5-2014

COMPUTATIONAL MODELS OF INTRACELLULAR AND INTERCELLULAR PROCESSES IN DEVELOPMENTAL BIOLOGY

Ahmadreza Ghaffarizadeh
Utah State University

COMPUTATIONAL MODELS OF INTRACELLULAR AND INTERCELLULAR PROCESSES IN DEVELOPMENTAL BIOLOGY

by

Ahmadreza Ghaffarizadeh

A dissertation submitted in partial fulfillment
of the requirements for the degree

of

DOCTOR OF PHILOSOPHY

in

Computer Science

Approved:

Dr. Nicholas S. Flann Major Professor	Dr. Gregory J. Podgorski Committee Member
Dr. Minghui Jiang Committee Member	Dr. Xiaojun Qi Committee Member
Dr. Ilya Shmulevich Committee Member	Dr. Mark R. McLellan Vice President for Research and Dean of the School of Graduate Studies

UTAH STATE UNIVERSITY
Logan, Utah

2014

ABSTRACT

Computational Models of Intracellular and Intercellular Processes in Developmental Biology

by

Ahmadreza Ghaffarizadeh, Doctor of Philosophy

Utah State University, 2014

Major Professor: Dr. Nicholas S. Flann
Department: Computer Science

Systems biology takes a holistic approach to biological questions as it applies mathematical modeling to link and understand the interaction of components in complex biological systems. Multiscale modeling is the only method that can fully accomplish this aim. Mutliscale models consider processes at different levels that are coupled within the modeling framework. A first requirement in creating such models is a clear understanding of processes that operate at each level. This research focuses on modeling aspects of biological development as a complex process that occurs at many scales. Two of these scales were considered in this work: cellular differentiation, the process of in which less specialized cells acquired specialized properties of mature cell types, and morphogenesis, the process in which an organism develops its shape and tissue architecture. In development, cellular differentiation typically is required for morphogenesis. Therefore, cellular differentiation is at a lower scale than morphogenesis in the overall process of development. In this work, cellular differentiation and morphogenesis were modeled in a variety of biological contexts, with the ultimate goal of linking these different scales of developmental events into a unified model of development.

Three aspects of cellular differentiation were investigated, all united by the theme of how the dynamics of gene regulatory networks (GRNs) control differentiation. Two of the projects of this dissertation studied the effect of noise and robustness in switching between cell types during differentiation, and a third deals with the evaluation of hypothetical GRNs that allow the differentiation of specific cell types. All these projects view cell types as high-dimensional attractors in the GRNs and use random Boolean networks as the modeling framework for studying network dynamics.

Morphogenesis was studied using the emergence of three-dimensional structures in biofilms as a relatively simple model. Many strains of bacteria form complex structures during growth as colonies on a solid medium. The morphogenesis of these structures was modeled using an agent-based framework and the outcomes were validated using structures of biofilm colonies reported in the literature.

(110 pages)

PUBLIC ABSTRACT

Computational Models of Intracellular and Intercellular Processes in Developmental Biology

Ahmadreza Ghaffarizadeh

Because living systems arise from coupling different organizational scales (e.g., molecular networks govern cellular states and behaviors, which in turn determine multicellular structures and characteristics), a suitable multiscale modeling framework that incorporates this coupling is needed to predict the effects of molecular level perturbations (such as mutations) on cellular and multicellular behaviors. Such models will be vitally important for understanding the molecular basis of embryogenesis and physiology as well as diseases, particularly cancer. In turn, this deeper understanding is essential for developing rational therapeutic strategies intended to alter clinical outcomes, such as reduction of tumor invasiveness or angiogenesis, through molecular targeting.

Having a clear understanding of processes at different scales is the first step in multiscale modeling. The research presented in this dissertation focuses on studying different aspects of biological development at two scales to provide some useful platforms for the construction of future multiscale models of developmental biology. Specifically, this work studied network dynamics-driven cellular differentiation at lower scale (for example, in myeloid and pancreatic cell differentiation) and examination of intracellular interactions in morphogenesis (emergence of three dimensional structures in biofilms) at higher scale.

ACKNOWLEDGMENTS

Research reported in this dissertation was supported by the National Institute Of General Medical Sciences of the National Institutes of Health under Award Number P50GM076547, Luxembourg Centre for Systems Biomedicine, the University of Luxembourg and the Institute for Systems Biology, Seattle, USA. Thanks to Ilya Shmulevich and Merja Heinaniemi for helpful discussions and to Jan-Ulrich Kreft for providing the original version of iDynoMiCs. The content is solely the responsibility of the authors and does not necessarily represent the official views of the National Institutes of Health.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

1.1 Multiscale modeling

A wide variety of processes occur at different levels of resolution and complexity. This difference can stem from the nature of these processes or just as a consequence of the measurements used for quantifying interconnected levels. Physical laws that work at each level define the temporal and spatial scales. As an example, consider a river flowing into its delta. At the macro scale, the flow of the river can be described by physical laws of the motion of fluids like Navier-Stokes equations using minutes and meters as time and space scales. Then the focus can be shifted to the water drops and their interactions with the environment as the meso scale. Here the time and space scales are of the order of magnitude 10^{-9} second and 10^{-3} meter, respectively. Continuing this hierarchy, at the micro scale, a water molecule can be described where the time and space scale become very small [7]. Note that one may define many other scales for the river flow process ranging from different temporal and spatial scales or change the view by describing water drops at the macro scale, water molecules as the meso scale and the covalent bonds in a water atom as the micro scale.

Multiscale modeling of complex systems has gained a lot of attention in the past decades as the result of significant progress of computational capacity. A broad range of topics is covered by multiscale modeling including the economy [8,9], ecology [10–12], material engineering [13–15], and biology [16–18]. The main challenge in all multiscale modeling approaches is tradeoffs between accuracy and computational complexity. The more details included in the model, the more accurate model will likely be, but at the cost of computational complexity.

Multiscale modeling approaches can fall in two main groups:

1. Sequential multiscale modeling: in this approach, sometimes referred to as microscopically-informed modeling, the result of one scale is obtained and then fed into an upper scale. Thus, there is no signaling or communication between distinct scales. The pros of this approach center around the independency of scales where the output of each level can be verified and validated before being passed to the upper level; also it makes it easier to implement and debug the system. On the other hand, the independence of scales turns to be its weak point too: this approach suffers from a lack of integrity. This type of modeling has proven effective where different scales work independently; as a result the vast majority of multiscale simulations that are in use are sequential. For some examples, see [19–22].

2. Parallel multiscale modeling: this approach, also called as concurrent multiscale modeling, is a more realistic one where different scales run simultaneously. There are feedback loops between different levels which result in an integrated system. Each change at each scale will gradually be passed to the neighbor scales and subsequently to the whole system. While this approach is closer to how the nature works, it is difficult to implement. Usually the scales are widely far from each other in terms of time or space. Therefore, the challenge of this approach is to find the proper inter-scale modeling to bridge the gaps between different scales. For some examples, see [17, 23–25].

1.2 Multiscale Modeling in Biology

Biological systems, at any level of abstraction, have roots in the complex interactions of cell systems. The behavior of each cell is regulated by its genome. In turn, the genome follows what is dictated by its molecular networks. Thus, a biological phenomenon can be seen as a collection of components which function at different scales. Studying the interaction of these components and the way they give rise to the function of a system is one of the main conceptual challenges in biology. Figure 1.1 shows a paradigm of multiscale modeling in biology that spans from micro seconds (10^{-6}s) to years ($10^{8}s$), and from nano meters (10^{-9}m) to meters. Each scale is coupled to the strata above and below it. Understanding

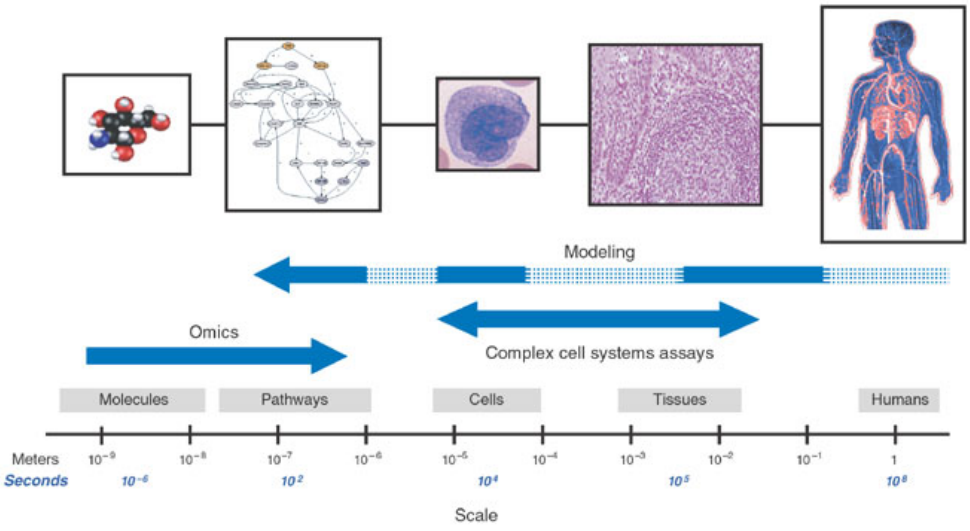


Figure 1.1. Multiscale interactions in biology (from [1]).

the inter-scale and intra-scale interactions between these diverse scales is critical to studying physiology and the treatment of diseases. Martins et al. [26] suggested categorizing events in the human body into 3 different scales starting with the microscopic scale. Gene regulation, signaling, metabolic pathways, and cell cycles are some sample of events at the microscopic scale. The next level is the mesoscopic scale which mainly deals with cell-cell and cell-matrix interactions. The last scale is the macroscopic level which includes processes at the tissue level.

The complexity of biological systems makes their simulation very complicated in terms of computation and interpretation. That is why most of the multiscale modeling frameworks in biology are implemented as sequential multiscale systems where discrete scales are separated by gaps which are usually bridged with mathematical prediction and estimation techniques [16].

1.2.1 Top-down and Bottom-up Approaches

Biological systems are usually modeled and interpreted in two ways: top-down or bottom-up [27]. The bottom-up approach starts with simulation of system components

in isolation. These components are later integrated to study the emergent behavior of the model and compare it with the observed behavior of the target system. As an example, consider the process of morphogenesis that causes an organism to develop its shape. One may employ Turing's 2-component reaction-diffusion theory of morphogenesis [28] to study the non-uniformities commonly seen in nature [29–33]. This theory explains how the concentration of two substances distributes and reaches a steady state over time. In this theory, one component, termed activator, stimulates the production of activator and inhibitor, while the inhibitor prevents their production. To create such a model, the activation/inhibition functions first need to be constructed, and then be combined to give rise to pattern formation.

Top-down approach starts by investigating the macroscopic behaviors of the system. These behaviors are then used to construct a model which can describe these high level properties. The main advantage of top-down approaches is their simplicity; they usually use a high level abstract description of the underlying mechanisms of the modeled phenomena, thus they do not consider many irrelevant details that have small or no effect on system behavior. Studying the formation of extracellular matrix (ECM) in multicellular organisms is an example of a top-down approach. Extracellular matrix is a non-cellular component that provides structural support to the cells and also plays important roles in activating signalling pathways and many critical biochemical interactions. Without getting trapped in modeling the complicated process of biomechanical adhesion, one may use the emergent behavior of this adhesion to describe the stages in formation of extracellular matrix.

There is another emerging approach called middle-out [34] which focuses on one scale and then starts linking the lower and upper scales to the collection of scales used in the modeling. In this dissertation, we use a middle-out approach to model and study some important phenomena in developmental biology that are important in human physiology and disease treatment. This dissertation focuses on the linkage of subcellular to cellular scales in modeling cellular differentiation and in turn on the coupling the cellular and multicellular scales to model morphogenesis.

Regulation type inference in gene regulatory networks, modeling and visualizing cell type switching, and studying the role of multistable switches in providing robustness to terminally differentiated cell types are the three research topics at the level of cellular differentiation that are covered in this dissertation. In a complementary work, wrinkle formation in biofilms was modeled as an example of morphogenesis and the outcomes of the modeling were validated against experimental findings. The studies presented in this dissertation can serve as steps towards multiscale studies in systems biology.

CHAPTER 2
REGULATION TYPE INFERENCE IN GENE REGULATORY NETWORKS

2.1 Abstract

Inferring gene regulatory networks from high-throughput microarray expression data remains a challenging problem in systems biology, mostly due to underdetermined nature of GRN inference. In order to cope with this problem, we constrain the solution space of GRNs by employing nested canalyzing functions. We use a genetic algorithm with nested canalyzing functions to search for candidate GRNs that allow the differentiation of specific cell types. The set of candidate GRNs is then used to infer the type of regulation (activation or inhibition) between particular genes. We use this method to evaluate two hypothetical GRNs previously proposed for myeloid and pancreas cellular differentiation. Finally, we propose a set of updating rules that can be used for future studies of myeloid and pancreas gene regulatory networks.

2.2 Introduction

The ultimate goal of systems biology is to obtain a blueprint of gene and protein interactions at the subcellular scale that give rise to the characteristics observed at cellular or organismal levels. Although it seems that we are at the beginning of a long way towards reaching this ultimate goal, the availability of high-throughput micro array data and the effectiveness of automated literature mining tools have provided hope that information may be coming available for systems biologists to reach this goal sooner rather than later. With this rapid growth of knowledge sources, the essence of accurate inference of gene regulatory networks (GRNs) seems more crucial than ever; the burst in the number of research studies

recently carried out in this field is an evidence for this claim.

Selection of a proper network architecture is the first step in modeling gene regulatory interactions. There are many network architectures suggested for modeling GRNs. In all of these networks nodes are genes or transcription factors. However, the formalism behind each architecture is different. Generally, these modeling approaches fall into 4 groups [35]: 1) Information theory models, 2) Bayesian Networks, 3) Boolean Networks, and 4) systems of equations. Each of these approaches has its own principals, assumptions and limitations. Based on the criteria defined for a problem, any of these approaches may be employed for GRN modeling. The tradeoffs are between scalability, simplicity, parameterization and the amount of required information for modeling [36]. Due to the simplicity and the capability of Boolean networks in modeling GRN dynamics without needing knowledge of any kinetic parameter [3], we used Boolean networks as our modeling framework in this study.

First proposed by Kauffman [37], Boolean networks are a major contributor to our knowledge of gene regulatory networks. They are dynamical networks which use discretized values for gene expression levels of each node with values set to 1 or 0. Each node is associated with a Boolean function which specifies the output of the node as on (1) or off (0) based on the corresponding inputs. The main advantage of Boolean networks is their simplicity; the influences are simply inhibition or activation without any quantification.

Kauffman [37] suggested that cell types are attractors in dynamics of Boolean networks simulating GRNs. He later proposed that these Boolean networks need to be robust against the intrinsic noise of the system in order to be able to model GRNs appropriately. Noise robustness comes from canalyzing functions, where the value of one input determines (canalyzes) the output no matter what other inputs are. Canalyzing functions shift the dynamics of the system from the chaotic to the critical domain where GRNs are believed to operate [38,39]. Nested canalyzing functions (NCFs) are a generalization of canalyzing functions where all inputs canalyze the output with an ordering defined for the function [40].

Identifying GRNs for known differentiation systems is a complex and challenging task. Too much information of physical and regulatory interactions between genes should be inte-

grated and compiled to a network whereas the dynamics of this network should simulate the differentiation process. Extracting data from sometimes contrasting reports makes this task even more complicated. In this work we use a searching mechanism to find Boolean networks with nested canalyzing functions which can reconstruct known GRNs in developmental biology. We employ a genetic algorithm to match the attractors of our inferred GRN against the expression profiles obtained from experimental data. We use an ensemble of candidate solutions to identify the type of each regulatory interaction (inhibition or activation) in our target GRN.

2.3 Boolean Networks as GRNs

Boolean networks [37] have been used for representing GRN structure and dynamics in many systems, including *Drosophila* development [41, 42], angiogenesis [43], eukaryotic cell dynamics [44], and yeast transcription networks [40]. Each node in a network represents a gene whose activity is regulated by an internal function based on inputs from other nodes. The output of each node is either the value true, representing an expressed gene, or the false, representing a non-expressed gene.

A Boolean network with K genes has 2^K possible states, denoted as $\hat{S}$. At each step in the simulation, the next state $\hat{s}_{t+1} \in \hat{S}$ is determined by applying each gene's logic function (representing the regulatory interactions) to the current value of the genes in $\hat{s}_t$. Let this computation be defined as $\hat{s}_{t+1} \leftarrow D(\hat{s}_t)$ where $D(\hat{s}_t)$ is the deterministic mapping function that finds the next state of the network given the current state. As the network is executed by repeated applications of $D(\hat{s})$, the state will reach a previously visited state, and thus, since the dynamics are deterministic, enter into an attractor which represents a fixed point of the system. Attractors can be single states, called point attractors, or consist of more than one state that the network continuously transitions between, called cyclic attractors. Let $\hat{a} = D(\hat{s})^*$ be the resulting network attractor state reached when starting at $\hat{s}$ and applying the logic functions until the attractor state $\hat{a}$ is reached [45].

In this work, attractors of the high dimensional state space of possible gene expression profiles are interpreted as distinct cell types [46]. The dynamical behavior of system in

transitioning from one attractor to another is employed to model cell differentiation [47].

2.4 Problem Description

Suppose that there is a cellular differentiation subtree that contains n cell types, each referred to as $C_i, i = 1...n$. Cell types are represented by expression profiles of a set of K genes identified that play a role in differentiation of these cell types $C_i \leftarrow \{e_1, ..., e_K\}$. The general problem is how to find a Boolean network that generates single or cyclic attractors that can be mapped to all the cell types in the target differentiation tree. In this way, the functions of the nodes specify the regulatory interactions between the genes of the regulatory network. Thus the objective of this problem is to minimize the error of model fitting as the difference between the observed gene expression levels of cell types from experimental data (C) and modeled GRNs (C').

$$objFun = \sum_{i=1}^{n} dist\left(C_i, C_i'\right) \tag{2.1}$$

Note that the ordering of attractors are not important and each cell type C is evaluated against its corresponding attractor.

2.4.1 Approach

A search mechanism is needed to find a Boolean network that can model a GRN. Since the state space of all possible Boolean networks is huge, the search mechanism must be efficient. As mentioned before, we use a genetic algorithm (GA) for our optimization algorithm. To use the GA, many settings need to be adjusted. Two of the most important are: 1) how to represent a GRN as a solution; and 2) How to evaluate a solution.

Solution Representation

Suppose that the GRN is composed of K interacting genes. To simplify the problem, we apply the maximum connectivity degree U constraint so that each gene can have U inputs. One intuitive way for representing a solution is to let the function of each gene have 2^U

entries. The function specifies the output of the gene as expressed or repressed for each input combination. The following table shows a sample gene function with 3 inputs.

inp_1	inp_2	inp_3	f
0	0	0	0
0	0	1	0
0	1	0	1
0	1	1	0
1	0	0	1
1	0	1	1
1	1	0	1
1	1	1	0

Each gene also needs to be sensitive to where its connection come from, so there are K binary numbers, each showing a connection from kth gene ($k = 1, ..., K$). For each gene we have 2^U entries for representing the function plus K connections, therefore the size of solution representation is $K * (2^U + K)$ bits.

Though this representation seems straightforward and has been used in some previous studies for inferring GRNs, we found it not suitable in our approach. In the Results section, we show that GRN inference is an under-determined problem; therefore, to direct our search method we need to impose constraints on the solution structure. We consider two constraints in our approach: 1) we fix the structure of GRN based on reports from the literature, and 2) we use NCFs. Even with these constraints, there are often many possible solutions.

To encode a solution considering the added constraints, we use the scheme shown in Figure 2.1. In this scheme, instead of directly encoding each gene's function to the solution, we use metafunctions. The size of solution is $\sum_i^K |MF_i|$ where $|MF_i|$ is the number of bits needed to represent the meta function for ith gene. As Figure 2.1 shows, metafunction i has a size $2m_i + [log_2(m_i!)] + 1$ where m_i is the number of inputs to the gene. The first $2m_i$ bits in MF_i create a mapping between the function's inputs (canalyzers) and its output. We also need to know how canalyzing effects are nested, e.g., what is the order in which the inputs are affecting the output? Since there are $m_i!$ possible permutations in which inputs can be arranged, we need $[log_2(m_i!)] + 1$ bits to choose one of these permutations (starting from $[1, 2, ..., m_i]$ to $[m_i, m_i - 1, ..., 1]$). The values of bits at locations j and $m_i + j (1 \leq j \leq m)$ of

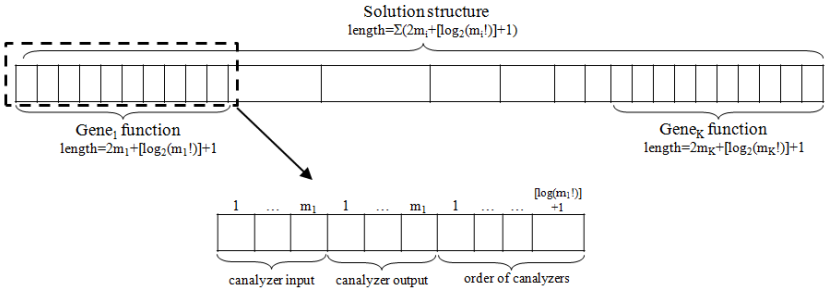


Figure 2.1. Solution structure representation using nested canalyzing functions.

ith metafunction indicate how the jth input affects the output of the ith gene. For example, if for an arbitrary gene, the first $2m$ bits of MF are 010110, it indicates that the value 0 for the first input turns on the gene; the value 1 for the second input turns on the gene too; and the value 0 for the third gene turns the gene off; so we can deduce that the first input has an inhibitory effect while the next two are activators of this gene.

Solution Evaluation

We use the function shown in Equation 2.1 to evaluate the fitness of the each solution. The City Block distance is employed as the measure of similarity between expression profile of a cell type and its corresponding modeled profile as an attractor in the dynamics of an inferred GRN. Note that the mapping of cell types is a one-to-one mapping and each cell type is mapped to the attractor with the least distance with tie broken randomly. Extra numbers of attractors are punished with a penalty P_{ext}. This constraint is added to prevent the overfitting by the model. Therefore, the *error* function (the difference between experimental data and modeled profiles) is formulated as:

$$error = \sum_{i=1}^{n} \sum_{k=1}^{K} \left| e_{ik} - e'_{ik} \right| + |n - n'| \times P_{ext} \tag{2.2}$$

where e_{ik} is the expression level of the kth gene in ith cell type from the experimental data, and e'_{ik} is the expression level of the modeled cell type, n is the number of cell types in the target differentiation process and n' is the number of attractors in the inferred GRN. The

maximum error is nK (considering $P_{ext} = 0$) while the minimum is 0 for a perfect fit. The GA solves the minimization problem.

