013. **16**(Pt 2): p. 403-10.

212. Garcia Molina, J.F., et al., *Incremental learning with SVM for multimodal classification of prostatic adenocarcinoma.* PLoS One, 2014. **9**(4): p. e93600.

213. Hashimshony, T., et al., *Spatiotemporal transcriptomics reveals the evolutionary history of the endoderm germ layer.* Nature, 2014.

214. Murray, J.I. and Z. Bao, *Automated lineage and expression profiling in live Caenorhabditis elegans embryos.* Cold Spring Harb Protoc, 2012. **2012**(8).

215. Rubel, O., et al., *Integrating data clustering and visualization for the analysis of 3D gene expression data.* IEEE/ACM Trans Comput Biol Bioinform, 2010. **7**(1): p. 64-79.

216. Ning, F., et al., *Toward automatic phenotyping of developing embryos from videos.* IEEE Trans Image Process, 2005. **14**(9): p. 1360-71.

217. Zhang, X. and M. Boutros, *A novel phenotypic dissimilarity method for image-based high-throughput screens.* BMC Bioinformatics, 2013. **14**: p. 336.

218. Cai, X., et al., *Joint stage recognition and anatomical annotation of Drosophila gene expression patterns.* Bioinformatics, 2012. **28**(12): p. i16-24.

219. Pruteanu-Malinici, I., D.L. Mace, and U. Ohler, *Automatic annotation of spatial expression patterns via sparse Bayesian factor models.* PLoS Comput Biol, 2011. **7**(7): p. e1002098.

220. Ye, J., et al., *Classification of Drosophila embryonic developmental stage range based on gene expression pattern images.* Comput Syst Bioinformatics Conf, 2006: p. 293-8.

221. Kurosawa, H., *Methods for inducing embryoid body formation: in vitro differentiation system of embryonic stem cells.* J Biosci Bioeng, 2007. **103**(5): p. 389-98.

222. Bratt-Leal, A.M., R.L. Carpenedo, and T.C. McDevitt, *Engineering the embryoid body microenvironment to direct embryonic stem cell differentiation.* Biotechnol Prog, 2009. **25**(1): p. 43-51.

223. Eiraku, M., et al., *Self-organizing optic-cup morphogenesis in three-dimensional culture.* Nature, 2011. **472**(7341): p. 51-6.

224. Lancaster, M.A., et al., *Cerebral organoids model human brain development and microcephaly.* Nature, 2013. **501**(7467): p. 373-9.

225. Koehler, K.R., et al., *Generation of inner ear sensory epithelia from pluripotent stem cells in 3D culture.* Nature, 2013. **500**(7461): p. 217-21.

226. Spence, J.R., et al., *Directed differentiation of human pluripotent stem cells into intestinal tissue in vitro.* Nature, 2011. **470**(7332): p. 105-9.

227. Suga, H., et al., *Self-formation of functional adenohypophysis in three-dimensional culture.* Nature, 2011. **480**(7375): p. 57-62.

228. Kramer, J., et al., *Embryonic stem cell-derived chondrogenic differentiation in vitro: activation by BMP-2 and BMP-4.* Mech Dev, 2000. **92**(2): p. 193-205.

229. Chinzei, R., et al., *Embryoid-body cells derived from a mouse embryonic stem cell line show differentiation into functional hepatocytes.* Hepatology, 2002. **36**(1): p. 22-9.

230. Przybyla, L.M., et al., *Matrix Remodeling Maintains ESC Self-Renewal by Activating Stat3.* Stem cells (Dayton, Ohio), 2013.

231. Sun, Z. and N.L. Komarova, *Stochastic modeling of stem-cell dynamics with control.* Math Biosci, 2012. **240**(2): p. 231-40.

232. Lagerlöf, J.H., J. Kindblom, and P. Bernhardt, *3D modeling of effects of increased oxygenation and activity concentration in tumors treated with radionuclides and antiangiogenic drugs.* Medical physics, 2011. **38**(8): p. 4888-4893.

233. Bratt-Leal, A.M., et al., *Incorporation of biomaterials in multicellular aggregates modulates pluripotent stem cell differentiation.* Biomaterials, 2011. **32**(1): p. 48-56.

234. Pappu, V. and P. Bagchi, *3D computational modeling and simulation of leukocyte rolling adhesion and deformation.* Comput Biol Med, 2008. **38**(6): p. 738-53.

235. Hester, S.D., et al., *A multi-cell, multi-scale model of vertebrate segmentation and somite formation.* PLoS Comput Biol, 2011. **7**(10): p. e1002155.