2.5 Results

2.5.1 Test Case 1: Pancreatic Cell Differentiation

The pancreas is as an endocrine gland that secretes many important hormones in the digestive and endocrine systems. By regulating blood glucose level through secreting insulin, pancreas beta cells play a key role in controlling metabolism; deficiency in the operation of these cells results in the diabetes. The interesting properties of the pancreas cellular differentiation have made it the focus of many studies [2, 48–52]. A partial subtree of pancreatic differentiation is shown in Figure 2.2. This figure shows the value of 5 key transcription factors (TFs) in exocrine, β/δ progenitor, and α/PP progenitor cells suggested by [2]. $Pdx1$ determines the onset of pancreatic development, however, this gene is only regulated in β/δ progenitor during the later stages of development. $Ngn3$ and $Ptf1a$ form a bistable switch which determines cell fate in the exocrine/endocrine bifurcation. For cells committed to endocrine lineage, the mutual interaction of $Pax4$ and Arx in another bistable switch determines whether a cell becomes either a β/δ progenitor or ab α/PP progenitor. The solid edges in Figure 2.3 represent experimentally confirmed interactions while the dashed lines are the proposed gene interactions.

We used the discretized gene expression levels of each cell type in the pancreatic differentiation subtree (shown in Figure 2.2) to infer a gene regulatory network using the method described in the previous section. As the first step, we ran our search algorithm 100 times using the original solution presentation (unconstrained scheme), and recorded solutions the algorithm converged to. To see how similar these solutions are, we compared each result with a hypothetical inferred GRN from [2] shown in Figure 2.3. We assigned each solution a score based on the number of false positive and false negative errors relative to the target GRN. Note that we just consider the presence or absence of an interaction for computing the score, not whether this interaction is positive or negative. The results showed (not presented here)

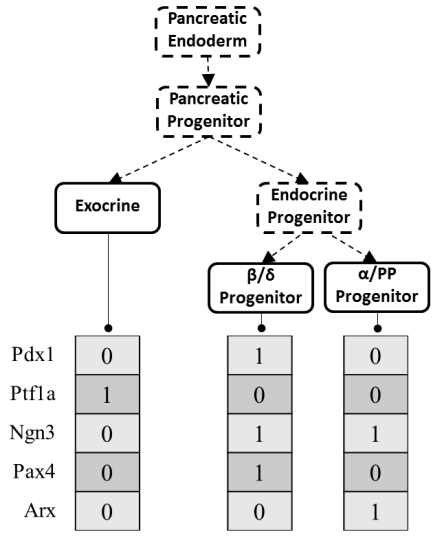


Figure 2.2. Pancreas cellular differentiation subtree and discretized expression values for 5 important genes that play a role in this differentiation.

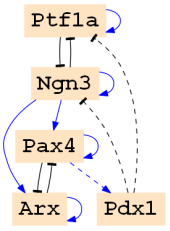


Figure 2.3. Hypothetical pancreas gene regulatory network from [2]. Arrows and closed lines show activation and inhibition influence respectively. The solid lines represents experimentally confirmed interactions while the dashed lines are proposed gene interactions by [2].

that there are variety of solutions with similar scores but very different structures. These structural differences are also present in cases with perfect matches of attractors to target cell types. This variability stems from the fact that the problem is so underdetermined.

To address the problem of underdetermination, we used the constrained solution scheme described in Section 2.4.1. We reduced the problem to search for nested canalyzing functions that can form attractors with a close match to target cell types in pancreatic differentiation. Note that a solution will determine both the order and type of influence (activa-

Table 2.1. List of research studies that confirm/propose the type of interactions between pancreatic transcription factors shown in Figure 2.2. Red references have reported inhibitory influence while blue ones have reported activation. Star shows theoretically proposed influences in previous studies that are not experimentally validated. [S] denotes our proposed corrections on previously reported influences.

	Pdx1	**Ptf1a**	**Ngn3**	**Pax4**	**Arx**
Pdx1	-	[2]*	[2]* [53] [S]	-	-
Ptf1a	-	[54]	[52]	-	-
Ngn3	-	[52]	[55] [56]	[52]	[49]
Pax4	[2]*	-	-	[2] [57]	[58]
Arx	-	-	-	[58]	[2]

tion/inhibition) each gene imposes on the other. Again we ran our search algorithm 100 times. It was interesting to see that all runs converged to a prefect fit solution in less than 30 iterations After comparin the solutions and removing the identical ones, there were 95 distinct solutions. To integrate the results obtained in this step, we extracted the inhibition and activation interactions from the results and compared the number of times in which an interaction was considered as activatiing versus the number of times it was considered as inhibiting. Results are shown in Figure 2.4. We hypothesize that for each regulatory interaction, the most frequent influence can predict the nature of this regulation in the inferred GRN. Figure 2.4 demonstrates the power of our method in evaluating the hypothetical GRN for pancreatic cell differentiation proposed by Zhou et al. [2]. Our results confirm all experimentally validated interactions. Table 2.1 lists the previous research studies that support each particular interaction; references that have reported the interaction as inhibitory are colored red and references that have reported activation are colored blue (our proposed influences are noted with [S]). It is striking that there are contradictory reports for some interactions.

For the interactions proposed by Zhou et al. [2] (dashed edges in Figure 2.3), our method confirms inhibitory interaction between Pdx1 and Ptf1a and also the activation of Pdx1 by Pax4; in contrast, based on our results, we predict that the inhibitory interaction between Pdx1 and Ngn3 is actually an activation. Significantly, our prediction was validated in a previous report by Oliver-Krasinski et al. [53].

For a small test case like the GRNs for pancreatic cell differentiation, an exhaustive

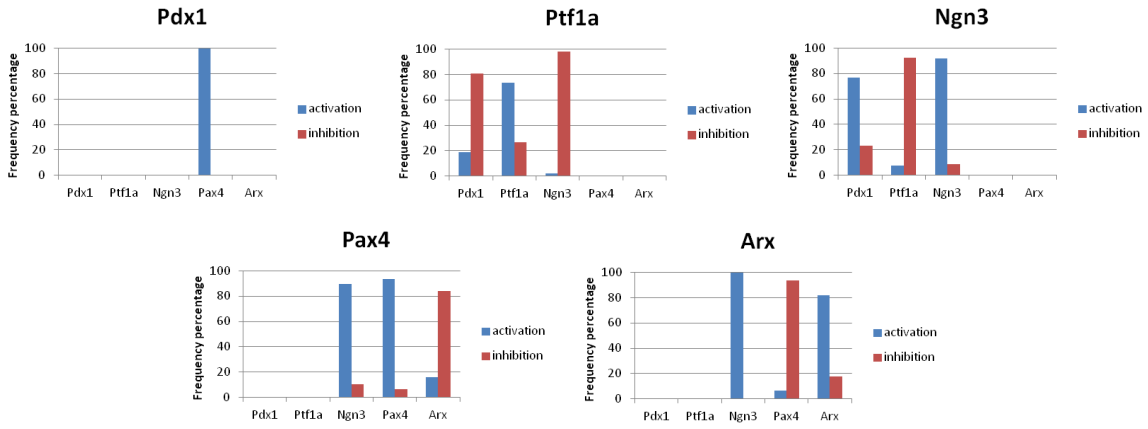


Figure 2.4. Percentage of presence of regulatory interactions either as activatory or inhibitory influence in pancreas differentiation. Results are averaged for 95 Boolean networks that their dynamics have attractors that can be matched to cell types shown in Figure 2.2.

search instead of a genetic algorithm is possible. In this example, the number of possible solutions is of the order 10^7 which also includes many redundant cases. However, for most of GRNs, like the next test case, the solution space is huge and it is not possible to examine all possible networks.

2.5.2 Test Case 2: Myeloid Cell Differentiation

Hematopoiesis is a well characterized example of cellular diversification. During hematopoiesis, hemocytoblast gives rise to blood precursor cells (common myeloid progenitors (CMPs)) and lymphocyte precursor cells. Figure 3.1 shows a sub-tree of myeloid lineage tree in which the CMP differentiates into megakaryocyte-erythrocyte precursor (MEP) cells and granulocyte-monocyte precursor (GMP) cells. Depending on extracellular environment, MEP can give rise to erythrocytes (red blood cells) or megakaryocytes (platelets). The GMP can differentiate to granulocytes or monocytes [3, 59, 60].

Informed by work of Krumsiek et al. [3], we picked 11 transcription factors known to play a role in myeloid differentiation: GATA-1, GATA-2, FOG-1, EKLF, Fli-1, SCL, C/EBPα, PU.1, cJun, Gfi-1, and EgrNab (EgrNab represents an integration of Egr-1, Egr-2 and Nab-

2). Figure 3.1 shows a subtree of myeloid differentiation and the discretized expression levels for 4 differentiated cell types. Using these expression profiles, we repeated the study done in the test case 1. Again, for the unconstrained search, we discovered a set of solutions with a very low similarity (results not shown). Similar to test case 1, we reduced the problem using the described constrained solution scheme. We also added information from a knockout experiment to the problem. GATA1-PU.1 is a tristable switch that controls the bifurcation of CMP cells between MEP and GMP lineages. Expression of GATA-1 represses PU.1 and pushes the cell to the MEP lineage; in contrast, when PU.1 is expressed it represses GATA-1 and the cell commits to the GMP lineage. We expected to see that knocking out PU.1 leads to disappearance of granulocytes and monocytes from the dynamics of our inferred GRN while GATA-1 knockout removes megakaryocytes and erythrocytes from the inferred GRN. Here, we rewrite Equation 2.2 to integrate information coming from multiple sources.

$$error = \sum_{e=1}^{E} \left(\sum_{i=1}^{n_e} \sum_{k=1}^{K_e} \left| e_{ik_e} - e'_{ik_e} \right| + |n_e - n'_e| \times P_{ext} \right) \tag{2.3}$$

In this equation, E is the total number of experiments used as different sources of information. Using this new formulation, we ran the constrained minimization problem 100 times. For this test case, only 11 runs converged to a perfect fit. We repeated the experiment multiple times to get 100 distinct solutions and averaged the results to identify the type of interaction between transcription factors in our target GRN. Figure 2.7 shows the averaged results of influences between genes. The results show excellent agreement with interaction information extracted from literature. Our results propose three corrections to regulatory interactions reported in the previous studies: a) the inhibitory influence of GATA-1 and its cofactor (FOG-1) on GATA-2 should be replaced with an activating influence. This activation might account for the partial expression of GATA-2 in erythrocytes where GATA-1 is fully expressed; b) the autoregulation of PU.1 should be negative; and c) the influence that Fli-1 imposes on EKLF may be of type of activation. This also might account for partial expression of EKLF during megakaryocyte differentiation where Fli-1 is fully expressed. Table 2.2 summarizes references that confirm each interaction. Again, references that have

Table 2.2. List of research studies that confirm/propose the type of interactions between myeloid transcription factors shown in Figure 3.3. Red references have reported inhibitory influence while blue ones have reported activation. Star shows theoretically proposed influences in previous studies that are not experimentally validated. [S] denotes our proposed corrections on previously reported influences.

	GATA-1	**GATA-2**	**FOG-1**	**EKLF**	**Fli-1**	**SCL**	**C/EBPα**	**PU.1**	**cJun**	**EgrNab**	**Gfi-1**
GATA1	[61] [62]	[63] [64][S]	[65]	[66]	[67]	[68]	[3]*	[69] [70] [71]	-	-	-
GATA2	[72]	[72] [63]	-	-	-	-	-	[69] [70] [71]	-	-	-
FOG-1	-	[72] [63][S]	-	-	-	-	[3]* [73]*	-	-	-	-
EKLF	-	-	-	-	[74]	-	-	-	-	-	-
Fli-1	[74] [75]	-	-	[74][S]	-	-	-	-	-	-	-
SCL	-	-	-	-	-	-	[3]*	-	-	-	-
C/EBPα	-	-	-	-	-	-	[3]*	[76] [77]	-	-	[78] [79]
PU.1	[70] [71]	[70] [71]	-	-	-	[80]	-	[81] [82][S]	[83]	[78]	-
cJun	-	-	-	-	-	-	-	-	-	[78]	-
EgrNab	-	-	-	-	-	-	-	-	-	-	[78]
Gfi-1	-	-	-	-	-	-	-	-	[84]	[78]	-

reported the interaction as an inhibitory influence are colored red and the references that have reported the activatory influence are colored blue.

2.5.3 Proposed GRN Updating Rules

Based on the results obtained for the test cases and from the interaction information extracted from literature, we propose a set of updating rules which can construct a GRN with attractors corresponding to the cell types of each test case. Table 2.3 and 2.4 present these updating rules for pancreas and myeloid differentiation, respectively. Note that these are not the only updating rules that can produce the same set of attractors; they are provided here simply to serve as a possible starting point for future studies. Compared to updating rules proposed in previous studies, our proposed rules do not produce extra attractors and the network dynamics are supported by the knockout experiments.

Table 2.3. A sample set of logical functions for pancreatic transcription factors shown in Figure 2.3.

Gene Name	**Update rule**
Pdx1	$Pax4$
Ptf1a	$\overline{Ngn3}.\overline{Pdx}$
Ngn3	$(Pdx1 + Ngn3).\overline{Ptf1a}$
Pax4	$(Ngn3 + \overline{Arx}).Pax4$
Arx	$Ngn3.\overline{Pax4}$

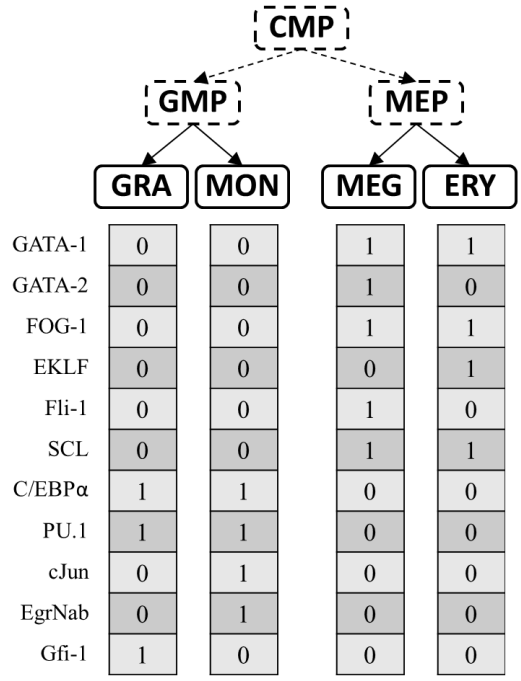


Figure 2.5. Myeloid differentiation tree and discretized expression values for 11 transcription factors that control the differentiation process. Terminal nodes are the mature cell types of Erythrocytes (ERY), Megakaryocytes (MEG), Monocytes (MON), and Granulocytes (GRA). Multipotent cells are the common myeloid progenitor (CMP), megakaryocyte-erythrocyte precursor (MEP), and granulocyte-monocyte precursor (GMP).

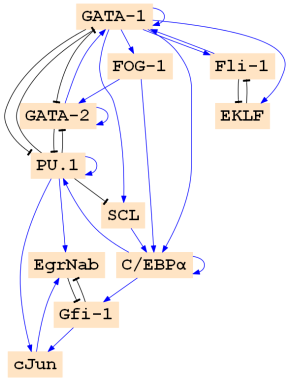


Figure 2.6. The inferred genetic regulatory network for myeloid differentiation. Nodes are eleven key transcription factors that control cell lineage and edges are the interactions between the genes.

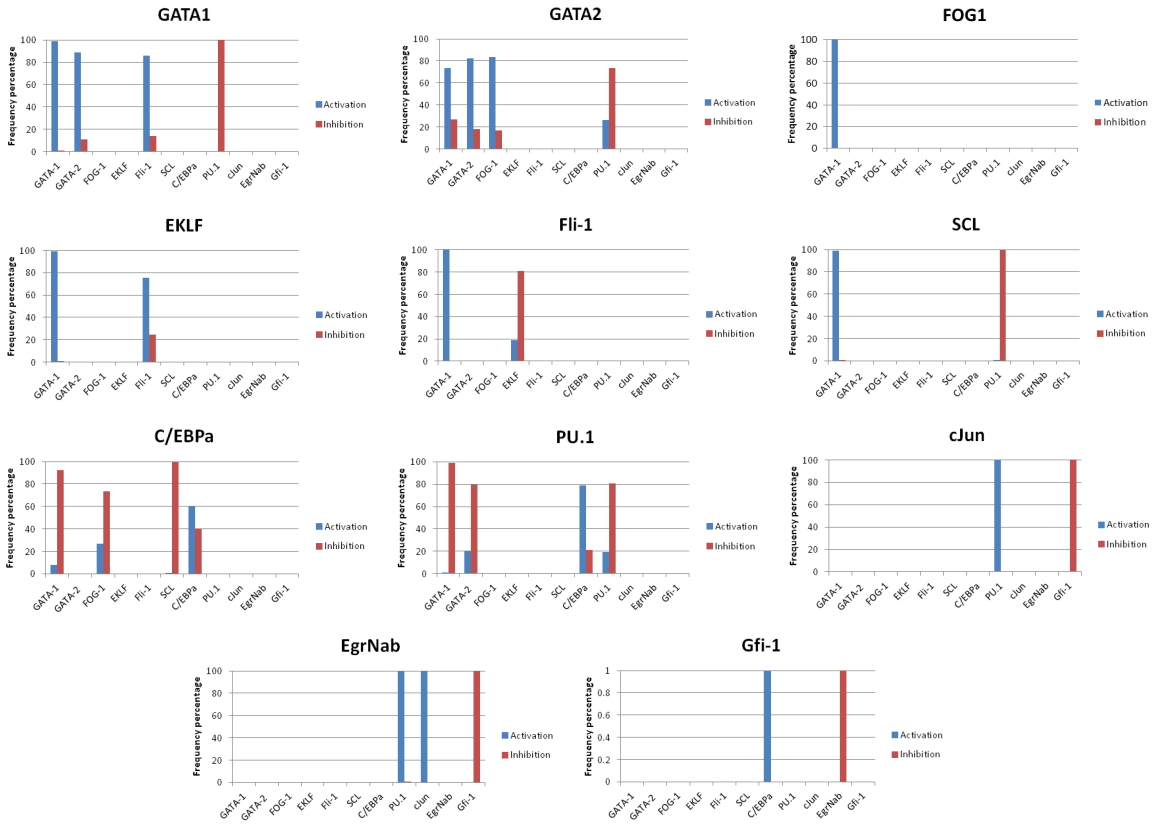


Figure 2.7. Percentage of presence of regulatory interactions either as activatory or inhibitory influence in myeloid differentiation process. Results are averaged for 95 Boolean networks that their dynamics have attractors that can be matched to cell types shown in Figure 3.1.

2.6 Conclusion

Boolean networks have proved effective in explaining and modeling cellular differentiation. They are very useful in studying the interactions between genes and for analyzing the dynamics of regulatory networks. Recent advances in the availability of high throughput data and extensive literature discussing different regulatory interactions have fostered interest in inferring GRNs based on this information. The major barrier to effective inference of GRNs is underdetermination. Even when many sources of information are integrated, there are many possible solutions that can simulate the dynamics of biological target GRN. In this work, we used the strong constraint of nested canalyzing functions on the space of possible

Table 2.4. A sample set of logical functions for myeloid transcription factors shown in Figure 3.3.

Gene Name	**Update rule**
GATA-1	$GATA\text{-}1 \vee GATA\text{-}2 \vee Fli\text{-}1 \vee \overline{PU.1}$
GATA-2	$GATA\text{-}1 \wedge GATA\text{-}2 \wedge FOG\text{-}1 \wedge \overline{PU.1}$
FOG-1	$GATA\text{-}1$
EKLF	$GATA\text{-}1 \wedge \overline{Fli\text{-}1}$
Fli-1	$GATA\text{-}1 \wedge \overline{EKLF}$
SCL	$\overline{PU.1}$
C/EBPα	$(\overline{GATA\text{-}1} \wedge \overline{SCL}) \vee (\overline{FOG\text{-}1} \wedge \overline{SCL} \wedge C/EBP\alpha)$
PU.1	$\overline{GATA\text{-}1}$
cJun	$PU.1 \wedge \overline{Gfi\text{-}1}$
EgrNab	$(PU.1 \wedge \overline{EKLF}) \vee (PU.1 \wedge cJun)$
Gfi-1	$C/EBP\alpha \wedge \overline{EgrNab}$

Boolean networks to reduce the inference problem and to find the regulatory interactions in a predefined GRN structure. A genetic algorithm is employed in this work to search the solution space of nested canalyzing functions. We averaged the solutions and compared the frequency in which a regulatory interaction was activating versus inhibitory. We hypothesized that the favored type of influence reveals the nature of this interaction. The two test cases we studied are in excellent agreement with our hypothesis. Finally, we proposed a set of updating rules that can be used for future studies of myeloid and pancreas differentiation.

CHAPTER 3

MODELING AND VISUALIZING CELL TYPE SWITCHING [1]

3.1 Abstract

Background: Understanding cellular differentiation is critical in explaining development and for taming diseases such as cancer. Conventionally, cellular differentiation is visualized as bifurcating lineage trees. However, these lineage trees cannot readily capture or quantify all the types of transitions between cell types that are now known to occur. For example, "terminally differentiated" cells can be reverted directly to a pluripotent state, adult cells can be induced to transdifferentiate, even across germ layer lineages, and in cancer, cells often acquire specialized properties not seen on any standard lineage tree. None of these transitions can easily be represented by a conventional hierarchically-arranged lineage tree and conventional trees do not show the likelihood of any transition.

Results: This work introduces a new analysis and visualization technique (with a supporting tool called *CellDiff3D*) that is capable of representing and visualizing all possible transitions between cell states compactly, quantitatively, and intuitively. *CellDiff3D* takes as input a regulatory network of transcription factors that control cell type switching, then performs an analysis of network dynamics to identify stable expression profiles and the potential cell types they may represent. *CellDiff3D* creates a three dimensional graph that shows the overall direction and likelihood of transitions between pairs of cell types within a lineage. In this visualization, the distance between a pair of cell types measures the likelihood of transitions between them, with greater distances indicating lower probabilities. Arrows between cell types show the favored direction of the transition, with the thickness

[1] A. Ghaffarizadeh, G. J. Podgorski, and N.S. Flann,"Modeling and Visualizing Cell Type Switching," *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 293980, 10 pages, 2014. doi:10.1155/2014/293980

of the arrow representing the relative rate. Therefore, probabilities and rates of transitions within a lineage are quantified in the *CellDiff3D* graph. In this work, the influence of gene expression noise and mutational changes in myeloid cell differentiation are presented as a demonstration of the *CellDiff3D* technique. The supporting software can be downloaded from www.CellDiff3D.org.

Conclusions: As new complexities in cellular differentiation are being recognized, more powerful analysis and visualization approaches are needed. Our technique is an innovative approach that quantities, represents and visualizes all possible cell state transitions in any given regulatory network.

3.2 Introduction

During development, a complex system of tissues and organs emerges from a single cell by the coordination of cell division, morphogenesis, and differentiation. Understanding the differentiation of cell types is necessary to understanding development and its associated defects, for improved control of stem cell differentiation in therapeutic use, and for taming diseases such as cancer. Cellular differentiation occurs when a less specialized cell or its progeny become increasingly specialized by acquiring properties that allow specific functions. In animals, differentiation typically results in a terminally differentiated state in which a specialized cell can no longer acquire the properties of other specialized adult cells. Recent discoveries, however, have shown that terminally differentiated cells can be reprogrammed to revert back to multipotent and pluripotent stem cells which have the potential to differentiate into other cell types [5, 85] or to transdifferentiate into other specialized cell types [86].

Differentiating cells normally follow well defined paths to mature cell types. Taken together, these paths are referred to as a lineage tree. Pluripotent stem cells give rise to progeny that specialize into more constrained multipotent cells. In turn, multipotent cells produce a variety of stable, terminally differentiated cells. This process is usually depicted as a tree with a pluripotent cell at its root, multipotent cells as intermediate nodes, and the mature cell types as branch tips. As an example, a simplified portion of the myeloid cell lineage tree is illustrated in Figure 3.1. This figure shows that common myeloid progenitor

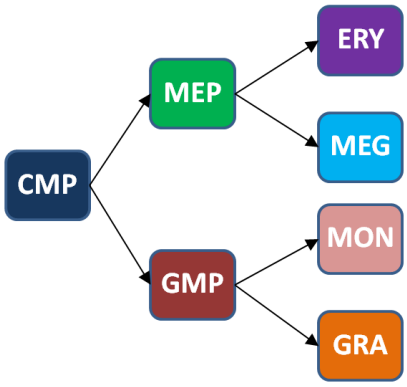


Figure 3.1. A simplified myeloid lineage tree from [3] where the terminal nodes are mature terminally differentiated erythrocytes (ERY), megakaryocytes (MEG), monocytes (MON), and granulocytes (GRA). Multipotent cells are the common myeloid progenitor (CMP), megakaryocyte-erythrocyte progenitor (MEP), and granulocyte-monocyte progenitor (GMP). The color assigned to each cell type in this figure is also used in the differentiation network shown in Figure 3.4.

stem cells produce two pluripotent cell types, a megakaryocyte-erythrocyte progenitor and a granulocyte-monocyte progenitor, that in turn produce terminally differentiated erythrocytes, megakaryocytes, monocytes and granulocytes.

Intracellular genetic regulatory networks (GRNs) control differentiation by responding to external (extracellular) and internal (intracellular) stimuli that reconfigure gene expression profiles and change cell physiology [87]. There is a growing body of evidence that cell types are determined by stable expression patterns of the regulatory networks, referred to as attractors. Switching between cell types amounts to transitioning from one attractor to another [47]. The attractor model explains how cell types can be stable under gene expression noise, and how changes in the expression of a small number of master regulators can shift the expression of hundreds of genes as cell types switch.

Regulatory network dynamics are driven by molecular events within the cell that are subject to noise [88]. Understanding the role of noise in gene expression and its effect on differentiation is essential to gaining insight into cellular specialization and its errors. If cell types are attractors of the GRN, these attractors must be robust to noise in order to maintain particular cell types and to stay on the correct branches of the lineage tree during differ-

entiation. Failure to do either can have dire consequences. For instance, cancer has been proposed to involve destabilization of attractor states due to changes in genetic regulatory network dynamics [89]. In this view, the attractors that correspond to normal cells switch to new, abnormal attractors characteristic of cancer cells. In addition to pathological states, transitions between attractor states of differentiated cells may lead to de-differentiation, in which a cell reverts to an earlier multipotent state, or trans-differentiation, in which a differentiated cell switches to another adult differentiated cell type [90]. Abnormal type switching may also result in off-differentiation in which a multipotent cell from one branch of a lineage tree is converted to a differentiated cell on another branch of the tree. Finally, to maintain a population of multipotent cells, at least some of these cells must resist differentiation to later stages within the lineage tree [91].

An early and influential way of viewing differentiation is Conrad Waddington's [4] epigenetic landscape. Waddington envisioned differentiation occurring on a rugged landscape of sloping ridges and valleys (see Figure 3.2). Waddington represented an undifferentiated cell as a ball at the uppermost point of the highest valley. Differentiation occurred as this ball rolled downhill, encountering the ends of ridges that define branch points between valleys. At each of these branch points the ball moved left or right to follow the new sloping valley to another ridge terminus that separates yet another pair of valleys. Each ridge terminus represents a progenitor cell in a conventional lineage tree and the movement right or left into a new valley from this branch point represents a commitment of the progenitor to one or another lineage. The ridges represent barriers that maintain a cell state once it is chosen.

In the decades since Waddington proposed his model, many investigators have used the concept of an epigenetic landscape and tailored it to explain a variety of developmental processes. Waddington himself cautioned that the epigenetic landscape is an abstraction that could not be rigorously interpreted [4]. Some recent work has tried to enhance Waddington's epigenetic landscape to move it from metaphor to rigorous model [5, 92–95]. However, even with these extensions, the ridge-and-valley topography of the epigenetic landscape places a fundamental limit on the number and kinds of cell type transitions that can be shown. For

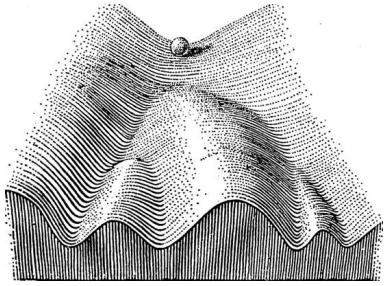


Figure 3.2. Waddington's classic model of an epigenetic landscape [4]. A developmentally immature cell, represented as a ball at the top rolls downhill and is deflected right or left at each branch point until it reaches a catch basin (not shown in this diagram) that corresponds to a terminally differentiated cell.

example, representing trans-differentiation between non-adjacent lineages in Waddington's model requires jumping over two or more ridges, and showing dedifferentation requires uphill movement. Conventional two-dimensional lineage trees suffer similar problems. Even more significant than difficulties in visually representing non-standard, yet documented transitions between cell types is that Waddington's epigenetic landscape and conventional lineage trees both fail to provide quantification of the probability of any transition. Finally, epigenetic landscapes and conventional lineage trees show only a small fraction of the possible transitions between cell types. Many of these transitions were previously considered hypothetical, but with ability to induce pluripotent stem cells from adult differentiated cells and to induce trans-differentiation between lineages, these changes in cell type are well known. To illustrate the limitations of standard representations of cell lineages, a generalized epigenetic landscape like that shown in Figure 3.2 that considers m cell type attractors can only represent a maximum of $2m - \log_2(m+1) - 1$ cell type transitions. This formulation considers the expected differentiation transitions within the lineage tree $(m-1)$ and trans-differentiation events between adjacently arranged cell types on the tree $(m - log_2(m+1))$. As the number of cell types in a system increases, the limitations of the epigenetic landscape become more acute: the number of representable transitions grows with $O(m)$, while the number of possible transitions grows with $O(m^2)$. Given that non-standard attractor type transitions play key roles in cancer and disease development, coupled with the ability to experimentally

induce de-differentiation and trans-differentiation, and the possibility of off-differentiation events, improvements are needed in the visualization of cellular differentiation.

In this work we present a new method that generates a three dimensional graph of attractors and all possible transitions between them to overcome the limitations of a conventional representation of cellular differentiation. Our technique, implemented by a tool called *CellDiff3D*, analyzes the network of attractors generated by a random Boolean GRN. In this work, the GRN that simulates myeloid cell differentiation is used as a demonstration. A noise analysis of the network dynamics is performed to identify m attractors and the likelihood of all the possible $m(m-1)$ transitions between them. This information determines the layout of the graph. The graph is easy to interpret and qualitatively represents the likelihood of transitions between cell types, their overall direction and rate under the influence of noise. Visualization of the results of *CellDiff3D* is achieved by Virtual Reality Modeling Language (VRML), that allows the user to zoom and rotate the three dimensional lineage network.

3.3 *CellDiff3D* Design and Visualization

3.3.1 Separation and Flux between Attractors

We use the mean first passage time (MFPT) [96] between the attractors of any given GRN, represented qualitatively as a Boolean network [37]. MFPT determines the probability and directionality of each theoretically possible transition between all pairs of network states. Introduced by Shmulevich et al. [96], MFPT(a_i, a_j) between a pair of attractors, a_i and a_j, is an estimate of the average number of state update steps of a Boolean network that are required to transition from an attractor state a_i to an attractor state a_j when the network operates under uniform random noise. Noise is modeled by having each bit (gene expression value) have a probability of changing states (a bit flip, from expressed to non-expressed or vice versa) at each state update step. Low MFPTs indicate a high likelihood of a transition between cell states and high MFPTs indicate low likelihood for this transition. Once MFPT between two attractors of a network is estimated, then two useful derived measures of the

epigenetic barrier between attractors can be determined: the separation between attractors and the flux of transitions between them. Let the separation between two attractors i, j be:

$$separation(i, j) = \min(MFPT(i, j), MFPT(j, i)) \tag{3.1}$$

Higher separation implies a lower likelihood of transition between attractors. Note that separation is symmetric. Flux, establishes the directionality of the transition by quantifying the difference between the rates (MFPTs) of forward and reverse transitions between a pair of attractors. The flux between attractors i, j is defined as:

$$flux(i, j) = MFPT(i, j) - MFPT(j, i) \tag{3.2}$$

Note that flux establishes overall direction of the transition between cell states and is asymmetric.

3.3.2 Network Dynamics Visualization

An important element of GRNs is their behavior under gene expression noise. By definition, attractors are stable expression states of a genetic regulatory network, but this stability is relative and expected to vary depending on the network structure and dynamics. For example, terminally differentiated cell states are expected to be more stable than progenitor cells that may be more sensitive to noise-driven changes in states. High levels of gene expression noise may cause unexpected or pathological cell state transitions, with these transitions categorized based on the relative positions of the source and sink cell types in the normal lineage tree. Table 3.1 summarizes five kinds of transitions between cell types and provides an example of each case with respect to the cell types in the simplified myeloid lineage tree shown in Figure 3.1.

Two of these five transition types are represented easily in Waddington's epigenetic landscape: differentiation (moving "downhill" in the landscape toward more specialized cell types) and de-differentiation (loss of specialization shown by upward movement). Two

Table 3.1. Summary of different kinds of cell type transitions with possible examples from myeloid differentiation tree shown in Figure 3.1.

Transition	**Example**	**Definition**
Spontaneous-differentiation	CMP to MEP	Cell switches to a more specialized state
Spontaneous-*de*differentiation	MON to GMP	Cell reverts to an earlier multipotent state
Off-differentiation	GMP to ERY	Cell switches to a more specialized state but on a wrong branch of the lineage tree
Off-dedifferentiation	MEG to GMP	Differentiated cell reverts to an earlier multipotent state but on a wrong branch of the lineage tree
Trans-differentiation	GRA to ERY	Differentiated cell switches to another differentiated state

other transition types cannot be shown in the classic epigenetic landscape representation: off-differentiation (differentiation to a cell type not on the normal lineage path); and off-*de*differentiation (loss of specialization to a cell type off the normal lineage path). Additionally, the epigenetic landscape limits visualization of trans-differentiation events (a switch from one adult differentiated cell type to another) to only those events that occur between adjacently arranged cell types. As discussed earlier, it is important to have a way of representing all possible transition types because off-differentiation and de-differentiation are likely to play central roles in cancer [86, 89], and because recent evidence suggests that trans-differentiation may occur during normal development [97] as well as being induced in cultured cells [98].

Our method visualizes the different attractor transition kinds by constructing a 3-dimensional graph in which the distances between pairs of cell types are their separation (the minimum MFPTs between each pair) and the favored direction of the transition is shown by an arrow with a thickness proportional to the flux. In this way, the graph provides a quantitative view of these important parameters. To reach this result, the following steps are taken. First, the attractors of a given network are determined. Next, noise analysis

(described later) is performed for each attractor pair and the separation and flux values are calculated. This is followed by mapping separation and flux values to a weighted directional graph in which attractors are shown as nodes. Mapping is done using *Graphviz*, an open source graphing application [99]. All these procedures are described in detail in the Methods below. Plotting separation and flux values using *Graphviz* produces 3-dimensional layouts of the graph which can be rotated freely in any web browser and that are easy to understand and analyze.

The graphical layout problem for showing cell type switching is defined in the following way: Let $i_{x,y,z}$ be the $\langle x, y, z\rangle$ coordinate of attractor i in the graph visualization, and $dist(i, j)$ be the Euclidean distance between points $i_{x,y,z}$ and $j_{x,y,z}$. Then given a graph of m attractors defined as a set of $separation(i, j)|1 \leq i, j \leq m$, the layout is defined by determining the set of coordinates for each attractor such that the following summation is minimized:

$$\sum_{1\leq i,j\leq m} (dist(i_{x,y,z}, j_{x,y,z}) - separation(i, j))^2$$

After determining the location of attractors (nodes) in 3D space, flux between pairs of attractors is represented by arrows (directed edges) of variable width between them with arrow width proportional to flux. The edge direction is given by the relationship between $MFPT(i, j)$ and $MFPT(j, i)$: if $MFPT(i, j) < MFPT(j, i)$ then the edge is from i to j. The 3D graph is viewable in any web browser using the VRML viewer plugin (such as *Cartona3D*) and allows the user to rotate and zoom the graph to aid viewing, analyzing, and understanding the relationships between attractors within complex networks.

3.3.3 Visualizing the Myeloid Differentiation Network

We modeled the simplified myeloid lineage network that is shown in Figure 3.1 to demonstrate the utility of the visualization technique. The modeling was based on the work of Krumsiek et al. [3] who considered a network of eleven transcription factors known to be important in myeloid cell differentiation. We extended this work by applying a novel search technique (manuscript in preparation) to discover a new Boolean regulatory network

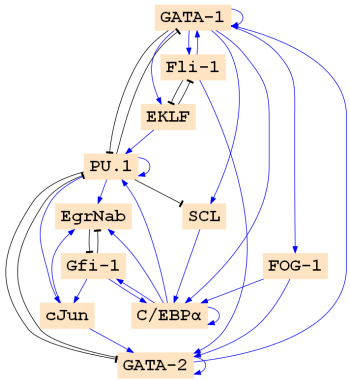


Figure 3.3. The genetic regulatory network used in this work for modeling myeloid differentiation. Nodes are eleven transcription factors that control cell lineage and edges are regulatory interactions between the transcription factors. An arrow signifies activation and a closed line signifies inhibition. The Boolean regulatory control functions are not shown. This network was discovered using a new search algorithm (manuscript in preparation) that uncovers networks that can produce a particular set of cell types, but it does not necessarily find the actual biological network.

that is both supported by the literature and whose dynamics produce all the attractors in the lineage tree: three attractors representing pluripotent cells, along with an additional 4 attractors representing the terminally differentiated cell types. The transcription factor expression pattern of each of these attractors corresponds to a myeloid cell type shown in Figure 3.1. Our GRN discovery method searches the space of Boolean GRNs converging to a specific GRN that minimizes the difference between the attractor's Boolean expression values and the experimental expression values of the corresponding cell types. The new inferred Boolean GRN is illustrated in Figure 3.3. The essential point for demonstrating the value of the *CellDiff3D* approach is that this network produces transitions between cell types that cannot be visualized using Waddington's epigenetic landscape or conventional lineage trees but can easily be seen and analyzed using *CellDiff3D*.

Figure 3.4 shows some outputs of the visualization method applied to simulated myeloid differentiation GRN. Running the myeloid GRN resulted in four attractors with gene expression levels that closely match the four terminally differentiated cell types (erythrocytes (ERY), megakaryocytes (MEG), monocytes (MON), and granulocytes(GRA)). In addition,

there are three attractors that correspond to the MEP and GMP progenitors and the CMP stem cell (expression data is given in [3]).

Each row of Figures 3.4 and 3.5 show three different orthographic projections of the 3D graph of the attractor network. The inferred Boolean network generated the seven stable attractors produced during normal myeloid differentiation (labeled wild type in Figures 3.4 and 3.5). Rows below the wild type network show how network modifications (equivalent to mutations) alter the attractor landscape and how the technique described here can readily visualize these changes. These mutated GRNs were created by knocking out the forward interaction link between a transcription factor and one of its targets by always assigning this link a value of false then running the network to compute the MFPT. For example, in the second row of Figure 3.4, we fix the value of the link from transcription factor EgrNab to transcription factor Gfi-1 in the network shown in Figure 3.3.

A key point in interpreting the visualized lineage networks is understanding flux and separation. For example, in the wild type network of Figure 3.4, note the wide spacing between the granulocyte (GRA; orange) and megakaryocyte-erythrocyte precursor (MEP; green) cells and the narrowness of the arrow that connects these cells. The large distance indicates that there is a low probability for this cell type transition, the direction of the arrow shows the overall direction of this infrequent transition, and the narrow width of the arrow indicates that there is relatively little difference between the forward and reverse rates of the transitions between these cells. Therefore, this is an infrequent and low flux transition. Similarly, the wide separation and lack of an arrow (signalling a very low flux) indicates that granulocyte (GRA; orange) and monocyte (MON; pink) terminal differentiation is stable and trans-differentiation is rare.

Contrast this with the arrow connecting the monocytes(MON; pink) and common myeloid precursor (CMP; dark blue) cells shown in the same row of the figure. The separation between these cell types is small, indicating a low MFPT and a high probability of this transition and the thick arrow connecting the CMP to the MON cells indicates both the overall direction of the cell state transition (CMP to MON) and that the rate of the CMP

to MON forward transition far exceeds the rate of the reverse transition. Therefore, this is a frequent and high flux transition. The ability to rotate this graph freely using the VRML viewer tool adds to the utility of the visualization as the viewer can explore the relationships between all pairs of cell types within this, or any other, lineage network.

Comparisons of the wild type network with mutated networks in which one of the interactions between transcription factors is blocked reveals strong differences in lineage network organization. For instance, in the bottom panel of Figure 3.4, our visualization method immediately demonstrates major alterations in the lineage tree due to blocking Fli-1's regulation of EKLF. In this case, two cell types, megakaryocyte-erythrocyte progenitor (MEP) and erythrocytes (ERY), are no longer present.

Finally, the technique developed here is able to reveal many different kinds of transitions between cell states (Table 3.2). Although a GRN that produces attractors that correspond to myeloid cell types was used in this initial study, any GRN and its resulting attractors/cell types can be explored using this approach. Significantly, non-standard transitions such as de-differentiation, off-differentiation, and trans-differentiation, are increasingly recognized in normal and disease states, many of which cannot be shown using conventional lineage trees. Our method allows their representation in 3-dimensional space and provides important information on their likelihood under either gene expression noise as shown here, or other driving forces in GRN dynamics.

3.4 Methods

3.4.1 Cell Differentiation and Attractor Dynamics

First proposed by Kauffman [37], Boolean networks are one of the main contributors to our current knowledge of gene regulatory networks. They have proved effective in representing many biological systems including *Drosophila* development [41, 42], angiogenesis [43], eukaryotic cell dynamics [44], and yeast transcription networks [40]. Boolean networks consist of nodes and directed edges. In GRN modeling, nodes represent the genes and edges represent the regulatory influences between the genes. These regulatory influences are fully

Table 3.2. Cell type transitions discovered and visualized in the myeloid differentiation network shown Figure 3.3 and in mutationally altered forms of this network.

Figure	**Network**	**Cell type switch**	**Kind**
3.4(a)	Wild Type	CMP ⇒ MON	Spontaneous-differentiation
		MEP ⇒ GMP	Off-differentiation
		MEG ⇔ CMP	High separation
		MEP ⇒ GMP	Off-differentiation
		MEP ⇔ MEG	Low separation
3.4(b)	EgrNab/Gfi-1	CMP ⇒ MEG	Spontaneous-differentiation
		ERY ⇒ GRA	Trans-differentiation
		GRA ⇔ CMP	Low separation
3.4(c)	Fli-1/EKLF	MEG ⇒ CMP	Spontaneous-dedifferentiation
		MEG ⇒ MON	Trans-differentiation
		MEG ⇒ GMP	Off-differentiation
		MEG ⇔ MON	Low separation
		GMP ⇔ MON	High separation
3.5(b)	GATA-2/PU.1	GRA ⇒ CMP	Spontaneous-dedifferentiation
		MEP ⇒ MON	Off-differentiation
		GMP ⇒ ERY	Off-differentiation
		GMP ⇒ MEG	Off-differentiation
3.5(c)	GATA-1/PU.1	MON ⇒ GMP	Trans-differentiation

defined by the updating rules for each gene as a logic function of the inputs. A gene can be either expressed (the output is true) or not expressed (the output is false).

A Boolean network with n genes has 2^n possible states, denoted as $\hat{S}$. Each network state $\hat{s}_t$ is the collection of all gene values at time t, $\hat{s}_t = \{g_1, g_2, ..., g_n\}$. Given the current state $\hat{s}_t$, the next network state $\hat{s}_{t+1}$ is obtained by applying each gene's function to the the current gene values. The gene's logic functions are deterministic. Thus, the the mapping function $D(\hat{s}_t)$ that finds the next network state is also deterministic: $\hat{s}_{t+1} \leftarrow D(\hat{s}_t)$. By repeatedly applying deterministic updating, the network dynamics will eventually reach a previously visited state. This cycle is called an attractor ($\hat{a}$). Attractors can be single states, called point attractors or cyclic attractors in which the cycle consists of more than one state. Note that to find all attractors of a given network, all possible starting states need to be considered (the code can be obtained from `http://code.google.com/p/pbn-matlab-toolbox`).

In this work, cell types are considered attractors in the state space of possible gene expression profiles [46] and cell differentiation is modeled as the process of transitioning from one attractor to another [47].

3.4.2 Simulating and Measuring Noise Dynamics

Noise at the molecular level plays a key role in many biological processes including protein folding, transcription factor binding to DNA, and the rate of initiating transcription and translation [100, 101]. At the systems level, noise influences the likelihood of cell type transitions [46]. Noise can be modeled in Boolean regulatory networks by random bit flips during network operation, with these bit flips representing noise-driven changes in gene expression. Let $\hat{s}_j \leftarrow \eta(\hat{s}_i, r)$ be the spontaneous noise function that maps a state of the network $\hat{s}_i$ to a new state $\hat{s}_j$ with the addition of noise, implemented as r bit flips, each single bit flip occurring with probability p. Noise modifies the probability of state transitions as the states are updated and the switching among network attractors. Since attractors represent cell types, measures of noise tolerance can estimate the magnitude of the barrier between attractors; the so-called epigenetic barrier. In the following section three measures of the epigenetic barrier are introduced and compared.

Hamming Distance

Hamming distance is the direct measure of the difference between corresponding elements of two bit vectors. In GRNs, Hamming distance measures the differences in expression levels between two network states. Differences between gene expression profiles are used to identify cell type or cell physiology [102]. However, as a measure of the epigenetic barrier between states, Hamming distance does not utilize $\eta(\hat{s}, r)$ and also ignores the constraints that regulatory network dynamics impose upon state transitions $D(\hat{s})$. For these reasons, Hamming distance is a poor measure of the epigenetic barrier.

Transitory Perturbation (Single-bit-flip)

An alternative measure of the likelihood of attractor transition under expression noise

was introduced by Villani et al. [103]. Once the set of attractors is identified, this measure inserts noise as a single bit flip one-off event followed by deterministic updating. So given $\hat{a}_i$ as an attractor state, $\hat{s}_i \leftarrow \eta(\hat{a}_i, 1)$ is applied to a single bit, then the network-defined updating rules are applied determinatively until an attractor state $\hat{a}_j \leftarrow D^*(\hat{s}_i)$ is reached. For each attractor and each bit, the process is repeated. Let $c_{i,j}, 1 \leq i, j \leq m$ (where m is the number of attractors) be the count of when $\hat{a}_j \leftarrow D^*(\eta(\hat{a}_i, 1))$. Then, $P(\hat{a}_i, \hat{a}_j) = \frac{c_{i,j}}{m}$. For each pair of attractors $\{\hat{a}_i, \hat{a}_j\}$, $P(\hat{a}_i, \hat{a}_j)$ is the portion of single one-step bit flips (transitory perturbations) in the nodes of all states of attractor $\hat{a}_i$ which will result in a transition from $\hat{a}_i$ to $\hat{a}_j$ under noise-free dynamics.