236. Bibicu, D., L. Moraru, and A. Biswas, *Thyroid nodule recognition based on feature selection and pixel classification methods.* J Digit Imaging, 2013. **26**(1): p. 119-28.

237. Pan, G.J., et al., *Stem cell pluripotency and transcription factor Oct4.* Cell Res, 2002. **12**(5-6): p. 321-9.

238. Nichols, J., et al., *Formation of pluripotent stem cells in the mammalian embryo depends on the POU transcription factor Oct4.* Cell, 1998. **95**(3): p. 379-91.

239. Prudhomme, W., et al., *Multivariate proteomic analysis of murine embryonic stem cell self-renewal versus differentiation signaling.* Proc Natl Acad Sci U S A, 2004. **101**(9): p. 2900-5.

240. Haque, A., et al., *Characterization and neural differentiation of mouse embryonic and induced pluripotent stem cells on cadherin-based substrata.* Biomaterials, 2012. **33**(20): p. 5094-106.

241. Ramasamy, S.K. and N. Lenka, *Notch exhibits ligand bias and maneuvers stage-specific steering of neural differentiation in embryonic stem cells.* Mol Cell Biol, 2010. **30**(8): p. 1946-57.

242. Davey, R.E., et al., *LIF-mediated control of embryonic stem cell self-renewal emerges due to an autoregulatory loop.* FASEB J, 2007. **21**(9): p. 2020-32.

243. Sakai, Y., Y. Yoshiura, and K. Nakazawa, *Embryoid body culture of mouse embryonic stem cells using microwell and micropatterned chips.* J Biosci Bioeng, 2011. **111**(1): p. 85-91.

244. Task, K., M. Jaramillo, and I. Banerjee, *Population based model of human embryonic stem cell (hESC) differentiation during endoderm induction.* PLoS One, 2012. **7**(3): p. e32975.

245. Banerjee, I., et al., *An integer programming formulation to identify the sparse network architecture governing differentiation of embryonic stem cells.* Bioinformatics, 2010. **26**(10): p. 1332-9.

246. Alcobia, I., et al., *Portrayal of the Notch system in embryonic stem cell-derived embryoid bodies.* Cells Tissues Organs, 2011. **193**(4): p. 239-52.

247. Moledina, F., et al., *Predictive microfluidic control of regulatory ligand trajectories in individual pluripotent cells.* Proc Natl Acad Sci U S A, 2012. **109**(9): p. 3264-9.

248. Zandstra, P.W., et al., *Leukemia inhibitory factor (LIF) concentration modulates embryonic stem cell self-renewal and differentiation independently of proliferation.* Biotechnol Bioeng, 2000. **69**(6): p. 607-17.

249. Van Winkle, A.P., I.D. Gates, and M.S. Kallos, *Mass Transfer Limitations in Embryoid Bodies during Human Embryonic Stem Cell Differentiation.* Cells Tissues Organs, 2012.

250. Kinney, M.A., R. Saeed, and T.C. McDevitt, *Systematic analysis of embryonic stem cell differentiation in hydrodynamic environments with controlled embryoid body size.* Integr Biol (Camb), 2012. **4**(6): p. 641-50.

251. Pruteanu-Malinici, I., W.H. Majoros, and U. Ohler, *Automated annotation of gene expression image sequences via non-parametric factor analysis and conditional random fields.* Bioinformatics, 2013. **29**(13): p. i27-35.

252. Puniyani, K., C. Faloutsos, and E.P. Xing, *SPEX2: automated concise extraction of spatial gene expression patterns from Fly embryo ISH images.* Bioinformatics, 2010. **26**(12): p. i47-56.

253. Zhang, W., et al., *A mesh generation and machine learning framework for Drosophila gene expression pattern image analysis.* BMC Bioinformatics, 2013. **14**: p. 372.

254. Trisnadi, N., et al., *Image analysis and empirical modeling of gene and protein expression.* Methods, 2013. **62**(1): p. 68-78.

255. Tomer, R., et al., *Quantitative high-speed imaging of entire developing embryos with simultaneous multiview light-sheet microscopy.* Nat Methods, 2012. **9**(7): p. 755-63.

256. Keller, P.J., *Imaging morphogenesis: technological advances and biological insights.* Science, 2013. **340**(6137): p. 1234168.

257. Morelli, L.G., et al., *Computational approaches to developmental patterning.* Science, 2012. **336**(6078): p. 187-91.

258. Herberg, M., et al., *Image-based quantification and mathematical modeling of spatial heterogeneity in ESC colonies.* Cytometry A, 2015.