This single-bit-flip measure of likelihood of network transition under noise efficiently estimates the epigenetic barrier (since it is $O(nm)$), but it assumes that expression noise is an infrequent event during network dynamics.

Mean First Passage Time

Introduced by Shmulevich et al. [96], mean first passage time ($MFPT$) is the the average time it takes to reach state y from state x in the presence of noise. Mathematically, first passage time (FPT) is defined as $F_k(\hat{s}_x, \hat{s}_y)$: the probability that starting in state $\hat{s}_x$, the first time the system visits a state $\hat{s}_y$ will be at time k; in Boolean networks, time is measured as the number of state updates. MFPT is then defined as:

$$MFPT(\hat{s}_x, \hat{s}_y) = \sum_k k F_k(\hat{s}_x, \hat{s}_y) \tag{3.3}$$

Where the F_k itself is formulated as:

$$F_k(\hat{s}_x, \hat{s}_y) = \sum_{\hat{s}_z \in \{0,1\}^n, z \neq y} p_{xz} F_{k-1}(\hat{s}_z, \hat{s}_y) \tag{3.4}$$

In this recursive formula $F_1(\hat{s}_x, \hat{s}_y)$ is the probability of direct transition from state $\hat{s}_x$ to $\hat{s}_y$. p_{xz} is the probability of transition from state $\hat{s}_x$ to state $\hat{s}_z$. Probabilistically, there are two ways to reach state $\hat{s}_z$ from $\hat{s}_x$; either $\hat{s}_z$ is a deterministic target for $\hat{s}_s$ and no bit flips occur due to the noise, or an aggregate of bit flips drive the transition from $\hat{s}_x$ to $\hat{s}_z$.

When the *MFPT* between two states is low, it implies that starting from the first state, the second state is easily reached by molecular noise. Figure 3.6 shows F_k and kF_k for the transition between two arbitrary attractors. As this figure shows, the b to a transition has a lower MFPT compared to a to b. Note that when an attractor has more than one state i.e it is a cyclic attractor, the MFPT is calculated for each state separately and then is averaged over all states of that attractor.

At each network state update $D(\hat{s})$ there is a probability that the state will change as a function of the Hamming distance (h) between the current state and the subsequent state $\hat{s}_{t+1} \leftarrow D(\eta(\hat{s_t}, r))$. *MFPT* models uniform expression noise by considering probabilistic bit flips at every possible state of the network and deriving the distribution of passage times from analysis of the corresponding Markov process. Statistically, the probability distribution of bit flips can be seen as a binomial distribution, thus the probability of r bit flips, $\eta(\hat{s}_a, r)$ is $\binom{n}{r}p^r(1-p)^{n-r}$, where p is the probability of a single bit flip and n is the total number of bits.

Mean first passage time quantifies the epigenetic barriers between all attractor states during network execution. Therefore, this work only considers *MFPT* because of its realism in modeling expression noise. However, the time required for MFPT computation is an exponential function of the number of genes, so if the number of genes in the network is large, calculating MFPT may become intractable. In this case, transitory perturbation can be used as a possible alternative.

3.5 Summary

In this work, we developed a technique and a supporting method for visualization, *CellDiff3D*, that estimates the likelihood and directionality of noise-driven transitions between different cell types and allows the three dimensional visualization of these relationships. A Boolean network model of myeloid cell differentiation [3] was used as a demonstration system for this research.

The metric of mean first passage time (MFPT) assesses the likelihood that noise in the GRN for myeloid differentiation will trigger a transition between cell types. Low MFPT

values indicate a high probability of a cell type transition. The difference in MFPTs for forward (cell type A to cell type B) and reverse (cell type B to cell type A) transitions provides a measure termed flux. Flux is analogous to the difference in forward and reverse rates of a chemical reaction and it gives the anticipated direction and the strength of the directionality in transitions between cell types.

Our technique calculated the MFPT, separation and flux between all pairs of cell types in a simplified myeloid lineage tree that included one multipotent stem cell, two intermediate cells, and four terminal cell types to produce a graph to display all 42 pairwise relationships $m(m-1)$ where $m = 7$ between the myeloid cell types. A VRML-based graphics tool was employed as part of *CellDiff3D* to visualize all attractor type transitions by placing all pairs of different cell types in 3 dimensional space. It shows the likelihood of a transition between cell types as the separation between each pair and the directionality of the transition as arrows with a width proportional to the flux. The VRML output, viewable in any web browser (with the proper plugin), allows the free rotation and zooming of the differentiation network to reveal its features. It can be used for any cell differentiation network, can include many more than the 7 cell types considered here, and is capable of showing all possible transitions (for example, de-differentiation and trans-differentiation) between different types of cells. Our technique readily revealed changes in the dynamics of mutationally altered myeloid differentiation networks, the loss of cell types, and unusual cell type transitions that included dedifferentiation, trans-differentiation, and off-differentiation.

This work has introduced a 3D graph approach to visualize the influence of noise on cell type switching of wild type and mutated regulatory networks. However, the system is not limited to noise analysis and can incorporate other influences that drive cell type switching.

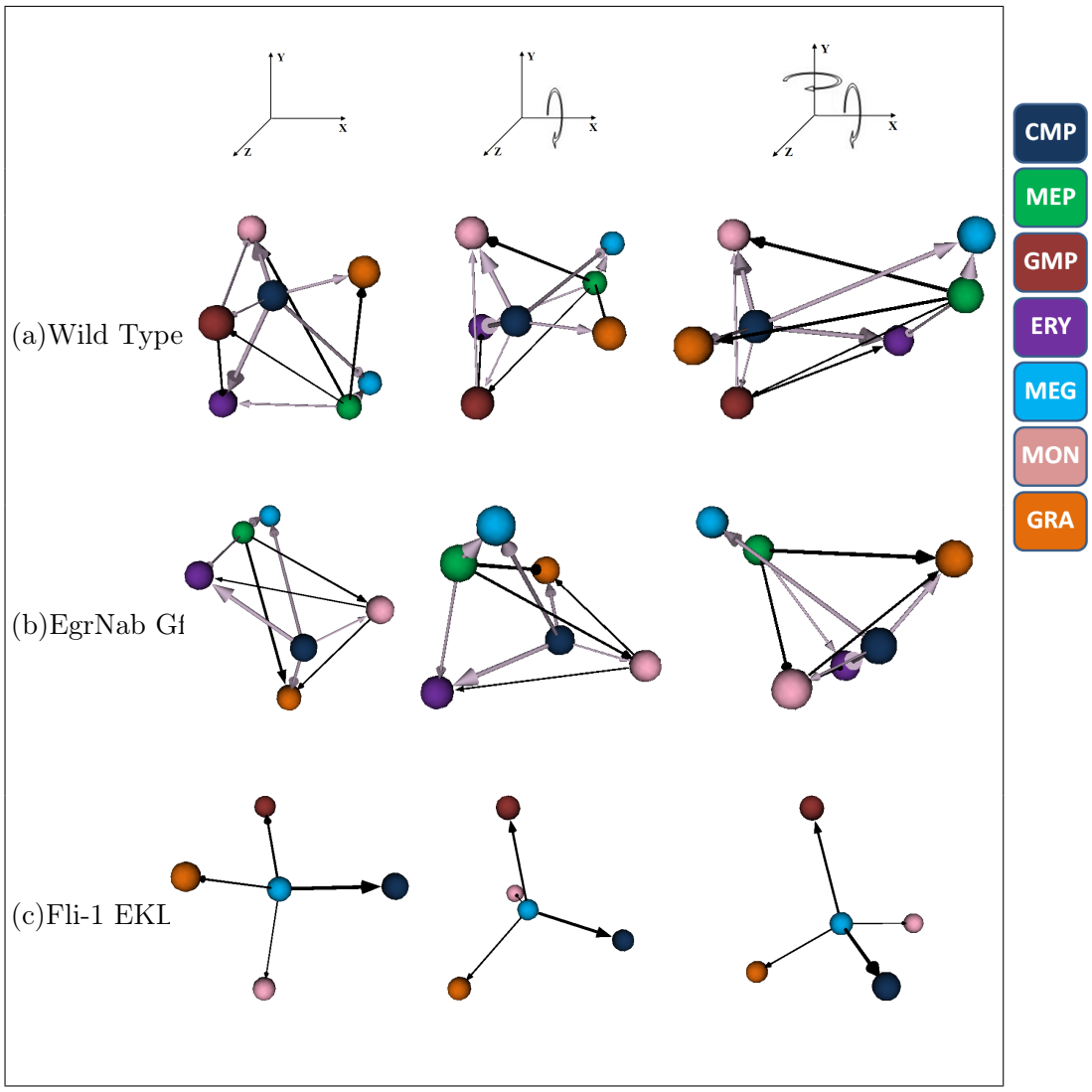


Figure 3.4. *CellDiff3D* visualization of the simulated myeloid differentiation network. Each image is a still taken from renderings of VRML code produced by the modeling method. The transcription factors and their regulatory interactions that comprise the GRN are shown in Figure 3.3. Each sphere is one of the myeloid cell types shown in Figure 3.1. Each row shows three orthographic views of cell type transitions derived from runs using the wild type transcription factor network (top row of panel) or with transcription factor mutations in which the first transcription factor listed does not interact with the second transcription factor (lower rows of panel). The distance between each pair of cell types is the separation and the arrow direction and thickness is flux. For clarity low flux edges are not shown. Lavender arrows show normal differentiation or de-differentiation along the standard lineage tree from a specialized cell to its immediate progenitor; black arrows show trans-differentiation, off-differentiation, or off-dedifferentiation.

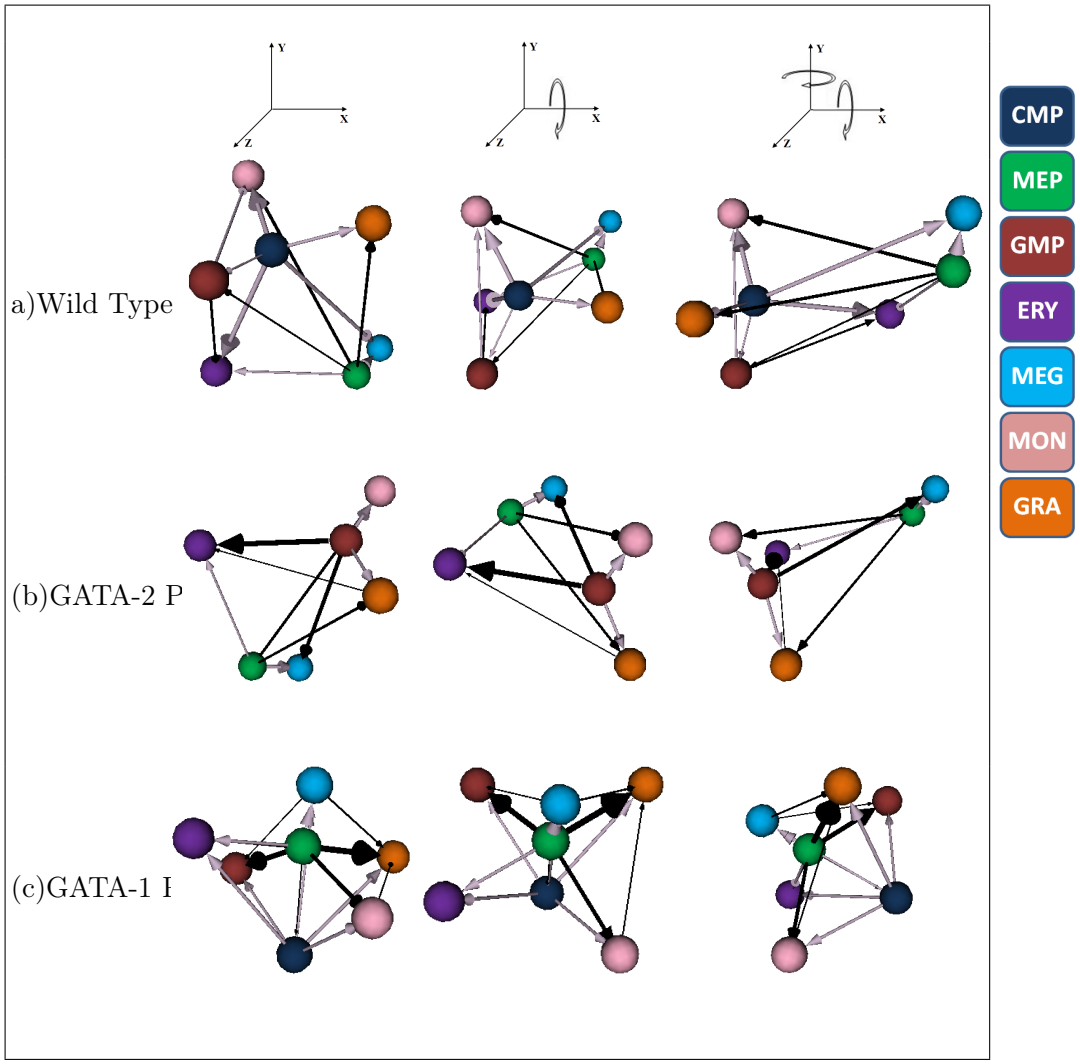
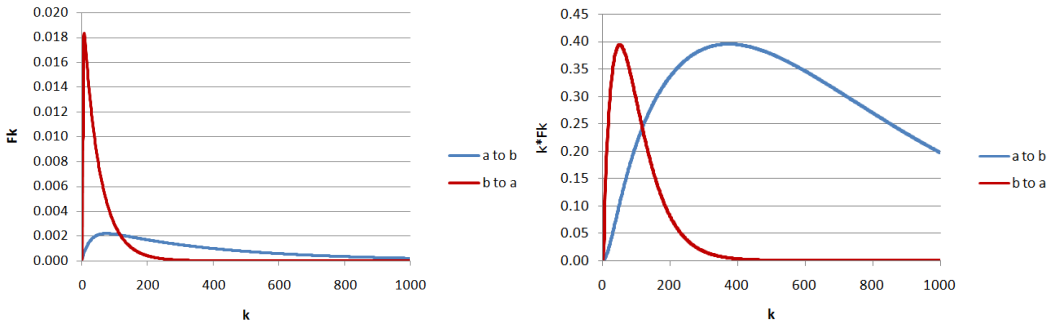


Figure 3.5. *CellDiff3D* illustration of the effects of two additional mutations that disrupt the myeloid differentiation network. There are interactions between GATA-2 to PU.1 (middle row) and GATA-1 to PU.1 (bottom row). See Figure 3.4 for extended caption.



(a) (b)

Figure 3.6. (a) F_k (probability of first visit at time step k) plotted for two arbitrary attractors, called a and b in a random Boolean network for 1000 steps (k). The red curve is for the transition from b to a that has a low MFPT compared to the reverse transition, a to b is shown with the blue curve; (b) kF_k plotted for the F_k curves in (a). Note that MFPT is the centroid of the area under the kF_k curve.

CHAPTER 4

MULTISTABLE SWITCHES AND THEIR ROLE IN CELLULAR DIFFERENTIATION NETWORKS [1]

4.1 Abstract

Background: Cellular differentiation during development is controlled by gene regulatory networks (GRNs). This complex process is always subject to gene expression noise. There is evidence suggesting that commonly seen patterns in GRNs, referred to as biological multistable switches, play an important role in creating the structure of lineage trees by providing stability to cell types.

Results: To explore this question a new methodology is developed and applied to study (a) the multistable switch-containing GRN for hematopoiesis and (b) a large set of random boolean networks (RBNs) in which multistable switches were embedded systematically. In this work, each network attractor is taken to represent a distinct cell type. The GRNs were seeded with one or two identical copies of each multistable switch and the effect of these additions on two key aspects of network dynamics was assessed. These properties are the barrier to movement between pairs of attractors (separation) and the degree to which one direction of movement between attractor pairs is favored over another (directionality). Both of these properties are instrumental in shaping the structure of lineage trees. We found that adding one multistable switch of any type had a modest effect on increasing the proportion of well-separated attractor pairs. Adding two identical switches of any type had a much stronger effect in increasing the proportion of well-separated attractors. Similarly, there was an increase in the frequency of directional transitions between attractor pairs when two

[1] A. Ghaffarizadeh, N.S. Flann, and G. J. Podgorski, "Multistable Switches and their Role in Cellular Differentiation Networks," *BMC Bioinformatics* 2014, 15 (Suppl 7), S7.

identical multistable switches were added to GRNs. This effect on directionality was not observed when only one multistable switch was added.

Conclusions: This work provides evidence that the occurrence of multistable switches in networks that control cellular differentiation contributes to the structure of lineage trees and to the stabilization of cell types.

4.2 Introduction

Understanding differentiation is critical to knowing how normal development unfolds and for taming diseases, such as cancer, that are associated with defects or reversals in differentiation. In animals, the process of differentiation typically results in cells reaching a terminally differentiated state. However, recent discoveries have shown that "terminal differentiation" may be a misnomer as fully differentiated cells can be reprogrammed to revert back to a pluripotent state, with these pluripotent cells having the potential to differentiate into other cell types.

Transitions between cell types can be mapped as a directed tree of cell types, known as a lineage tree, with embryonic stem cells at the root, various classes of precursor cells as internal nodes, and terminally differentiated cells as branch tips. Gene regulatory networks (GRNs) that respond to both external stimuli and to gene expression noise control transitions between cell types and determine the structure of lineage trees [5]. Given that differentiation is driven by the output of dynamic gene regulatory networks, a useful, network-based perspective for envisioning different stable cell types is as basins in an attractor landscape [39, 87]. In this dynamical systems view, differentiation is the process of moving between the different attractor basins that are generated by the dynamics of the gene regulatory network.

The GRNs that control differentiation are complex, but these larger networks can be decomposed into smaller modules of simpler, frequently appearing regulatory motifs that consist of only a few genes that interact in characteristic patterns [5]. For example, a common feature of many regulatory motifs is a pair of genes coupled by either positive or negative feedback loops [104]. These couplings result in different network outputs, with positive

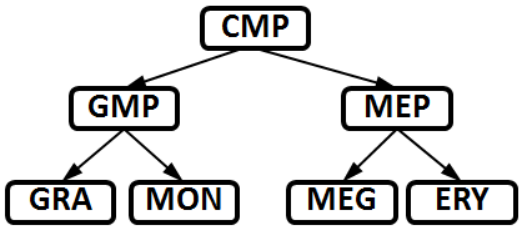


Figure 4.1. A simplified myeloid lineage tree (from [3]) where the terminal nodes are the mature cell types of erythrocytes (ERY), megakaryocytes (MEG), monocytes (MON), and granulocytes (GRA). Multipotent cells are the common myeloid progenitor (CMP), megakaryocyte-erythrocyte progenitor (MEP), and granulocyte-monocyte progenitor (GMP).

feedback loops often producing two or more stable attractor states, and negative feedback loops often enhancing attractor stability [104]. The generation of two or more attractors is referred to as multistability, with the special case of generating only two attractors termed bistability.

In this work, we investigated four regulatory motifs, termed multistable switches, that operate in differentiating cells [5,104]. Each of these motifs results in multistability when the motif operates in isolation [5]. These multistable switches were added singly or in identical pairs to larger GRNs to understand how they affect the structure of lineage trees and the stability of different cell types. These studies were done by generating random Boolean GRNs that produce five or more attractors. These networks were then seeded with the multistable switches. We found that the addition of identical pairs multistable switches of any of the four different types increased the stability of attractors produced by the GRNs. Adding a single multistable switch of any type had little effect on attractor stability. The addition of two multistable switches to a randomly generated GRN also increased the proportion of directional transitions between attractors. In terms of differentiation, this contributes to the structure of a lineage tree by favoring particular pathways that lead between different cell types.

4.3 Approach and results

This work studied three key properties of cellular differentiation [103]: (a) differentiation of multipotent cells can be driven by gene expression noise; (b) there is a strong directionality to differentiation, with transitions between cell types occurring from less to more differentiated cells; and (c) terminally differentiated cells are stable.

The simplified myeloid linage tree illustrated in Figure 4.1 provides an example of these key properties. This lineage tree includes only favored transitions between cell types that involve progenitor cells giving rise to two different, more differentiated cell types, and the establishment of barriers between cell types that prevent transdifferentiation and dedifferentiation.

4.3.1 Cellular differentiation and attractor dynamics

In this work, differentiation is viewed as a set of transitions between attractor basins produced by a dynamical genetic regulatory network. This model of differentiation was pioneered by Kaufman and extended by many others [37, 39, 103, 105]. Borrowing from early work by Waddington [4], the landscape created by these attractor basins has been termed an epigenetic landscape [5]. A conceptual model of such an epigenetic landscape is shown in Figure 4.2. In this view, each cell type occupies an attractor basin at a particular level of a potential energy landscape. A cell can be moved out its attractor basin in response to an external signal or to gene expression noise. Once it crosses the barrier that delimits the basin, it moves down to another attractor basin lower in the epigenetic landscape. There are at least two possible paths leaving each attractor basin, with each downhill path leading to a different basin that represents a distinct, more specialized cell type. Once a cell descends into a new basin, the large potential energy barrier between the new lower basin and upper starting basin makes it unlikely for a more specialized cell to make the transition back to a progenitor cell. This process of cells moving out of an attractor basin in response to external signals or to gene expression noise and descending into attractor basins of lower potential energy that correspond to more differentiated cells is repeated at each level of the lineage tree.

This potential energy barrier that must be crossed to move between attractor basins is called the epigenetic barrier. Shmulevich et al. [96] proposed a method of quantifying this barrier termed the mean first passage time (MFPT), defined as the average number of state transitions needed to move from one attractor basin to another during the noisy operation of a Boolean regulatory network. The MFPT provides a measure of the probability of a particular transition between two attractor basins, with low MFPTs indicating a high likelihood of the transition, and high MFPTs indicating a low likelihood for this transition. Details on the calculation of MFPT values and all other aspects of the procedures are given in Methods; this section will only provide an overview.

The forward and reverse MFPT values between two attractor basins (simply called attractors from this point forward), att_1 and att_2, provide information on the directionality of the transition. Directionality is a key element of differentiation, as under normal circumstances, cells transition from less to more mature states, but not in the reserve direction. For the pair of attractors att_1 and att_2, we define a **directional transition** to occur if $att_1 \rightarrow att_2$ (reaching att_2 from att_1) has a significantly larger MFPT than the MFPT of $att_2 \rightarrow att_1$.

Another important aspect of cellular differentiation captured by MFPT is the probability of making a transition between any pair of different cell types. This is important in shaping the structure of a lineage tree and in stabilizing cell types. For example, progenitor cell types should not differentiate into cell types off the normal lineage path, and terminally differentiated cells must be prevented from dedifferentiation or transdifferentiating into other cell types. Therefore, the MFPT should be high in both directions for unfavored transitions between attractors. We term this **separation**, with high separation occurring when the MFPTs of $att_1 \rightarrow att_2$ and $att_2 \rightarrow att_1$ are both large.