259. Minka, T.P., *Automatic Choice of Dimensionaility for PCA.* NIPS, 2000: p. 598-604.

260. Frey, B.J. and D. Dueck, *Clustering by passing messages between data points.* Science, 2007. **315**(5814): p. 972-6.

261. Dorin Comanicicu, P.M., *Mean Shift: A robust approach toward feature space analysis.* IEEE Transactions on Pattern Analysis and Machine Intelligence, 2002: p. 603-619.

262. M. Ester, H.P.K., J. Sander, and X. Xu, *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise.* AAAI Press, 1996: p. 226–231.

263. White, D.E., et al., *Spatial pattern dynamics of 3D stem cell loss of pluripotency via rules basedcomputational modeling.* PLoS Comput Biol, 2013. **9**(3): p. e1002952.

264. Bassi, A., B. Schmid, and J. Huisken, *Optical tomography complements light sheet microscopy for in toto imaging of zebrafish development.* Development, 2015.

265. Baker, R.P., et al., *A combined light sheet fluorescence and differential interference contrast microscope for live imaging of multicellular specimens.* J Microsc, 2015.

266. Keller, P.J. and M.B. Ahrens, *Visualizing Whole-Brain Activity and Development at the Single-Cell Level Using Light-Sheet Microscopy.* Neuron, 2015. **85**(3): p. 462-483.

267. Chen, B.C., et al., *Lattice light-sheet microscopy: imaging molecules to embryos at high spatiotemporal resolution.* Science, 2014. **346**(6208): p. 1257998.

268. Krzic, U., et al., *Multiview light-sheet microscope for rapid in toto imaging.* Nat Methods, 2012. **9**(7): p. 730-3.

269. Keller, P.J., *In vivo imaging of zebrafish embryogenesis.* Methods, 2013. **62**(3): p. 268-78.

270. Setty, Y., et al., *A model of stem cell population dynamics: in silico analysis and in vivo validation.* Development, 2012. **139**(1): p. 47-56.

271. Kazemian, M., et al., *Quantitative analysis of the Drosophila segmentation regulatory network using pattern generating potentials.* PLoS Biol, 2010. **8**(8).

272. Mizutani, C.M., et al., *Formation of the BMP activity gradient in the Drosophila embryo.* Dev Cell, 2005. **8**(6): p. 915-24.

273. Lembong, J., N. Yakoby, and S.Y. Shvartsman, *Pattern formation by dynamically interacting network motifs.* Proc Natl Acad Sci U S A, 2009. **106**(9): p. 3213-8.

274. Tamulonis, C., et al., *A cell-based model of Nematostella vectensis gastrulation including bottle cell formation, invagination and zippering.* Dev Biol, 2011. **351**(1): p. 217-28.

275. Crane, M.M., et al., *Autonomous screening of C. elegans identifies genes implicated in synaptogenesis.* Nat Methods, 2012. **9**(10): p. 977-80.

276. Zacharia, E., et al., *Segmentation of zebrafish embryonic images using a geometric atlas deformation.* Conf Proc IEEE Eng Med Biol Soc, 2012. **2012**: p. 3998-4001.

277. Castro-Gonzalez, C., et al., *Towards a digital model of zebrafish embryogenesis. Integration of cell tracking and gene expression quantification.* Conf Proc IEEE Eng Med Biol Soc, 2010. **2010**: p. 5520-3.

278. Carpenter, A.E., et al., *CellProfiler: image analysis software for identifying and quantifying cell phenotypes.* Genome Biol, 2006. **7**(10): p. R100.

279. Niebruegge, S., et al., *Generation of human embryonic stem cell-derived mesoderm and cardiac cells using size-specified aggregates in an oxygen-controlled bioreactor.* Biotechnol Bioeng, 2009. **102**(2): p. 493-507.

280. Mohr, J.C., et al., *The microwell control of embryoid body size in order to regulate cardiac differentiation of human embryonic stem cells.* Biomaterials, 2010. **31**(7): p. 1885-93.

281. Bauwens, C.L., et al., *Control of human embryonic stem cell colony and aggregate size heterogeneity influences differentiation trajectories.* Stem Cells, 2008. **26**(9): p. 2300-10.

282. Hwang, Y.S., et al., *Microwell-mediated control of embryoid body size regulates embryonic stem cell fate via differential expression of WNT5a and WNT11.* Proc Natl Acad Sci U S A, 2009. **106**(40): p. 16978-83.