Given the directionality of differentiation and the large separation of the majority of cell types within a linage tree, a plot of the distribution of MFTPs of the forward (for example, $att_1 -> att2$) and reverse ($att_2 -> att_1$) transitions between all possible pairs of cell types within a lineage tree is expected to show clustering in the regions of directionality and

separation. This is shown in Figure 4.3. In this plot of forward and reverse MFPTs between all possible attractor pairs produced by a set of gene regulatory networks, the quadrant with a low forward MFPT and high reverse MFPT represents attractor pairs (cell types) that are linked with a strong directional transition. In contrast, the quadrant with high MFTPs in both the forward and reverse directions represents well separated attractor pairs. This region of high separation represents low probability transitions between cells types, such as transdifferentiation or differentiation off the normal lineage pathway. Using this reasoning, if adding a small multistable switch to a larger GRN enhances the directionality of transitions between attractors, then in a plot like the one shown in Figure 4.3, there should be an increase in frequency of attractor pairs in the regions labeled *directional*. Similarly, if a multistable switch added to a gene regulatory network increases the separation between pairs of attractors, then there should be an increase in the region of Figure 4.3 labeled *separate*. This is the basis of the approach followed in this work.

An important point to note is that in a MFPT representation of biologically realistic lineage trees, the proportion of attractor pair transitions in the *separate* region will far exceed the proportion in the directional quadrant. This is because the topology of actual linage trees leads to there being significantly fewer directional transitions than well separated transitions. Intuitively, this stems from the ideas that the number of favored transitions between different cell types is much smaller than the number of theoretically possible transitions, and that most of the theoretically possible transitions are unfavored events such as dedifferentiation and transdifferentiation. Mathematically, the possible number of well separated transitions is on the order $O(b^{2h})$ while the number of directional transitions is of the order $O(b^h)$, where b is the branching factor of differentiation tree (number of children for each node) and h is the height of the tree measured as the number of cell type transitions between a stem cell and a terminally differentiated cell. This expected difference in the proportions of *separate* and *directional* attractor pair transitions is important when interpreting the effects of adding multistable switches to random Boolean genetic regulatory networks (see below).

We investigated how the addition of the four multistable switches shown in Figure 4.4

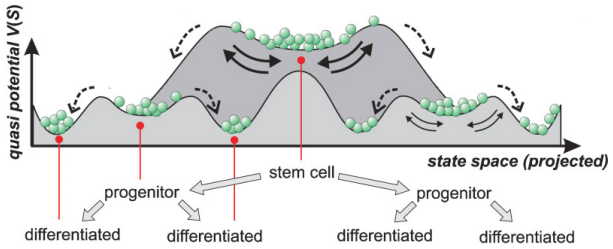


Figure 4.2. A hypothetical two dimensional epigenetic landscape of differentiation (modified from [5]). The horizontal axis shows the state space of different cell types and the vertical axis approximates potential energy differences between cell types. The basins are attractors that represent different cell types and the magnitude of potential energy differences between states provides a measure of the probability of transitions between states under gene expression noise.

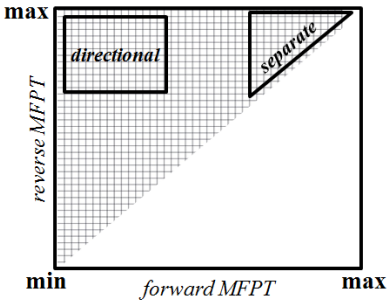


Figure 4.3. Forward and reverse MFPT plot showing *directional* and *separate* regions.

influenced the attractor landscape produced by randomly generated Boolean regulatory networks. A conventional node-and-edge diagram of each multistable switch used in biological literature is depicted in the figure, followed by a more informative logic circuit representation. The first logic circuit (Figure 4.4.a) is usually referred to as a bistable switch (BS) or toggle switch [106]. We call the second logic switch (4.4.b) a mutual inhibition switch (MI00). Note how the less informative node-and-edge diagrams for these two distinct logic circuits are identical. The next two multistable switches extend mutual inhibition with the addition of one (*MI+0*) or two (*MI++*) positive feedback loops. *MI++* is sometimes referred to as tristable switch.

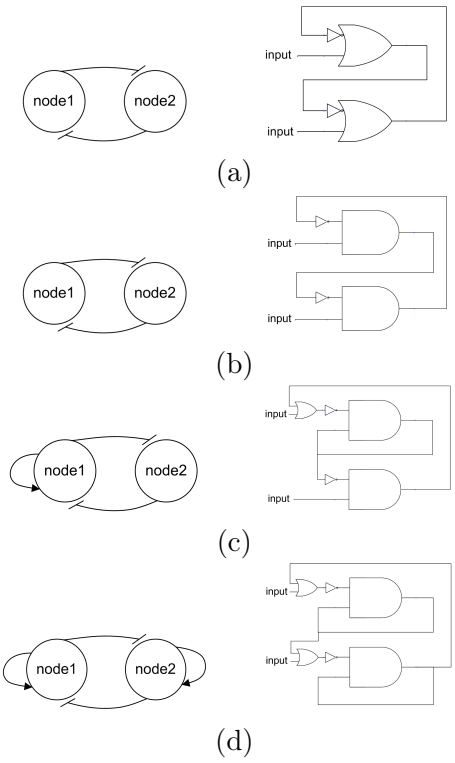


Figure 4.4. Multistable switches used in this work. The diagrams in the left show the node and edge representation and the diagrams at the right show the logic gate representation of each switch. The truth table of the functions are [1,1,0,1] for a and [0,1,0,0] for b, c, and d for binary numbers [00,01,10,11], respectively. In this work, the multistable switches are referred to as: (a) bistable switch (*BS*), (b) mutual inhibition with zero positive feedback loops (*MI00*), (c) mutual inhibition with one positive feedback loop (*MI0+*), and (d) mutual inhibition with two positive feedback loops (*MI++*).

4.3.2 Multistable switches in myeloid differentiation

An important example of cellular diversification is the well studied system of hematopoiesis. During hematopoiesis, multipotent stem cells (hemocytoblasts) differentiate into either myeloid or lymphoid progenitors [3]. A sub-tree of the myeloid lineage tree is illustrated in Figure 4.1. This figure shows that hemocytoblast stem cells produce two pluripotent cell types (megakaryocyte-erythrocyte progenitor (MEP) cells and granulocyte-monocyte progenitor (GMP) cells) that in turn produce terminally differentiated erythrocyte, megakaryocyte, monocyte and granulocyte cells.

To construct a GRN that simulates the dynamics of the myeloid differentiation, we extracted a set of regulatory gene expression levels of all cell types in Figure 4.1 from three datasets of distinct experiments available at ArrayExpress database (`http://www.ebi.ac.uk/microarray-as/ae/`): E-GEOD-5606, E-GEOD-8407, and E-GEOD-18483. Motivated by Krumsiek et al. [3], we picked 11 transcription factors that play important roles in myeloid differentiation: GATA-1, GATA-2, FOG-1, EKLF, Fli-1, SCL, C/EBPα, PU.1, cJun, EgrNab, and Gfi-1; note that the EgrNab, represents an integration of Egr-1, Egr-2 and Nab-2. Using these genes and their expression profiles, we utilized a search tool to infer a GRN for myeloid differentiation as a Boolean network (manuscript in preparation). This network includes 4 well-known gene interactions that represent multistable switches [3, 5]: a) An MI++ switch between GATA-1 and PU.1; b) An MI++ switch between GATA-2 and PU.1; c) A bistable switch between Fli-1 and EKLF; and d) A bistable switch between Gfi-1 and EgrNab. We computed the MFPT between attractors of this network that represent the cell types of the myeloid lineage tree. The pairwise forward and reverse MFPT values between all pairs of attractors of this network are depicted in the Figure 4.5 (red circles); we also included the MFPT values for the attractors of the original network proposed by Krumsiek and colleagues (green diamonds) that contains only four attractors as the terminally differentiated cell types. This figure shows that the majority of transitions in myeloid differentiation fall in either the separation or directionality regions shown in Figure 4.3.

4.3.3 Multistable switches in random networks

We showed that the myeloid differentiation network, with its multistable switches, generates directional transitions and well separated attractors. How general is this result? We extended our study to examine the role of these switches in a large space of cellular differentiation networks.

The outline of this approach was to:

1. Construct a random Boolean network (only networks that are expected to operate in the critical domain were generated (see Methods)).

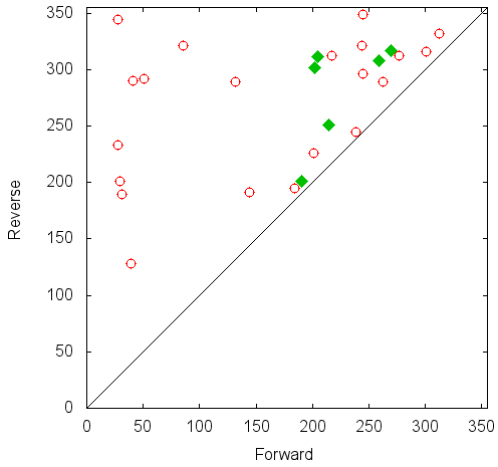


Figure 4.5. Forward and reverse MFPT plot for the myeloid differentiation network. Red circles are the MFPT values of our inferred network. This network has all 7 attractors of the myeloid lineage tree shown in Figure 4.1, including multipotent cells. Green diamonds show the MFPT values for the network proposed by Krumsiek et al. which only has the 4 terminally differentiated cell types [3]. Including multipotent cells illustrates additional attractor relations, including directionality.

2. Embed zero, one or two copies of a given multistable switch within the network.

3. Run the network and identify attractors; if the number of attractors is less than 5, go back to step 1.

4. Compute the forward and reverse MFPT between all pairs of attractors.

5. Map the forward and reverse MFPT of each pair of attractors to a point in a MFPT density plot like the one shown in Figure 4.3.

6. Repeat for 5000 random Boolean networks to create each MFPT density plot.

Density plots were generated for 9 different types of networks: RBN networks without any added multistable switch and RBNs with one or two identical copies of each of the four types of multistable switches. Figure 4.6 shows these density plots. Each plot shows the forward and reverse MFPT between all attractor pairs generated by 5000 networks of a single type.

The MFPT density distribution produced by RBNs without any added multistable switch (Figure 4.6 a) shows no clustering in the *directional* or *separate* regions of the plot. Instead, the forward and reverse MFPTs of most of the transitions are equal and of intermediate values and therefore fall in the mid-range of the diagonal. Adding a single multistable switch of any type to the RBN had a modest effect of increasing the density of attractor pairs in the *separate* region. Adding two multistable switches of the same type to the RBN had a much stronger effect on increasing the frequency of well separated attractor pairs. This is reflected in an increased density in the *separate* region of the MFPT plots. The particular kind of multistable switch had little impact on this effect; instead, the critical element was adding two rather than one multistable switch to the RBN.

There was a modest increase in the density of attractor pairs in the *directional* regions of the MFPT plot when two identical multistable switches were added. However, as discussed above, a major clustering of MFPT values in the *directional* region is not expected in networks that produce lineage trees. The modest increase in directionality gained by adding multistable switches is likely to be significant. In contrast to the effect on separation, there was a difference between the multistable switch types in increasing directionality: The MI++ switch type did not increase directionality, but all three of the other types did. To better illustrate these enrichments in *directional* and *separate* regions, Figure 4.7 shows the difference between the MFPT distribution of networks with two embedded multistable switches and the base-line random network distribution.

4.4 Conclusion

This work examined how the attractor structure generated from random Boolean regulatory network dynamics was influenced by the addition of multistable switches that are commonly found in biological networks that control differentiation. The results show that the addition of multistable switches increases the resilience of genetic regulatory networks to gene expression noise. This is seen by the increase in the proportion of well separated attractors. In a biological context, this separation of attractors has the effect of stabilizing determined cells and of helping to establish well defined pathways between differentiating

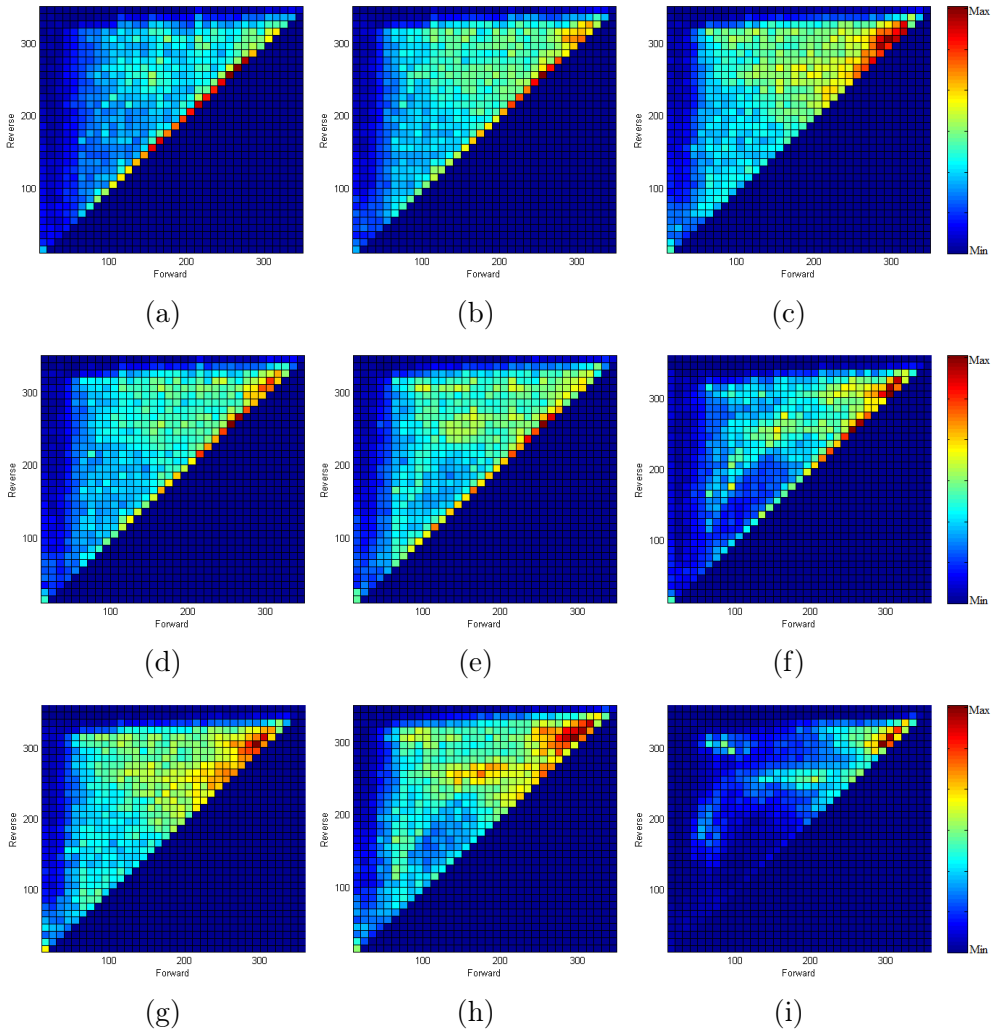


Figure 4.6. Distributions of MFPT values. The plots show the forward and reverse MFPTs for all transitions seen in 5000 critical networks of each type. (a) Networks with no added multistable motifs; (b) Networks with one embedded bistable switch; (c) Networks with two embedded bistable switches; (d) Networks with one embedded MI00 switch; (e) Networks with one embedded MI+0 switch; (f) Networks with one embedded MI++ switch; (g) Networks with two embedded MI00 switches; (h) Networks with two embedded MI+0 switches; (i) Networks with two embedded MI++ switches.

cells. Adding a single multistable switch to a random network had a relatively modest stabilizing effect, but adding two identical switches of any of the four types tested here produced much stronger barriers between different cell types. In parallel, there was also evidence that adding two multistable switches to a genetic regulatory network increased the frequency of

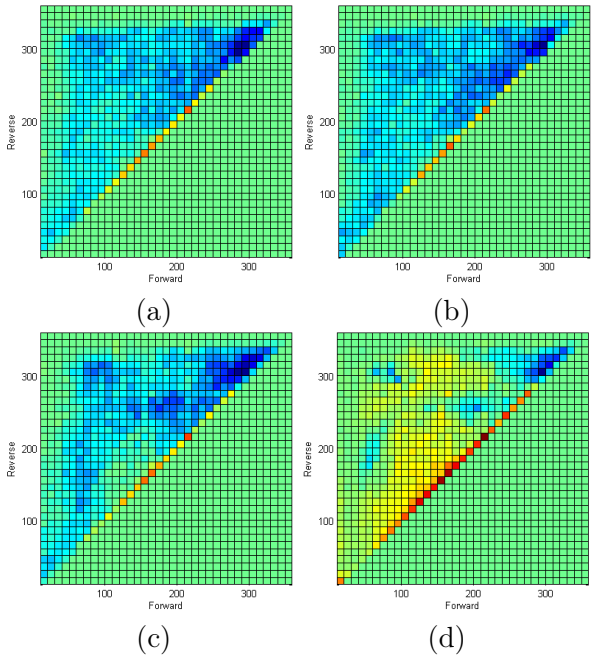


Figure 4.7. Difference of distributions of MFPT values for networks embedded with two identical motifs against the networks with no motifs. (a) Difference of network with no motifs and networks with two embedded bistable switches ; (b) Difference of network with no motifs and networks with two embedded MI00 switches; (c) Difference of network with no motifs and networks with two embedded MI+0 switches; (c) Difference of network with no motifs and networks with two embedded MI++ switches.

directional transitions between attractors. From a biological perspective, this structures a linage tree by favoring one-way transitions between particular cell types. Therefore, the pervasive occurrence of multistable switches in networks that control cellular differentiation is likely to contribute to the structure of lineage trees and to the stabilization of cell types.

4.5 Detailed methods

4.5.1 Cell differentiation and attractor dynamics

Boolean networks [37] have proved effective in representing GRN structure and dynamics in many systems, including *Drosophila* development [41,42], angiogenesis [43], eukaryotic cell dynamics [44], and yeast transcription networks [40]. Each gene in a network is repre-

sented as a node whose regulation by other genes is modeled using updating rules as logic functions. An expressed gene is assigned the value true and a non-expressed gene the value false.

A Boolean network with n genes has 2^n possible states, denoted as $\hat{S}$. At each step in the simulation, the next state $\hat{s}_{t+1} \in \hat{S}$ is determined by applying each gene's logic function (representing the regulatory interactions) to the current value of the genes in $\hat{s}_t$. Let this computation be defined as $\hat{s}_{t+1} \leftarrow D(\hat{s}_t)$ where $D(\hat{s}_t)$ is the deterministic mapping function that finds the next state of the network given the current state. As the network is executed by repeated applications of $D(\hat{s})$, the state will reach a previously visited state, and thus, since the dynamics are deterministic, enter into an attractor which represents a fixed point of the system. Attractors can be single states, called point attractors, or consist of more than one state that the network continuously transitions between, called cyclic attractors. Let $\hat{a} = D^*(\hat{s})$ be the resulting network attractor state reached when starting at $\hat{s}$ and applying the logic functions until the attractor state $\hat{a}$ is reached.

In this work, cell types are considered attractors in the state space of possible gene expression profiles [46] and cell differentiation is modeled as the process of transitioning from one attractor to another [47].

4.5.2 Network construction

A random Boolean regulatory network is generated by randomly connecting a varying number of nodes, then instantiating each node with a randomly generated logic function. To replicate networks found in natural systems, we created only networks that operate in the critical domain, rather than ordered or chaotic. Critical networks implement maximal information flow [39] and have the lowest attractor basin entropy [107]. Evidence that GRN's tend to be critical is given in [38]. To generate critical networks, the parameters are set according to $s = 2qp_N(1-p_N)$ where s is the sensitivity of the network to perturbations in gene values, p_N is the probability of the output of each Boolean function being 1, and q is the count of inputs to each Boolean function [108]. When $s = 1$ a single bit change is on average propagated to one other node and the network is in the critical domain. In

an ordered network, $s < 1$ and perturbations tend to die out, while in a chaotic network, $s > 1$ and perturbations tend to grow. In this work s was fixed at 1 and p_N was adjusted depending upon the value of q.

The attractors of each random Boolean regulatory network are determined, then Markov chain analysis is performed to determine the transition probabilities between all possible states. This allows determination of the MFPTs between each pair of attractors [109]. The MFPTs allow the construction of a graph whose nodes are attractors and weighted edges are the MFPT value between different nodes. Figure 4.8.a shows a sample graph. MFPT graphs for cellular differentiation are expected to have a small MFPT value for forward edges (moving from less to more specialized cell types), large values for reverse edges, and large values in both directions for transitions between attractors at the same level of tree (level is the number of transitions from the root). In [103] a method was introduced that applied successively higher MFPT thresholds to prune edges from this complete MFPT tree as a means to identify separation among subsets of close attractor states as illustrated in Figure 4.8(b). The effects of changing the threshold from low to high was proposed as a possible mechanism for cellular differentiation with the low threshold representing pluripotency and the process of raising the threshold as type specialization as attractors become more and more isolated . This model proposes that cells differentiate by actively controlling their sensitivity of expression noise and can account for the observation that terminally differentiated cell states tend to be more stable than pluripotent states.

4.5.3 Network search

We perform a uniform Monte Carlo search over the space of critical random Boolean networks. For each network we find the attractors and compute the MFPT between all possible attractor pairs (extended from code posted at `http://code.google.com/p/pbn-matlab-toolbox` [109]) . Using the MFPT values, for each type of multistable switch added to the network, we draw a density plot where the x-axis is the forward MFPT and the y-axis is the reverse MFPT (we consider the edge with lower MFPT as forward). The acquired density plots are used to determine the distribution of directional, non-directional, separated and non-separated

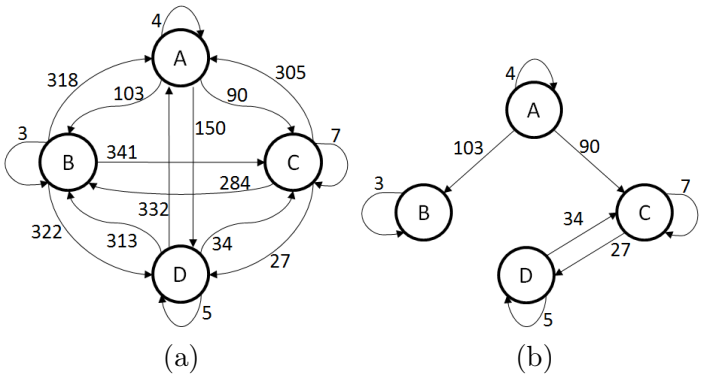


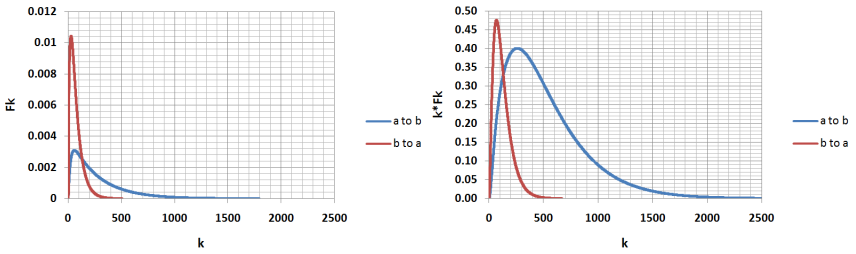
Figure 4.8. (a) A sample MFPT graph. Nodes are attractors and the weights of edges are proportional to MFPT values between attractors. (b) Same graph as in (a) with high (> 103) MFPT edges eliminated.

probability transitions between attractors in each network.