283. Zandstra, P.W., et al., *Leukemia inhibitory factor (LIF) concentration modulates embryonic stem cell self-renewal and differentiation independently of proliferation.* Biotechnology and bioengineering, 2000. **69**(6): p. 607-617.

284. Paul, S. and A.J. O'Malley, *Hierarchical longitudinal models of relationships in social networks.* J R Stat Soc Ser C Appl Stat, 2013. **62**(5): p. 705-722.

285. Young, H.P., *The dynamics of social innovation.* Proc Natl Acad Sci U S A, 2011. **108 Suppl 4**: p. 21285-91.

286. Ruths, J. and D. Ruths, *Control profiles of complex networks.* Science, 2014. **343**(6177): p. 1373-6.

287. Di Deco, J., et al., *Machine learning and social network analysis applied to Alzheimer's disease biomarkers.* Curr Top Med Chem, 2013. **13**(5): p. 652-62.

288. Warmflash, A., et al., *A method to recapitulate early embryonic spatial patterning in human embryonic stem cells.* Nat Methods, 2014.

289. Mian, A., et al., *Image set based face recognition using self-regularized non-negative coding and adaptive distance metric learning.* IEEE Trans Image Process, 2013. **22**(12): p. 5252-62.

290. Dowell, K.G., et al., *Novel insights into embryonic stem cell self-renewal revealed through comparative human and mouse systems biology networks.* Stem Cells, 2014. **32**(5): p. 1161-72.

291. Hirai, H., P. Karian, and N. Kikyo, *Regulation of embryonic stem cell self-renewal and pluripotency by leukaemia inhibitory factor.* The Biochemical journal, 2011. **438**(1): p. 11-23.

292. Suwinska, A. and M.A. Ciemerych, *Factors regulating pluripotency and differentiation in early mammalian embryos and embryo-derived stem cells.* Vitamins and hormones, 2011. **87**: p. 1-37.

293. Sasai, M., et al., *Time scales in epigenetic dynamics and phenotypic heterogeneity of embryonic stem cells.* PLoS Comput Biol, 2013. **9**(12): p. e1003380.

294. Lakatos, D., et al., *Autocrine FGF feedback can establish distinct states of Nanog expression in pluripotent stem cells: a computational analysis.* BMC Syst Biol, 2014. **8**: p. 112.

295. Sengupta, T.K., et al., *Rapid inhibition of interleukin-6 signaling and Stat3 activation mediated by mitogen-activated protein kinases.* Proc Natl Acad Sci U S A, 1998. **95**(19): p. 11107-12.

296. Kang, M., et al., *FGF4 is required for lineage restriction and salt-and-pepper distribution of primitive endoderm factors but not their initial expression in the mouse.* Development, 2013. **140**(2): p. 267-79.

297. Abranches, E., et al., *Stochastic NANOG fluctuations allow mouse embryonic stem cells to explore pluripotency.* Development, 2014. **141**(14): p. 2770-9.

298. Chickarmane, V., V. Olariu, and C. Peterson, *Probing the role of stochasticity in a model of the embryonic stem cell: heterogeneous gene expression and reprogramming efficiency.* BMC Syst Biol, 2012. **6**: p. 98.

299. Dietrich, J.E. and T. Hiiragi, *Stochastic patterning in the mouse pre-implantation embryo.* Development, 2007. **134**(23): p. 4219-31.

300. Navarro, P., et al., *OCT4/SOX2-independent Nanog autorepression modulates heterogeneous Nanog gene expression in mouse ES cells.* The EMBO Journal, 2012. **31**(24): p. 4547-4562.

301. Brawand, D., et al., *The genomic substrate for adaptive radiation in African cichlid fish.* Nature, 2014. **513**(7518): p. 375-81.

302. Ding, B., et al., *Quantitative Genetic Analyses of Male Color Pattern and Female Mate Choice in a Pair of Cichlid Fishes of Lake Malawi, East Africa.* PLoS One, 2014. **9**(12): p. e114798.

303. Heule, C., et al., *Genetics and timing of sex determination in the East African cichlid fish Astatotilapia burtoni.* BMC Genet, 2014. **15**(1): p. 140.

304. Kudo, Y., et al., *A microsatellite-based genetic linkage map and putative sex-determining genomic regions in Lake Victoria cichlids.* Gene, 2015.

305. Machado-Schiaffino, G., et al., *Parallel evolution in Ugandan crater lakes: repeated evolution of limnetic body shapes in haplochromine cichlid fish.* BMC Evol Biol, 2015. **15**(1): p. 9.