4.5.4 Network types

We investigated 9 types of networks. Approximately $5 * 10^4$ networks of each type were explored to find $5,000$ networks of each type with five or more attractors. The different network types come from the use of the 4 multistable switches that are shown in Figure 4.4. The first switch (Figure 4.4 a) is a bistable switch, a small local circuit with feedback loops. This is a common switch in biological networks and it controls binary branch points between two mutually exclusive cell lineages [5,106]. The truth table of the functions in this switch is [1,1,0,1] for binary numbers [00,01,10,11], respectively. The other three switches all encode mutual inhibition between two genes. The first is *MI00* and is based on the network synthesized in [106]. *MI00* includes two incoherent feedback loops. The final two switches extend mutual inhibition with the addition of one *MI+0* or two *MI++* positive (coherent) feedback loops. These two switches were explored in [110], where it was shown that the positive feedback loops can introduce additional shallow attractor basins in continuous ODE network models.

The four switches were used as described above to construct nine different types of networks: no motif, one BS, one MI00, one MI0+, one MI++ and then four more network



(a) (b)

Figure 4.9. (a) F_k (probability of first visit at time step k) plotted for two arbitrary attractors, called a and b, in a random Boolean network for 2500 steps (k). The red curve is for the transition from a to b that has a low MFPT compared to the reverse transition, b to a (shown with the blue curve); (b) kF_k plotted for the F_k curves in (a). Note that MFPT is the centroid of area under the kF_k curve.

classes each with two of the same switch. Note that when a motif defined in Figure 4.4 is embedded, two nodes of the original RBN are selected randomly, their logic functions replaced and inputs and outputs rewired. For illustration, consider how a MI0+ motif is embedded into a RBN. Starting with a RBN (see Figure 4.4 (a)), two nodes are selected randomly and their truth tables are changed to [0,1,0,0]. Then, the 2^0 input of the second node is wired to the output of first node and, conversely, the 2^0 input of first node is wired to the output of the second node. The small or-gate and not-gate are not considered in wiring, because they were previously considered in the truth tables of their respective nodes.

4.5.5 Mean first passage time

The first-passage time (FPT), also called first hitting time, is the time taken by a stochastic system for the first visit of a specific state. Mathematically, FPT is defined as $F_k(\hat{s}_x, \hat{s}_y)$: the probability that starting in state $\hat{x}$, the first time the system visits a state $\hat{y}$ will be at time k. In the case of Boolean networks, time is the path length of state transitions. Considering p_{xy} as the probability of transition between states x and y, then $F_1(\hat{s}_x, \hat{s}_y) = p_{xy}$. As equation 4.1 shows, for $k \geq 2$, F_k is calculated by a recursive iteration over all transitive relations: for all z states in the network dynamics, $F_k(\hat{s}_x, \hat{s}_y)$ is the probability of a one step transition from state x to z times the FPT from state z to y

in $k-1$ steps.

$$F_k(\hat{s}_x, \hat{s}_y) = \sum_{\hat{s_z} \in \{0,1\}^n, z \neq y} p_{xz} F_{k-1}\left(\hat{s_z}, \hat{s}_y\right) \tag{4.1}$$

Probabilistically, there are two possibilities to reach state y from x ; either y is a deterministic target for x and no bit flips occur due to the noise, or an aggregate of bit flips drive the transition from x to y. So the equation for p_{xy} can be written as follows.

$$p_{xy} = \begin{cases} (1-p_e)^n & y \leftarrow D(\hat{s}_x) \\ p_e^{h_{xy}}(1-p_e)^{n-h_{xy}} & y \leftarrow \eta(\hat{s}_x, h_{xy}), \hat{s}_x \neq \hat{s}_y \end{cases} \tag{4.2}$$

where d_{ij} is equal to 1 if there is a deterministic transition from x to y in the network dynamics, otherwise it is 0; p_e is the probability of a single bit flip resulting from noise and h_{xy} is the Hamming distance between two states; n is the total number of nodes in the network.

Although the *FPT* is a valuable measure, the average time it takes to reach state y from state x, termed Mean First Passage Time (*MFPT*), is of greater interest. MFPT in Boolean networks was introduced by Shmulevich et al. [96] and is defined as:

$$MFPT\left(\hat{s}_x, \hat{s}_y\right) = \sum_k k F_k\left(\hat{s}_x, \hat{s}_y\right) \tag{4.3}$$

A low *MFPT* between two states indicates that starting from the first state, the second state is easily reached by gene expression noise. Figure 4.9 shows F_k, kF_k, and MFPT for the transition between two arbitrary attractors. As this figure shows, the a to b transition has a lower MFPT compared to the other.

At each network state update $D(\hat{s})$ there is a probability that the state will change as a function of the Hamming distance (h) between the current state and the subsequent state $\hat{s}_{t+1} \leftarrow D(\eta(\hat{s_t}, r))$. *MFPT* models uniform gene expression noise by considering probabilistic bit flips at every possible state of the network and deriving the distribution of passage times from analysis of the corresponding Markov process. Statistically, the probability dis-

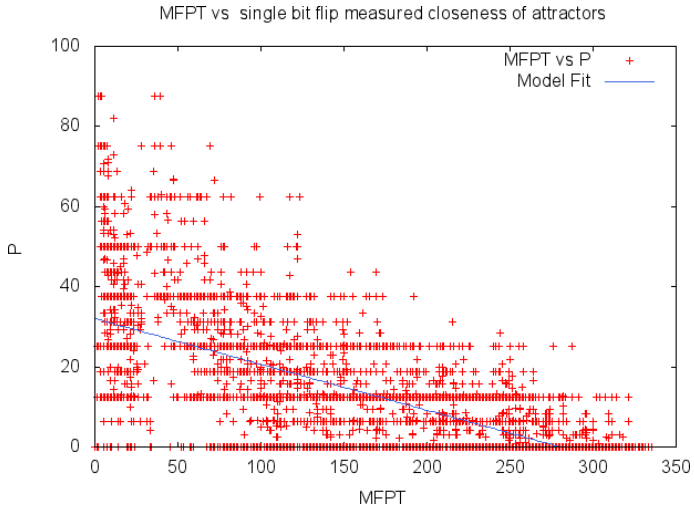


Figure 4.10. Relationship between MFPT and P for 100 critical RBNs.

tribution of bit flips can be seen as a binomial distribution, thus the probability of r bit flips, $\eta(\hat{s}_a, r)$ is $\binom{h}{r}p^r(1-p)^{h-r}$, where p is the probability of a single bit flip and h is the total number of bits.

4.6 Comparisons of epigenetic barrier measures

There are a number of possible ways to measure epigenetic barriers that separate two attractor basins. In this part of the work, the utility of three of these measures, MFPT, transitory bit flips, and Hamming distance, were compared.

4.6.1 Evaluating epigenetic barriers: MFPT vs. transitory bit flips

Villani et al. [103] studied noise-driven network transitions in RBNs. They introduced a measure of the probability of network transitions as the likelihood of attractor transition under expression noise. In this measure, for each pair of attractors $\{a_i, a_j\}$, $P(i,j)$ is the portion of single one-step bit flips (transitory perturbations) in the nodes of all states of attractor a_i which will result in a transition from a_i to a_j under noise-free dynamics. The measure of likelihood of network transition under noise is similar to MFPT, but it does not consider gene expression variability throughout the network. MFPT better models global

expression noise by considering probabilistic bit flips at every possible state of the network and deriving the distribution of passage times from analysis of the corresponding Markov process.

Since one-off bit flips consider noise only as a single bit changes and only when the network has reached its attractor states, it could serve as an efficient yet heuristic measure of the MFPT. To test this idea, a study was performed on a set of small critical networks where for each network and each pair of attractors, $P(i, j)$ was compared with MFPT(i, j). Figure 4.10 depicts the relationship between MFPT and P for 100 arbitrary Boolean networks that have 5 or more attractors. Each point represents the epigenetic barrier between two attractors measured in MFPT and P. Since the networks studied in these experiments are small and do not have many attractors, many points are located in the line $P = 0$. The regression line in this figure shows that as MFPT increases P tends to decrease. P and MFPT are modestly correlated for these small networks and it is unclear how well one-off bit flips can accurately estimate MFPT when network size grows. Since the networks in our experiments are small, we only consider MFPT because of its realism in modeling expression noise.

4.6.2 Evaluating epigenetic barriers: MFPT vs. Hamming distance

An intuitive idea is that MFPT between attractors has a direct relationship to the Hamming distance that separates these attractors. However, we found that this is not the case. Instead, network dynamics, not the Hamming distance, is the main contributor to the MFPT between attractors. As an example of the limitations of Hamming distance, consider that the MFPT(a_i, a_j) and MFPT(a_j, a_i) can be different, but that the Hamming distance between these attractors is the same. However, even though there is not a strong relationship between MFPT and Hamming distance, a weak correlation between the average of the forward and reverse MFPT between attractors and their Hamming distance can be detected. This is depicted in Figure 4.11, which shows MFPT versus the Hamming distance obtained from 100 RBNs containing 8 nodes. As the Hamming distance increases, the upper-bound of MFPT values also increases ($r = 0.1027$ for Hamming distance and average

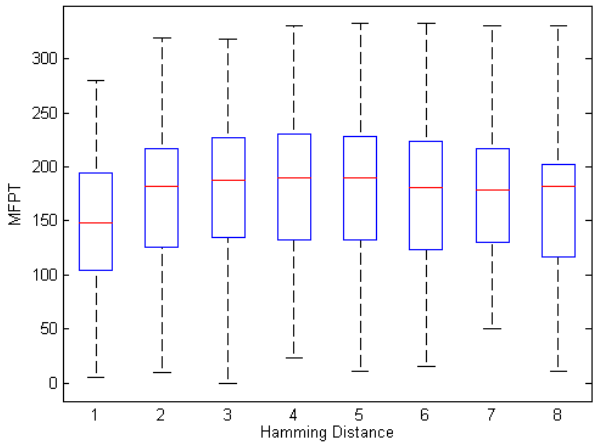


Figure 4.11. Relationship between average MFPT between attractor pairs and Hamming distance for 100 critical BNs.

MFPT). In Figure 4.11, the box represents the central 50% of the points and the red bar shows the median of data.

CHAPTER 5

IN SILICO MODEL OF MORPHOGENESIS IN BIOFILMS

5.1 Abstract

Complex organization of connecting wrinkles observed in many biofilms plays a critical role in survival of these microbial communities. While the underlying genetic causes of wrinkling are not well-understood, recent discoveries have proposed the counterintuitive idea that wrinkles are formed by localized cell death rather than cell growth. This work aims to explore whether the accumulation of vertical and horizontal forces at the areas of cell death initiates the formation of wrinkles.

5.2 Introduction

Bacteria live in almost every environment on the Earth. While they have proved useful in some contexts, for example in microbial fuel production, they are the source of many threats to human health [111–113]. Bacteria can attach to a surface and form a bacteria community encased in an extracellular matrix, referred to as biofilm, which increases the survival rate of their aggregation [111]. Biofilms are responsible for many infections caused by implanted medical devices [114]. Complex organization of connecting wrinkles observable in many biofilms, play two important roles in survival of the microbial community. First, they maximize liquid transport in biofilms by forming high permeable channels connected in a radial network [115]. The aqueous liquid carries nutrients, waste, and signalling molecules. A second role of wrinkles is to increase the waste disposal rate by maximizing the surface area exposed to air [116]. Wrinkles and their formation process have been the target of many microbial research studies [6, 116–119].

A recent study by Asally et al. [6] proposed that localized programmed cell death

initiates wrinkle formation in *B. subtilis* colonies by unlinking cells and the substratum, eliminating anchors that hold the colony in stasis. The well-formed extra cellular matrix (ECM) keeps the cells in place by providing a bond between cells that resists movement against the compressive force from the other cells' growth and division. Cell death disrupts the integrated network of cells and ECM to provide an outlet for compressive stress release [6].

In this work, we aim to model the process of wrinkle formation triggered by cell death using an *in silico* setup. This research employs an agent-based framework to explore whether the accumulation of vertical and horizontal forces at the areas of cell death initiates the wrinkle formation.

5.3 Results

The complex organization of bioflims starts from a single bacterium adhering to a surface. The bacterium secrets a glue-like protein to attach itself to the substratum. Then during division process, the cell cement itself to its daughter [118]. Formation of these cell-cell and cell-surface bonds, along with the pressure arising from population growth, push the colony system to a quasi stable state where there are potential unrelaxed forces dampened by the rigid structure of biofilm. Figure 5.1 demonstrates the potential energy function between two particles. There is an equilibrium point where all competing forces are balanced so the net force applied to particles is zero. If there is no bond, all the particles are expected to be at this point. However, the formation of bonds will gradually prevent relaxation of the repulsive forces. The accumulation of unreleased repulsive forces moves the system to a new quasi stable state as shown in Figure 5.1. The disruption of bonds at the cell-substratum interface, triggered by cell death, provides an outlet for the release of the accumulated forces [6]. This relief of these lateral forces causes the formation of wrinkles at the areas of cell death.

In this study, we used an agent based model extended from [120] to validate the idea that the cell death and the triggered biomechanical forces are sufficient for forming the wrinkles in biofilms. We simulated simple colonies and configured them to be initially compressed

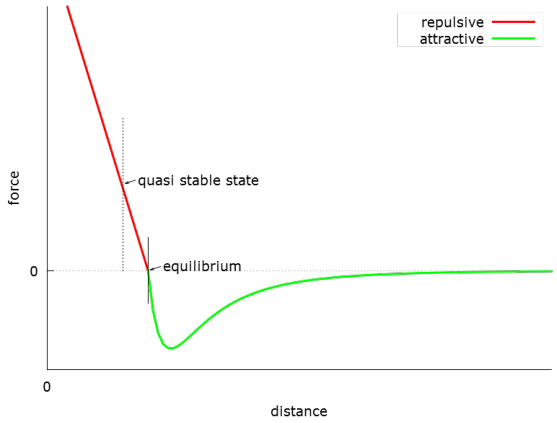


Figure 5.1. Potential energy function: Repulsive and attractive forces between two particles based on the mutual distance. At the equilibrium point the net force between particles is zero. The stiff junctions formed between particles in addition to the pressure resulting from the population growth push the system to the quasi stable state.

having their cell-cell interactions comprised of particles that interact biomechanically in two states: bonded or unassociated. At the next step, we disrupt the connections at the interface of colony and substratum to model cell death. As Figure 5.2 shows, the subsequent movements of the cells cause vertical buckling at the location of cell death. In the left panel of Figure 5.2 (images from [6]), cell death is measured by Sytox Green, a fluorescent cell death marker. The right panel demonstrates our 2D simulation of this process where green color represents the area where we kill the cells to disrupt the connection of biofilm and substratum. Here, particles are the cells and the connections between particles shows the bonds between cells.

To quantitatively track the movement of particles during wrinkle formation, we simulated fluorescent beads placed on the surface of colony. We determined the trajectory of each bead and computed the velocity vectors to replicate the experimental setup in [6]. Figure 5.3 depicts the process of this simulation. Starting from a block of cells in the quasi stable state, we mapped the cell death pattern (CDP) generated by Sytox reporters, as shown in Figure 5.3.A (pattern adopted from [6]), to the bottom layer of this block killing the cells in the mapped area (Figure 5.3.B). As expected, wrinkles form on the top of the areas of cell

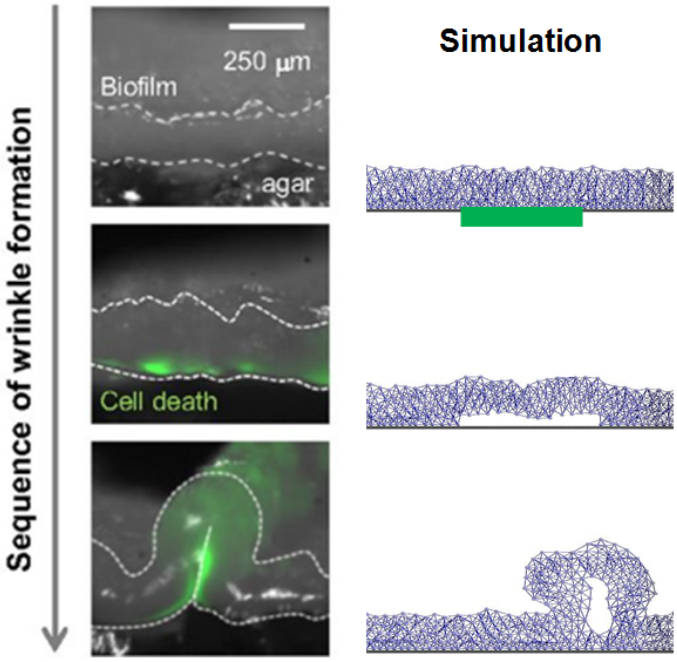


Figure 5.2. Sequence of wrinkle formation originating from cell death at the cell-substratum interface. Left column shows cross-sectional images of a wrinkle from [6] while the right column shows the simulated process. Green color shows the area of cell death.

death as shown in Figure 5.3.C (video can be found in Supplementary materials section). We quantized the surface and by summing the trajectory of the particles in each square, we computed the velocity vector for that partition (blue arrows in Figure 5.3.D). These velocity vectors determine the convergence (negative divergence) of vector fields, demonstrating the aggregate material directional movement. The colored areas in Figure 5.3.D show the convergence; the more intense color, the higher convergence. The observed convergence along with the velocity vectors confirm the counterintuitive idea that wrinkles are formed by cell death rather than local cell growth [118]. Figure 5.3.E shows how cell death areas and wrinkles spatially correlate.

The last experiment is the simulation of the "smiley face" that arises from an artificial CDP designed by Suel group at University of California San Diego [6]. They manually applied cells to regions of a colony to initiate cell death due to high density of cells. The designed cell pattern forms a smiley face on the surface of the biofilm. We simulated this

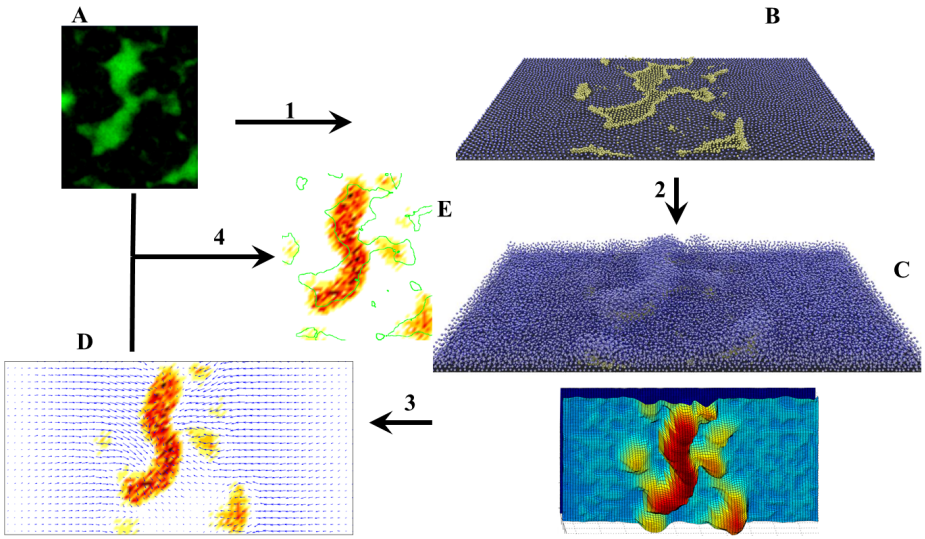


Figure 5.3. Experimental steps taken in this study: A) Cell death pattern (CDP) adopted from [6]. B) CDP mapped to the bottom layer of a colony in which the cells are in a quasi stable state. Note that the upper layer of the colony is not shown in this image. C) Relief of lateral pressure at the CDP area gives rise to the wrinkles. D) Velocity vectors and convergence of vector fields computed from material movement. E) Spatial correlation of CDP and wrinkles.

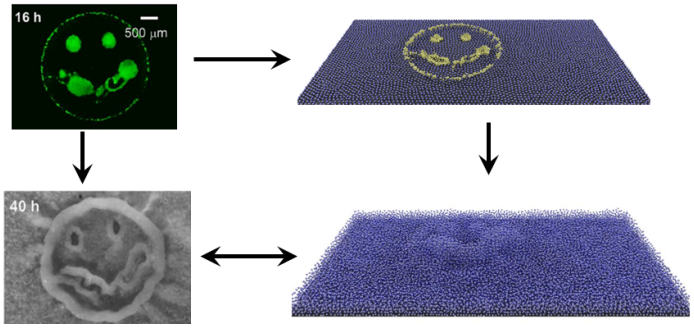


Figure 5.4. Smiley face simulation: an artificially designed CDP, the resulting wrinkled biofilm (images from [6]), and *in silico* formed wrinkles (our simulation).

experiment in our framework shown in Figure 5.4. The simulated wrinkles are in an excellent agreement with the wrinkles in bacterial biofilm.

5.4 Methodology

5.4.1 Domain independent agent-based framework

In silico computer-based modeling has proved effective in many biological research studies [121–124], especially for experiments that are expensive in time and cost. To simulate the process of wrinkle formation, we extended the agent based modeling framework developed by the Kreft group at University of Birmingham, referred to as iDynoMiCs [120]. We modeled cells as particles that mechanically interact. A particle is composed of biomass and regulatory components; it is positioned in space and occupies the volume of a single cell, but does not commit to the cell's specific morphology. Particles grow, divide, and mechanically interact with each other through packing constraints, pressure relief, adhesion, and bonding. Cell state switching is defined through logical expressions that test properties of a particle's micro-environment and internal state. Shoving (where particles are pushed by neighboring particles to relieve packing constraints), bonding, and relief of mechanical stress fields are the main players in the process of wrinkle formation presented in this study. These mechanisms are explained in details in Supplementary Materials section.

5.4.2 Cell death

Cell death is implemented by state switching from a growing to a dying cell. Dying cells do not interact with the environment and gradually are removed from the system as their mass drops below a threshold. In this work cell death patterns are set from segmented Sytox reporter images. However, the spatiotemporal patterns of cell death can be mathematically modeled and be integrated to the framework.

5.4.3 Convergence and velocity fields computation

As mentioned in Results section, we partition the xy plane into a fine grid. For each particle we track the movement and compute the displacement vector. By subtracting the number of particles that enters a grid point from the number of particles that left we can find the convergence (negative divergence) at that particular grid point. The velocity vector

for each square can be computed as the average of all displacement vectors of particles which were initially located at this grid point.

5.4.4 Computation domain

Boundaries of our computation domain are set as rigid surfaces which cells can stick to. For each simulation shown in Figures 5.3 and 5.4 we used 29,250 particles structured in a $45(witdth) \times 130(length) \times 5(height)$ block. We arrange the particles without overlap. However, due to their close packing, there are latent repulsive forces among neighboring particles with the potential to rearrange them in a minimum mechanical interactions structure. The presence of cell-cell and cell-ECM bonds prevents this rearrangements and maintains the particle dynamics to a quasi stable state.

5.5 Discussion

To form a multicellular organism, cells interact via a complex interplay between biochemical signaling and biomechanical forces; however, these interactions are poorly understood. Investigating the morphogenesis of model organisms will provide insights into understanding and formalizing the common patterns seen in many organisms. Having this big plan in mind, we narrowed our research down to multiscale modeling of morphology formation of complex biofilms. Recent studies show that cell death triggered by biochemical stress combined with relaxation of biomechanical forces play a critical role in the initiation of wrinkles in biofilms by eliminating anchors that hold the colony in stasis. This research employs an agent based framework to simulate the wrinkles initiated by cell death events. The results validate the idea that accumulation of vertical and horizontal forces at the cell death area originates the wrinkle formation.

By tracking simulated beads on the surface of colony, we can compute the convergence of lateral movements. The discrepancy between this simulated convergence and experimentally determined convergence can act as an error function. This is the future work of this study to approximate the biomechanical parameters of a system model via minimizing this error function.

The wrinkles simulated in this study are specific morphological features of the whole colony. An important future work is the expansion of the current simulation to a larger spatial scale. We will use Biocellion [125], a high performance computational framework capable of simulating millions of cells, to simulate a whole colony, bringing the simulation close to the scale of the biological systems.

5.6 Supplementary materials

5.6.1 Particle dynamics

A configuration of particles represents the state of the biological system. Simulation configures the updates by first determining the net force acting on each particle, then moving the particles based on their determined force vector. The mechanical forces acting on a particle are computed by vector addition of force contributions of each mechanism in play during the simulation of the system. Since the particles are over damped, inertial effects are ignored and particle velocity is proportional to force. A weak stochastic force η is added to each particle to model underlying fluctuations in cell movement using a Gaussian distribution with a default value for Coefficient of Variation of 0.1; this stochastic movement is an essential component for reaching realexed cellular configurations of near-minimum energy.

The change in momentum of an arbitrary particle σ_i, denoted as $p(\sigma_i)$, is a function of noise and the forces acting on σ_i. These forces can be from pairwise interactions $F_k(\sigma_i, \sigma_j)$ like cell-cell adhesion or overlap, or can be acting on each individual particle $F_l(\sigma_i)$ like cell-surface adhesion (k and l refer to active mechanisms). Changes in momentum of particles due to force is defined as follows:

$$\Delta p(\sigma_i) = (\Sigma_{k \in M} \Sigma_j F_k(\sigma_i, \sigma_j) + \Sigma_{l \in M} F_l(\sigma_i) + \eta)\Delta t \tag{5.1}$$

where $p(\sigma_i)$ is the momentum of particle σ_i, and all neighboring particles are σ_j.

Once forces are generated, a relaxation algorithm is executed to determine the quasi-steady state that minimizes forces acting on the system. The process continues until the magnitude of particle movement drops below some threshold. The complete particle configuration is asynchronously updated by randomly selecting each particle then applying a small momentum based on equation 5.1.

5.6.2 Pairwise particle interaction

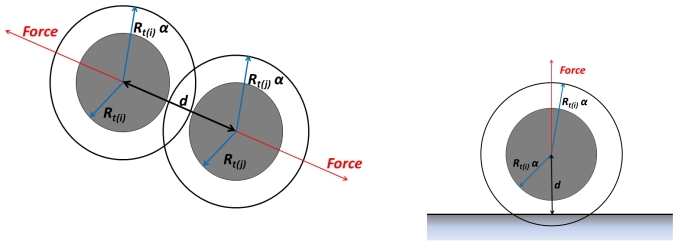
Packing constraints cause particles to exert repellant force (positive) on each other to prevent spatial overlapping caused by growth or cell movement. The process is illustrated in Figure 5.5. In this work, the repellant force magnitude is directly proportional to the overlap distance between each particle. So

$$\begin{aligned} F_{ov}(\sigma_i, \sigma_j) &= (\alpha_t |R_{t_i} + R_{t_j}| - d(\sigma_i, \sigma_j)), \\ d(\sigma_i, \sigma_j) &< \alpha_t |R_{t_i} + R_{t_j}| \end{aligned} \tag{5.2}$$

where R_{t_i} is the designated radius of particle i based on its state $t(\sigma_i)$. The shoving factor α_t determines the average packing density of particles of size R_{t_i} and R_{t_j}. Additionally, nearby particle experience attractive forces due to adhesion. This is represented as potential function applied when $d(\sigma_i, \sigma_j)$ greater than or equal $\alpha_t |R_{t_i} + R_{t_j}|$. Initially attractive forces increase from zero then fall off to zero as the distance increases. The potential function in this case is a generalized Morse function described in [126].

5.6.3 Stiff junctions

Given particles i and j the equilibrium distance between particles is $\alpha_t(R_{t(i)} + R_{t(j)})$. A stiff junction between two particles is modeled as an attractive force when the distance is greater than $\alpha_t(R_{t(i)} + R_{t(j)})$ and a repulsive force when the distance is less than $\alpha_t(R_{t(i)} + R_{t(j)})$. Unlike adhesion, the force between two joined particles does not fall off with distance, rather, as the distance grows, the attractive force between them grows until the bond between particles breaks and the particles become unassociated when their dis-



(a) particle shoving (b) particle/boundary shoving

Figure 5.5. Particle Shoving: $R_t(\sigma_i)$ is the radius of a particle of state $t(\sigma_i)$, α_t is the shoving factor for this state t and d is the distance between the objects. When two particles i and j are closer than $\alpha_t(R_{t_i} - R_{t_j})$ then a force is applied to push them apart. Similarly for an impregnable boundary, but the force is only applied to the particle.

tance passes a threshold. The magnitude of $F_s(\sigma_i, \sigma_j)$ is defined as follows, where $p_s(\sigma_i, \sigma_j)$ is the potential function:

$$\begin{aligned} F_s(\sigma_i, \sigma_j) &= \alpha_t |R_{t(i)} + R_{t(j)}| p_s(\sigma_i, \sigma_j) \\ p_s(\sigma_i, \sigma_j) &= -(x-1)\tanh(s_s|x-1|) \\ x &= \frac{d(\sigma_i, \sigma_j)}{\alpha_t |R_{t(i)} + R_{t(j)}|}, \end{aligned} \tag{5.3}$$

The stiffness of the bond between the particles is controlled by a parameter s_s whose effect is illustrated in Figure 5.6. As s_s grows the forces around the equilibrium point also grow to pull or push the particles back to the equilibrium point, so as to strongly enforce the distance constraint. With low s_s, the distance constraint is lax and the particles are allowed to separate away from the equilibrium distance even when the low forces are applied.

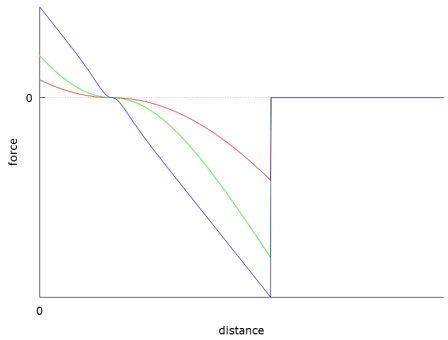


Figure 5.6. Stiff junctions: Stiffness potential function between particles based on the normalized distance between the particles for three different stiffness strength parameter s_s.

CHAPTER 6

CONCLUSIONS

While the international community has not yet reached an agreement on a formal definition of systems biology, there are two aspects which are conserved in all the endless definitions of this emergent field of study: 1) a holistic perspective; and 2) the need for mathematical modeling that links components of the system. The first characteristic separates systems biology from the classical reductionist approaches. The second aspect is essential to understand the big picture of how all the pieces interact in an organism. The only modeling approach that can assure the simultaneous consideration of these two characteristics is multi-scale modeling. This approach can construct a whole system, starting from a very fine scale where molecules are modeled as integrators and controllers of signaling pathways, to the largest scale where the behavior of an entire organ or animal is modeled. Having such a model – where all the inter- and intra-scale interactions are known – is the ultimate goal of systems biology. Although we are only at the beginning of a long path to reaching this goal, the burst in availability of -omics data can help in building data-based and experimentally validated computational models. These models can help in the transition from traditional population-based disease treatments to personalized medicines. For example, in cancer therapy, training a multi-scale model with patient-specific genomic, proteomic, physiologic, and pathological data has the potential to significantly improve the treatment outcomes.

This dissertation focuses on studying two fundamental aspects of developmental biology, cell differentiation and morphogenesis, as a starting point for the multiscale modeling. Three chapters of this dissertation are centered around studying gene regulatory networks (GRNs) as the the underlying controller of cellular differentiation. Random Boolean networks are used here as a tool for modeling GRNs where each network attractor is taken to represent a distinct cell type. Chapter 2 presents a methodology for validating the hypothetical GRNs

using the gene expression data. Chapter 4 introduces a technique and a supporting method for visualization, *CellDiff3D*, that estimates the likelihood and directionality of noise-driven transitions between different cell types. Chapter 4 examines how the attractor structure generated from random Boolean regulatory network dynamics are influenced by the addition of multistable switches commonly found in biological networks that control differentiation.

The last study of this dissertation studies the process of wrinkle formation in biofilms as an example of morphogenesis. This study employs an agent based framework to simulate the wrinkles initiated by localized cell death. The results reveal the importance of biomechanics forces in morphogenesis by showing the role of lateral pressures in the formation of the wrinkles at the cell death areas.

The topics covered in this dissertation can serve as principal steps towards future multiscale studies in systems biology.

REFERENCES

[1] E. C. Butcher, E. L. Berg, and E. J. Kunkel, "Systems biology in drug discovery," *Nat. Biotechnol.*, vol. 22, no. 10, pp. 1253–1259, Oct. 2004. [Online]. Available: http://dx.doi.org/10.1038/nbt1017

[2] J. X. Zhou, A. Samal, A. F. d'Hérouël, N. Price, and S. Huang, "Boolean network landscape for pancreas cell differentiation," *to be published*.

[3] J. Krumsiek, C. Marr, T. Schroeder, and F. J. Theis, "Hierarchical differentiation of myeloid progenitors is encoded in the transcription factor network," *PLoS ONE*, vol. 6, no. 8, pp. e22 649+, Aug. 2011. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0022649

[4] C. H. Waddington, *The Strategy of Genes*. London: George Unwin & Unwin, 1957.

[5] S. Huang, "Reprogramming cell fates: reconciling rarity with robustness." *BioEssays : News and Reviews in Molecular, Cellular and Developmental Biology*, vol. 31, no. 5, pp. 546–560, May 2009. [Online]. Available: http://dx.doi.org/10.1002/bies.200800189

[6] M. Asally, M. Kittisopikul, P. Rué, Y. Du, Z. Hu, T. Çağatay, A. B. Robinson, H. Lu, J. Garcia-Ojalvo, and G. M. Süel, "Localized cell death focuses mechanical forces during 3D patterning in a biofilm," *PNAS*, vol. 109, no. 46, pp. 18 891–18 896, Nov. 2012. [Online]. Available: http://dx.doi.org/10.1073/pnas.1212429109

[7] E. Weinan and B. Engquist, "Multiscale modeling and computation," *Notices Amer. Math. Soc*, vol. 50, no. 50, pp. 1062–1070, 2003. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.97.2226

[8] G. Naldi, L. Pareschi, and G. Toscani, Eds., *Mathematical modeling of collective behavior in socio-economic and life sciences.* Boston: Birkhäuser Boston, 2010. [Online]. Available: http://dx.doi.org/10.1007/978-0-8176-4946-3

[9] V. Khanna and B. R. Bakshi, "Integrated multiscale modeling of economic-environmental systems for assessing Biocomplexity of material use," in *2010 IEEE International Symposium on Sustainable Systems and Technology (ISSST).* IEEE, May 2010, pp. 1–6. [Online]. Available: http://dx.doi.org/10.1109/issst.2010.5507702

[10] M. Vasconcelos, B. Zeigler, and L. Graham, "Modeling multi-scale spatial ecological processes under the discrete event systems paradigm," *Landscape Ecol.*, vol. 8, no. 4, pp. 273–286, 1993. [Online]. Available: http://dx.doi.org/10.1007/bf00125133

[11] C. Bellamy, C. Scott, and J. Altringham, "Multiscale, presence-only habitat suitability models: fine-resolution maps for eight bat species," *J. Appl. Ecol.*, vol. 50, no. 4, pp. 892–901, Aug. 2013. [Online]. Available: http://dx.doi.org/10.1111/1365-2664.12117

[12] R. Costanza, B. Fisher, K. Mulder, S. Liu, and T. Christopher, "Biodiversity and ecosystem services: A multi-scale empirical study of the relationship between species richness and net primary production," *Ecol. Econ.*, vol. 61, no. 2-3, pp. 478–491, Mar. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.ecolecon.2006.03.021

[13] S. A. Baeurle, "Multiscale modeling of polymer materials using field-theoretic methodologies: a survey about recent developments," *J. Math. Chem.*, vol. 46, no. 2, pp. 363–426, Aug. 2009. [Online]. Available: http://dx.doi.org/10.1007/s10910-008-9467-3

[14] J. J. de Pablo, "Coarse-grained simulations of macromolecules: from DNA to nanocomposites." *Annu. Rev. Phys. Chem.*, vol. 62, no. 1, pp. 555–574, 2011. [Online]. Available: http://dx.doi.org/10.1146/annurev-physchem-032210-103458

[15] J. L. Bouvard, D. K. Ward, D. Hossain, S. Nouranian, E. B. Marin, and M. F. Horstemeyer, "Review of hierarchical multiscale modeling to describe the mechanical

behavior of amorphous polymers," *J. Eng. Mater. Technol.*, vol. 131, no. 4, pp. 041 206+, 2009. [Online]. Available: http://dx.doi.org/10.1115/1.3183779

[16] Z. Qu, A. Garfinkel, J. N. Weiss, and M. Nivala, "Multi-scale modeling in biology: How to bridge the gaps between scales?" *Prog. Biophys. Mol. Biol.*, vol. 107, no. 1, pp. 21–31, Oct. 2011. [Online]. Available: http://dx.doi.org/10.1016/j.pbiomolbio.2011.06.004

[17] S. Schnell, R. Grima, and P. Maini, "Multiscale modeling in biology," *Am. Sci.*, vol. 95, no. 2, pp. 134+, 2007. [Online]. Available: http://dx.doi.org/10.1511/2007.64.1018

[18] M. Meier-Schellersheim, I. D. Fraser, and F. Klauschen, "Multiscale modeling for biologists." *Wiley Interdiscip. Rev. Syst. Biol. Med.*, vol. 1, no. 1, pp. 4–14, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1002/wsbm.33

[19] N. Marzari and D. Vanderbilt, "Maximally localized generalized Wannier functions for composite energy bands," *Phys. Rev. B*, vol. 56, no. 20, pp. 12 847–12 865, Nov. 1997. [Online]. Available: http://dx.doi.org/10.1103/physrevb.56.12847

[20] G. Lu and E. Kaxiras, "An overview of multiscale simulations of materials," Jan. 2004. [Online]. Available: http://arxiv.org/abs/cond-mat/0401073.pdf

[21] H. J. Bungartz and M. Griebel, "Sparse grids," *Acta Numerica*, vol. 13, no. -1, pp. 147–269, May 2004. [Online]. Available: http://dx.doi.org/10.1017/s0962492904000182

[22] E. I. Volkova, I. A. Jones, R. Brooks, Y. Zhu, and E. Bichoutskaia, "Sequential multiscale modelling of sic/al nanocomposites reinforced with ws_2 nanoparticles under static loading," *Phys. Rev. B*, vol. 86, no. 10, Sep. 2012. [Online]. Available: http://dx.doi.org/10.1103/physrevb.86.104111

[23] Z. X. Li, T. H. T. Chan, Y. Yu, and Z. H. Sun, "Concurrent multi-scale modeling of civil infrastructures for analyses on structural deterioration—Part I: Modeling methodology and strategy," *Finite Elem. Anal. Des.*, vol. 45, no. 11, pp. 782–794, Sep. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.finel.2009.06.013

[24] M. Tawhai, J. Bischoff, D. Einstein, A. Erdemir, T. Guess, and J. Reinbolt, "Multiscale modeling in computational biomechanics." *IEEE Eng. Med. Biol. Mag.*, vol. 28, no. 3, pp. 41–49, 2009. [Online]. Available: http://dx.doi.org/10.1109/memb.2009.932489

[25] Z. C. Su, V. B. C. Tan, and T. E. Tay, "Concurrent multiscale modeling of amorphous materials in 3D," *Int. J. Numer. Meth. Engng*, vol. 92, no. 13, pp. 1081–1099, Dec. 2012. [Online]. Available: http://dx.doi.org/10.1002/nme.4369

[26] M. L. Martins, S. C. Ferreira, and M. J. Vilela, "Multiscale models for biological systems," *Curr. Opin. Colloid Interface Sci.*, vol. 15, no. 1-2, pp. 18–23, Apr. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.cocis.2009.04.004

[27] J. O. Dada and P. Mendes, "Multi-scale modelling and simulation in systems biology," *Integr. Biol.*, vol. 3, no. 2, pp. 86–96, Feb. 2011. [Online]. Available: http://dx.doi.org/10.1039/c0ib00075b

[28] A. M. Turing, "The chemical basis of morphogenesis," *Philos. Trans. R. Soc. London, Ser. B*, vol. 237, no. 641, pp. 37–72, Aug. 1952. [Online]. Available: http://dx.doi.org/10.1098/rstb.1952.0012

[29] A. Gierer and H. Meinhardt, "A theory of biological pattern formation," *Biol. Cybern.*, vol. 12, no. 1, pp. 30–39, Dec. 1972. [Online]. Available: http://dx.doi.org/10.1007/bf00289234

[30] G. Lintern, "Dynamic patterns: The self organization of brain and behavior," *Complexity*, vol. 2, no. 3, pp. 45–46, Jan. 1997. [Online]. Available: http://dx.doi.org/10.1002/(sici)1099-0526(199701/02)2:3\%3C45::aid-cplx8\%3E3.3.co;2-z

[31] S. A. Levin, "The problem of pattern and scale in ecology: The Robert H. MacArthur award lecture," *Ecology*, vol. 73, no. 6, pp. 1943–1967, Dec. 1992. [Online]. Available: http://dx.doi.org/10.2307/1941447

[32] H. C. Metz, M. Manceau, and H. E. Hoekstra, "Turing patterns: how the fish got its spots," *Pigm. Cell Melan. Res.*, vol. 24, no. 1, pp. 12–14, Feb. 2011. [Online]. Available: http://dx.doi.org/10.1111/j.1755-148x.2010.00814.x

[33] R. Liu, S. Liaw, and P. Maini, "Two-stage Turing model for generating pigment patterns on the leopard and the jaguar," *Phys. Rev. E*, vol. 74, no. 1, Jul. 2006. [Online]. Available: http://dx.doi.org/10.1103/physreve.74.011914

[34] D. C. Walker and J. Southgate, "The virtual cell—a candidate co-ordinator for 'middle-out' modelling of biological systems," *Briefings Bioinf.*, vol. 10, no. 4, pp. 450–461, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1093/bib/bbp010

[35] M. Hecker, S. Lambeck, S. Toepfer, E. van Someren, and R. Guthke, "Gene regulatory network inference: Data integration in dynamic models—A review," *Biosystems*, vol. 96, no. 1, pp. 86–103, Apr. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.biosystems.2008.12.004

[36] N. Berestovsky and L. Nakhleh, "An evaluation of methods for inferring Boolean networks from time-series data," *PLoS ONE*, vol. 8, no. 6, pp. e66 031+, Jun. 2013. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0066031

[37] S. A. Kauffman, "Metabolic stability and epigenesis in randomly constructed genetic nets." *J. Theo. Bio*, vol. 22, no. 3, pp. 437–467, Mar. 1969. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/5803332

[38] M. Nykter, N. D. Price, M. Aldana, S. A. Ramsey, S. A. Kauffman, L. E. Hood, O. Yli-Harja, and I. Shmulevich, "Gene expression dynamics in the macrophage exhibit criticality," *PNAS*, vol. 105, no. 6, pp. 1897–1900, Feb. 2008. [Online]. Available: http://dx.doi.org/10.1073/pnas.0711525105

[39] S. A. Kauffman, *The origins of order: Self-organization and selection in evolution*, 1st ed. Oxford University Press, USA, Jun. 1993. [Online]. Available: http://www.worldcat.org/isbn/0195079515

[40] S. Kauffman, C. Peterson, B. Samuelsson, and C. Troein, "Random Boolean network models and the yeast transcriptional network," *PNAS*, vol. 100, no. 25, pp. 14 796–14 799, Dec. 2003. [Online]. Available: http://dx.doi.org/10.1073/pnas.2036429100

[41] R. Albert and H. G. Othmer. (2003) The topology of the regulatory interactions predicts the expression pattern of the segment polarity genes in Drosophila melanogaster. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.13.3370

[42] J. W. Bodnar, "Programming the *Drosophila* embryo," *J. Theo. Bio*, vol. 188, no. 4, pp. 391–445, Oct. 1997. [Online]. Available: http://dx.doi.org/10.1006/jtbi.1996.0328

[43] A. L. Bauer, T. L. Jackson, Y. Jiang, and T. Rohlf, "Receptor cross-talk in angiogenesis: Mapping environmental cues to cell phenotype using a stochastic, Boolean signaling network model," *J. Theo. Bio*, vol. 264, no. 3, pp. 838–846, Jun. 2010. [Online]. Available: http://dx.doi.org/10.1016/j.jtbi.2010.03.025

[44] I. Shmulevich, S. A. Kauffman, and M. Aldana, "Eukaryotic cells are dynamically ordered or critical but not chaotic," *PNAS*, vol. 102, no. 38, pp. 13 439–13 444, Sep. 2005. [Online]. Available: http://dx.doi.org/10.1073/pnas.0506771102

[45] A. Ghaffarizadeh, N. Flann, and G. Podgorski, "Multistable switches and their role in cellular differentiation networks," *BMC Bioinf.*, vol. 15, no. Suppl 7, pp. S7+, 2014. [Online]. Available: http://dx.doi.org/10.1186/1471-2105-15-s7-s7

[46] A. Huang, L. Hu, S. Kauffman, W. Zhang, and I. Shmulevich, "Using cell fate attractors to uncover transcriptional regulation of HL60 neutrophil differentiation," *BMC Syst. Biol.*, vol. 3, no. 1, pp. 20+, 2009. [Online]. Available: http://dx.doi.org/10.1186/1752-0509-3-20

[47] H. H. Chang, M. Hemberg, M. Barahona, D. E. Ingber, and S. Huang, "Transcriptome-wide noise controls lineage choice in mammalian progenitor cells."