306. Sylvester, J.B., et al., *Brain diversity evolves via differences in patterning.* Proc Natl Acad Sci U S A, 2010. **107**(21): p. 9718-23.

307. Sanchez-Arrones, L., et al., *Sharpening of the anterior neural border in the chick by rostral endoderm signalling.* Development, 2012. **139**(5): p. 1034-44.

308. Bielen, H. and C. Houart, *BMP signaling protects telencephalic fate by repressing eye identity and its Cxcr4-dependent morphogenesis.* Dev Cell, 2012. **23**(4): p. 812-22.

309. Reichert, S., R.A. Randall, and C.S. Hill, *A BMP regulatory network controls ectodermal cell fate decisions at the neural plate border.* Development, 2013. **140**(21): p. 4435-44.

310. Vermot, J., S.E. Fraser, and M. Liebling, *Fast fluorescence microscopy for imaging the dynamics of embryonic development.* HFSP J, 2008. **2**(3): p. 143-55.

311. Tucker, J.A., K.A. Mintzer, and M.C. Mullins, *The BMP signaling gradient patterns dorsoventral tissues in a temporally progressive manner along the anteroposterior axis.* Dev Cell, 2008. **14**(1): p. 108-19.

312. Rowitch, D.H., *Glial specification in the vertebrate neural tube.* Nat Rev Neurosci, 2004. **5**(5): p. 409-19.

313. Jessell, T.M., *Neuronal specification in the spinal cord: inductive signals and transcriptional codes.* Nat Rev Genet, 2000. **1**(1): p. 20-9.

314. Zhou, Q. and D.J. Anderson, *The bHLH transcription factors OLIG2 and OLIG1 couple neuronal and glial subtype specification.* Cell, 2002. **109**(1): p. 61-73.

315. Qi, Y., et al., *Control of oligodendrocyte differentiation by the Nkx2.2 homeodomain transcription factor.* Development, 2001. **128**(14): p. 2723-33.

316. Wichterle, H., et al., *Directed differentiation of embryonic stem cells into motor neurons.* Cell, 2002. **110**(3): p. 385-97.

317. Snyder, J.L., C.A. Kearns, and B. Appel, *Fbxw7 regulates Notch to control specification of neural precursors for oligodendrocyte fate.* Neural Dev, 2012. **7**: p. 15.

318. Rabadan, M.A., et al., *Jagged2 controls the generation of motor neuron and oligodendrocyte progenitors in the ventral spinal cord.* Cell Death Differ, 2012. **19**(2): p. 209-19.

319. Gokoffski, K.K., et al., *Activin and GDF11 collaborate in feedback control of neuroepithelial stem cell proliferation and fate.* Development, 2011. **138**(19): p. 4131-42.

320. Dias, J.M., et al., *Tgfbeta signaling regulates temporal neurogenesis and potency of neural stem cells in the CNS.* Neuron, 2014. **84**(5): p. 927-39.

321. Kinney, M.A., R. Saeed, and T.C. McDevitt, *Mesenchymal morphogenesis of embryonic stem cells dynamically modulates the biophysical microtissue niche.* Sci Rep, 2014. **4**: p. 4290.

322. Janssens, T., et al., *Charisma: an integrated approach to automatic H&E-stained skeletal muscle cell segmentation using supervised learning and novel robust clump splitting.* Med Image Anal, 2013. **17**(8): p. 1206-19.

323. Schuffler, P.J., et al., *TMARKER: A free software toolkit for histopathological cell counting and staining estimation.* J Pathol Inform, 2013. **4**(Suppl): p. S2.

324. Bartels, P.H., J.E. Weber, and L. Duckstein, *Machine learning in quantitative histopathology.* Anal Quant Cytol Histol, 1988. **10**(4): p. 299-306.

325. Chen, Y.W., et al., *Computer-aided diagnosis and quantification of cirrhotic livers based on morphological analysis and machine learning.* Comput Math Methods Med, 2013. **2013**: p. 264809.

326. Cruz-Roa, A., J.C. Caicedo, and F.A. Gonzalez, *Visual pattern mining in histology image collections using bag of features.* Artif Intell Med, 2011. **52**(2): p. 91-106.

327. Maddah, M. and K. Loewke, *Automated, non-invasive characterization of stem cell-derived cardiomyocytes from phase-contrast microscopy.* Med Image Comput Comput Assist Interv, 2014. **17**(Pt 1): p. 57-64.

328. Nguyen, A.H., et al., *Gelatin methacrylate microspheres for controlled growth factor release.* Acta Biomater, 2015. **13**: p. 101-10.