Nature, vol. 453, no. 7194, pp. 544–547, May 2008. [Online]. Available: http://dx.doi.org/10.1038/nature06965

[48] J. X. Zhou, L. Brusch, and S. Huang, "Predicting pancreas cell fate decisions and reprogramming with a hierarchical multi-attractor model," *PLoS ONE*, vol. 6, no. 3, pp. e14 752+, Mar. 2011. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0014752

[49] J. F. Habener, D. M. Kemp, and M. K. Thomas, "Minireview: transcriptional regulation in pancreatic development," *Endocrinology*, vol. 146, no. 3, pp. 1025–1034, Mar. 2005. [Online]. Available: http://dx.doi.org/10.1210/en.2004-1576

[50] J. M. Oliver-Krasinski and D. A. Stoffers, "On the origin of the β cell," *Genes. Dev.*, vol. 22, no. 15, pp. 1998–2021, Aug. 2008. [Online]. Available: http://dx.doi.org/10.1101/gad.1670808

[51] K. S. Zaret and M. Grompe, "Generation and regeneration of cells of the liver and pancreas," *Science*, vol. 322, no. 5907, pp. 1490–1494, Dec. 2008. [Online]. Available: http://dx.doi.org/10.1126/science.1161431

[52] J. Jensen, "Gene regulatory factors in pancreatic development," *Dev. Dyn.*, vol. 229, no. 1, pp. 176–200, Jan. 2004. [Online]. Available: http://dx.doi.org/10.1002/dvdy.10460

[53] J. M. Oliver-Krasinski, M. T. Kasner, J. Yang, M. F. Crutchlow, A. K. Rustgi, K. H. Kaestner, and D. A. Stoffers, "The diabetes gene Pdx1 regulates the transcriptional network of pancreatic endocrine progenitor cells in mice," *J. Clin. Invest.*, vol. 119, no. 7, pp. 1888–1898, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1172/jci37028

[54] T. Masui, G. H. Swift, M. A. Hale, D. M. Meredith, J. E. Johnson, and R. J. Macdonald, "Transcriptional autoregulation controls pancreatic Ptf1a expression

during development and adulthood." *Mol. Cell. Biol.*, vol. 28, no. 17, pp. 5458–5468, Sep. 2008. [Online]. Available: http://dx.doi.org/10.1128/mcb.00549-08

[55] M. Ejarque, S. Cervantes, G. Pujadas, A. Tutusaus, L. Sanchez, and R. Gasa, "Neurogenin3 cooperates with Foxa2 to autoactivate its own expression," *J. Biol. Chem.*, pp. jbc.M112.388 173+, Mar. 2013. [Online]. Available: http://dx.doi.org/10.1074/jbc.m112.388173

[56] S. B. Smith, H. Watada, and M. S. German, "Neurogenin3 activates the islet differentiation program while repressing its own expression," *Mol. Endocrinol.*, vol. 18, no. 1, pp. 142–149, Jan. 2004. [Online]. Available: http://dx.doi.org/10.1210/me.2003-0037

[57] S. B. Smith, H. Watada, D. W. Scheel, C. Mrejen, and M. S. German, "Autoregulation and maturity onset diabetes of the young transcription factors control the human PAX4 promoter," *J. Biol. Chem.*, vol. 275, no. 47, pp. 36 910–36 919, Nov. 2000. [Online]. Available: http://dx.doi.org/10.1074/jbc.m005202200

[58] G. K. Gittes, "Developmental biology of the pancreas: A comprehensive review," *Dev. Biol.*, vol. 326, no. 1, pp. 4–35, Feb. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.ydbio.2008.10.024

[59] S. F. Gilbert, *Developmental Biology, Tenth Edition*, 10th ed. Sinauer Associates, Inc., Jun. 2013. [Online]. Available: http://www.worldcat.org/isbn/160535192X

[60] A. Ghaffarizadeh, G. J. Podgorski, and N. S. Flann, "Modeling and visualizing cell type switching," *Comput. Math. Methods Med.*, vol. 2014, pp. 1–10, 2014. [Online]. Available: http://dx.doi.org/10.1155/2014/293980

[61] S. F. Tsai, E. Strauss, and S. H. Orkin, "Functional analysis and in vivo footprinting implicate the erythroid transcription factor GATA-1 as a positive regulator of its own promoter." *Genes. Dev.*, vol. 5, no. 6, pp. 919–931, Jun. 1991. [Online]. Available: http://dx.doi.org/10.1101/gad.5.6.919

[62] C. D. Trainor, J. G. Omichinski, T. L. Vandergon, A. M. Gronenborn, G. M. Clore, and G. Felsenfeld, "A palindromic regulatory site within vertebrate GATA-1 promoters requires both zinc fingers of the GATA-1 DNA-binding domain for high-affinity interaction." *Mol. Cell. Biol.*, vol. 16, no. 5, pp. 2238–2247, May 1996. [Online]. Available: http://www.ncbi.nlm.nih.gov/pmc/articles/PMC231211/

[63] J. A. Grass, M. E. Boyer, S. Pal, J. Wu, M. J. Weiss, and E. H. Bresnick, "GATA-1-dependent transcriptional repression of GATA-2 via disruption of positive autoregulation and domain-wide chromatin remodeling," *PNAS*, vol. 100, no. 15, pp. 8811–8816, Jul. 2003. [Online]. Available: http://dx.doi.org/10.1073/pnas.1432147100

[64] E. H. Bresnick, H.-Y. Lee, T. Fujiwara, K. D. Johnson, and S. Keles, "GATA switches as developmental drivers," *J. Biol. Chem.*, vol. 285, no. 41, pp. 31 087–31 093, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1074/jbc.r110.159079

[65] H. Iwasaki, S.-i. Mizuno, R. A. Wells, A. B. Cantor, S. Watanabe, and K. Akashi, "GATA-1 converts lymphoid and myelomonocytic progenitors into the megakaryocyte/erythrocyte lineages," *Immunity*, vol. 19, no. 3, pp. 451–462, Sep. 2003. [Online]. Available: http://dx.doi.org/10.1016/s1074-7613(03)00242-5

[66] M. Crossley, A. P. Tsang, J. J. Bieker, and S. H. Orkin, "Regulation of the erythroid Kruppel-like factor (EKLF) gene promoter by the erythroid transcription factor GATA-1." *J. Biol. Chem.*, vol. 269, no. 22, pp. 15 440–15 444, Jun. 1994. [Online]. Available: http://www.jbc.org/content/269/22/15440.abstract

[67] B. Barbeau, C. Barat, D. Bergeron, and E. Rassart, "The GATA-1 and Spi-1 transcriptional factors bind to a GATA/EBS dual element in the Fli-1 exon 1." *Oncogene*, vol. 18, no. 40, pp. 5535–5545, Sep. 1999. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/10523830

[68] E. O. Bockamp, F. McLaughlin, A. M. Murrell, B. Gottgens, L. Robb, C. G. Begley, and A. R. Green, "Lineage-restricted regulation of the murine SCL/TAL-1

promoter," *Blood*, vol. 86, no. 4, pp. 1502–1514, Aug. 1995. [Online]. Available: http://bloodjournal.hematologylibrary.org/content/86/4/1502.abstract

[69] P. Zhang, G. Behre, J. Pan, A. Iwama, N. Wara-aswapati, H. S. Radomska, P. E. Auron, D. G. Tenen, and Z. Sun, "Negative cross-talk between hematopoietic regulators: GATA proteins repress PU.1," *PNAS*, vol. 96, no. 15, pp. 8705–8710, Jul. 1999. [Online]. Available: http://dx.doi.org/10.1073/pnas.96.15.8705

[70] S. T. Chou, E. Khandros, L. C. Bailey, K. E. Nichols, C. R. Vakoc, Y. Yao, Z. Huang, J. D. Crispino, R. C. Hardison, G. A. Blobel, and M. J. Weiss, "Graded repression of PU.1/Sfpi1 gene transcription by GATA factors regulates hematopoietic cell fate," *Blood*, vol. 114, no. 5, pp. 983–994, Jul. 2009. [Online]. Available: http://dx.doi.org/10.1182/blood-2009-03-207944

[71] N. Rekhtman, F. Radparvar, T. Evans, and A. I. Skoultchi, "Direct interaction of hematopoietic transcription factors PU.1 and GATA-1: functional antagonism in erythroid cells," *Genes. Dev.*, vol. 13, no. 11, pp. 1398–1411, Jun. 1999. [Online]. Available: http://genesdev.cshlp.org/content/13/11/1398.abstract

[72] K. Ohneda and M. Yamamoto, "Roles of hematopoietic transcription factors GATA-1 and GATA-2 in the development of red blood cell lineage," *Acta Haematologica*, vol. 108, no. 4, pp. 237–245, 2002. [Online]. Available: http://dx.doi.org/10.1159/000065660

[73] V. Chickarmane, T. Enver, and C. Peterson, "Computational modeling of the hematopoietic erythroid-myeloid switch reveals insights into cooperativity, priming, and irreversibility," *PLoS Comput Biol.*, vol. 5, no. 1, pp. e1 000 268+, Jan. 2009. [Online]. Available: http://dx.doi.org/10.1371/journal.pcbi.1000268

[74] J. Starck, N. Cohet, C. Gonnet, S. Sarrazin, Z. Doubeikovskaia, A. Doubeikovski, A. Verger, M. Duterque-Coquillaud, and F. Morle, "Functional cross-antagonism between transcription factors FLI-1 and EKLF," *Mol. Cell. Biol.*, vol. 23, no. 4,

pp. 1390–1402, Feb. 2003. [Online]. Available: http://dx.doi.org/10.1128/mcb.23.4.1390-1402.2003

[75] M. Athanasiou, G. Mavrothalassitis, L. Sun-Hoffman, and D. G. Blair, "FLI-1 is a suppressor of erythroid differentiation in human hematopoietic cells." *Leukemia*, vol. 14, no. 3, pp. 439–445, Mar. 2000. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/10720139

[76] A. D. Friedman, "C/EBPa induces PU.1 and interacts with AP-1 and NF-KB to regulate myeloid development," *Blood Cell. Mol. Dis.*, vol. 39, no. 3, pp. 340–343, Nov. 2007. [Online]. Available: http://dx.doi.org/10.1016/j.bcmd.2007.06.010

[77] C. Yeamans, D. Wang, I. Paz-Priel, B. E. Torbett, D. G. Tenen, and A. D. Friedman, "C/EBPa binds and activates the PU.1 distal enhancer to induce monocyte lineage commitment," *Blood*, vol. 110, no. 9, pp. 3136–3142, Nov. 2007. [Online]. Available: http://dx.doi.org/10.1182/blood-2007-03-080291

[78] P. Laslo, C. J. Spooner, A. Warmflash, D. W. Lancki, H.-J. Lee, R. Sciammas, B. N. Gantner, A. R. Dinner, and H. Singh, "Multilineage transcriptional priming and determination of alternate hematopoietic cell fates," *Cell*, vol. 126, no. 4, pp. 755–766, Aug. 2006. [Online]. Available: http://dx.doi.org/10.1016/j.cell.2006.06.052

[79] M. R. R. Lidonnici, A. Audia, A. R. R. Soliera, M. Prisco, G. Ferrari-Amorotti, T. Waldron, N. Donato, Y. Zhang, R. V. Martinez, T. L. Holyoake, and B. Calabretta, "Expression of the transcriptional repressor gfi-1 is regulated by c/ebpalpha and is involved in its proliferation and colony formation-inhibitory effects in p210bcr/abl-expressing cells." *Cancer Res.*, vol. 70, no. 20, pp. 7949–7959, Oct. 2010. [Online]. Available: http://dx.doi.org/10.1158/0008-5472.can-10-1667

[80] M. Le Clech, E. Chalhoub, C. Dohet, V. Roure, S. Fichelson, F. Moreau-Gachelin, and D. Mathieu, "PU.1/Spi-1 binds to the human TAL-1 silencer to mediate its activity," *J. Mol. Biol.*, vol. 355, no. 1, pp. 9–19, Jan. 2006. [Online]. Available: http://dx.doi.org/10.1016/j.jmb.2005.10.055

[81] Y. Okuno, G. Huang, F. Rosenbauer, E. K. Evans, H. S. Radomska, H. Iwasaki, K. Akashi, F. Moreau-Gachelin, Y. Li, P. Zhang, B. Göttgens, and D. G. Tenen, "Potential autoregulation of transcription factor PU.1 by an upstream regulatory element," *Mol. Cell. Biol.*, vol. 25, no. 7, pp. 2832–2845, Apr. 2005. [Online]. Available: http://dx.doi.org/10.1128/mcb.25.7.2832-2845.2005

[82] M. Leddin, C. Perrod, M. Hoogenkamp, S. Ghani, S. Assi, S. Heinz, N. K. Wilson, G. Follows, J. Schönheit, L. Vockentanz, A. M. Mosammam, W. Chen, D. G. Tenen, D. R. Westhead, B. Göttgens, C. Bonifer, and F. Rosenbauer, "Two distinct auto-regulatory loops operate at the PU.1 locus in B cells and myeloid cells," *Blood*, vol. 117, no. 10, pp. 2827–2838, Mar. 2011. [Online]. Available: http://dx.doi.org/10.1182/blood-2010-08-302976

[83] U. Steidl, F. Rosenbauer, R. G. W. Verhaak, X. Gu, A. Ebralidze, H. H. Otu, S. Klippel, C. Steidl, I. Bruns, D. B. Costa, K. Wagner, M. Aivado, G. Kobbe, P. J. M. Valk, E. Passegue, T. A. Libermann, R. Delwel, and D. G. Tenen, "Essential role of Jun family transcription factors in PU.1 knockdown-induced leukemic stem cells," *Nat Genet*, vol. 38, no. 11, pp. 1269–1277, Nov. 2006. [Online]. Available: http://dx.doi.org/10.1038/ng1898

[84] R. Dahl, S. R. Iyer, K. S. Owens, D. D. Cuylear, and M. C. Simon, "The transcriptional repressor GFI-1 antagonizes PU.1 activity through protein-protein interaction," *J. Biol. Chem.*, vol. 282, no. 9, pp. 6473–6483, Mar. 2007. [Online]. Available: http://dx.doi.org/10.1074/jbc.m607613200

[85] T. Masaki, J. Qu, J. Cholewa-Waclaw, K. Burr, R. Raaum, and A. Rambukkana, "Reprogramming adult Schwann cells to stem cell-like cells by leprosy bacilli promotes dissemination of infection," *Cell*, vol. 152, no. 1-2, pp. 51–67, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.cell.2012.12.014

[86] J. P. Thiery, "Epithelial–mesenchymal transitions in tumour progression," *Nat. Rev. Cancer*, vol. 2, no. 6, pp. 442–454, Jun. 2002. [Online]. Available: http://dx.doi.org/10.1038/nrc822

[87] S. Huang, G. Eichler, Y. B. Yam, and D. E. Ingber, "Cell fates as high-dimensional attractor states of a complex gene regulatory network," *Phys. Rev. Lett.*, vol. 94, no. 12, pp. 128 701+, Apr. 2005. [Online]. Available: http://dx.doi.org/10.1103/physrevlett.94.128701

[88] D. Orrell and H. Bolouri, "Control of internal and external noise in genetic regulatory networks," in *J. Theo. Bio*, 2004, pp. 301–312. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.85.229

[89] S. Huang, I. Ernberg, and S. Kauffman, "Cancer attractors: a systems view of tumors from a gene network dynamics and developmental perspective." *Semin. Cell Dev. Biol.*, vol. 20, no. 7, pp. 869–876, Sep. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.semcdb.2009.07.003

[90] L. T. Macneil and A. J. Walhout, "Gene regulatory networks and the role of robustness and stochasticity in the control of gene expression." *Genome Res*, vol. 21, no. 5, pp. 645–657, May 2011. [Online]. Available: http://dx.doi.org/10.1101/gr.097378.109

[91] Q. Hu, A. M. Friedrich, L. V. Johnson, and D. O. Clegg, "Memory in induced pluripotent stem cells: reprogrammed human retinal-pigmented epithelial cells show tendency for spontaneous redifferentiation." *Stem Cells (Dayton, Ohio)*, vol. 28, no. 11, pp. 1981–1991, Nov. 2010. [Online]. Available: http://dx.doi.org/10.1002/stem.531

[92] C. Furusawa and K. Kaneko, "A dynamical-systems view of stem cell biology," *Science*, vol. 338, no. 6104, pp. 215–217, Oct. 2012. [Online]. Available: http://dx.doi.org/10.1126/science.1224311

[93] N. Suzuki, C. Furusawa, and K. Kaneko, "Oscillatory protein expression dynamics endows Stem cells with robust differentiation potential," *PLoS ONE*, vol. 6, no. 11,

pp. e27 232+, Nov. 2011. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0027232

[94] J. X. Zhou, M. D. S. Aliyu, E. Aurell, and S. Huang, "Quasi-potential landscape in complex multi-stable systems," *J. R. Soc. Interface*, vol. 9, no. 77, pp. 3539–3553, Aug. 2012. [Online]. Available: http://dx.doi.org/10.1098/rsif.2012.0434

[95] S. Bhattacharya, Q. Zhang, and M. Andersen, "A deterministic map of Waddington's epigenetic landscape for cell fate specification," *BMC Syst. Biol.*, vol. 5, no. 1, pp. 85+, 2011. [Online]. Available: http://dx.doi.org/10.1186/1752-0509-5-85

[96] I. Shmulevich, E. R. Dougherty, and W. Zhang, "Gene perturbation and intervention in probabilistic Boolean networks," *Bioinformatics*, vol. 18, no. 10, pp. 1319–1331, Oct. 2002. [Online]. Available: http://dx.doi.org/10.1093/bioinformatics/18.10.1319

[97] Y. Xue, K. Ouyang, J. Huang, Y. Zhou, H. Ouyang, H. Li, G. Wang, Q. Wu, C. Wei, Y. Bi, L. Jiang, Z. Cai, H. Sun, K. Zhang, Y. Zhang, J. Chen, and X.-D. D. Fu, "Direct conversion of fibroblasts to neurons by reprogramming PTB-regulated microRNA circuits." *Cell*, vol. 152, no. 1-2, pp. 82–96, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1016/j.cell.2012.11.045

[98] T. Vierbuchen, A. Ostermeier, Z. P. Pang, Y. Kokubu, T. C. Södhof, and M. Wernig, "Direct conversion of fibroblasts to functional neurons by defined factors," *Nature*, vol. 463, no. 7284, pp. 1035–1041, Jan. 2010. [Online]. Available: http://dx.doi.org/10.1038/nature08797

[99] J. Ellson, E. R. Gansner, E. Koutsofios, S. C. North, and G. Woodhull, "Graphviz and dynagraph - static and dynamic graph drawing tools," in *Graph Drawing Software*, 2003, pp. 127–148. [Online]. Available: http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.96.3776

[100] A. Bar-Even, J. Paulsson, N. Maheshri, M. Carmi, E. O'Shea, Y. Pilpel, and N. Barkai, "Noise in protein expression scales with natural protein abundance,"

Nature Gen., vol. 38, no. 6, pp. 636–643, May 2006. [Online]. Available: http://dx.doi.org/10.1038/ng1807

[101] G. Tkačik, T. Gregor, and W. Bialek, "The role of input noise in transcriptional regulation," *PLoS ONE*, vol. 3, no. 7, pp. e2774+, Jul. 2008. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0002774

[102] L. Bruno, R. Hoffmann, F. McBlane, J. Brown, R. Gupta, C. Joshi, S. Pearson, T. Seidl, C. Heyworth, and T. Enver, "Molecular signatures of self-renewal, differentiation, and lineage choice in multipotential hemopoietic progenitor cells in vitro." *Mol. Cell. Biol.*, vol. 24, no. 2, pp. 741–756, Jan. 2004. [Online]. Available: http://dx.doi.org/10.1128/MCB.24.2.741

[103] M. Villani, A. Barbieri, and R. Serra, "A dynamical model of genetic networks for cell differentiation," *PLoS ONE*, vol. 6, no. 3, pp. e17 703+, Mar. 2011. [Online]. Available: http://dx.doi.org/10.1371/journal.pone.0017703

[104] J.-R. R. Kim, Y. Yoon, and K.-H. H. Cho, "Coupled feedback loops form dynamic motifs of cellular networks." *Biophys. J.*, vol. 94, no. 2, pp. 359–365, Jan. 2008. [Online]. Available: http://dx.doi.org/10.1529/biophysj.107.105106

[105] S. Huang, "Shape-dependent control of cell growth, differentiation, and apoptosis: switching between attractors in cell regulatory networks," *Exp. Cell. Res.*, vol. 261, no. 1, pp. 91–103, Nov. 2000. [Online]. Available: http://dx.doi.org/10.1006/excr.2000.5044

[106] T. S. Gardner, C. R. Cantor, and J. J. Collins, "Construction of a genetic toggle switch in Escherichia coli," *Nature*, vol. 403, no. 6767, pp. 339–342, Jan. 2000. [Online]. Available: http://dx.doi.org/10.1038/35002131

[107] P. Krawitz and I. Shmulevich, "Basin entropy in Boolean network ensembles." *Phys. Rev. Lett.*, vol. 98, no. 15, Apr. 2007. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/17501391

[108] B. Derrida and Y. Pomeau, "Random networks of automata: A simple annealed approximation," *Europhys. Lett.*, vol. 1, no. 2, pp. 45–49, Jan. 1986. [Online]. Available: http://dx.doi.org/10.1209/0295-5075/1/2/001

[109] I. Shmulevich and E. R. Dougherty, *Probabilistic boolean networks: The modeling and control of gene regulatory networks.* SIAM, 2009. [Online]. Available: http://portal.acm.org/citation.cfm?id=1734075

[110] J. Macía, S. Widder, and R. Solé, "Why are cellular switches Boolean? General conditions for multistable genetic circuits." *J. Theo. Bio*, vol. 261, no. 1, pp. 126–135, Nov. 2009. [Online]. Available: http://dx.doi.org/10.1016/j.jtbi.2009.07.019

[111] R. M. Donlan and J. W. Costerton, "Biofilms: survival mechanisms of clinically relevant microorganisms." *Clin. Microbiol. Rev.*, vol. 15, no. 2, pp. 167–193, Apr. 2002. [Online]. Available: http://dx.doi.org/10.1128/cmr.15.2.167-193.2002

[112] P. Hunter, "The mob response. The importance of biofilm research for combating chronic diseases and tackling contamination." *EMBO Reports*, vol. 9, no. 4, pp. 314–317, Apr. 2008. [Online]. Available: http://dx.doi.org/10.1038/embor.2008.43

[113] A. Edlund, Y. Yang, A. P. Hall, L. Guo, R. Lux, X. He, K. E. Nelson, K. H. Nealson, S. Yooseph, W. Shi, and J. S. McLean, "An in vitro biofilm model system maintaining a highly reproducible species and metabolic diversity approaching that of the human oral microbiome," *Microbiome*, vol. 1, no. 1, pp. 25+, 2013. [Online]. Available: http://dx.doi.org/10.1186/2049-2618-1-25

[114] J. W. Costerton, L. Montanaro, and C. R. Arciola, "Biofilm in implant infections: its production and regulation." *Int. J. Artif. Organs*, vol. 28, no. 11, pp. 1062–1068, Nov. 2005. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/16353112

[115] J. N. Wilking, V. Zaburdaev, M. De Volder, R. Losick, M. P. Brenner, and D. A. Weitz, "Liquid transport facilitated by channels in Bacillus subtilis

biofilms." *PNAS*, vol. 110, no. 3, pp. 848–852, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1073/pnas.1216376110

[116] L. E. P. Dietrich, C. Okegbe, A. Price-Whelan, H. Sakhtah, R. C. Hunter, and D. K. Newman, "Bacterial community morphogenesis is intimately linked to the intracellular redox state," *J. Bacteriology*, vol. 195, no. 7, pp. 1371–1380, Apr. 2013. [Online]. Available: http://dx.doi.org/10.1128/jb.02273-12

[117] S. Payne, B. Li, Y. Cao, D. Schaeffer, M. D. Ryser, and L. You, "Temporal control of self-organized pattern formation without morphogen gradients in bacteria," *Mol. Syst. Biol.*, vol. 9, no. 1, p. n/a, Jan. 2013. [Online]. Available: http://dx.doi.org/10.1038/msb.2013.55

[118] D. Schultz, J. N. Onuchic, and E. Ben-Jacob, "Turning death into creative force during biofilm engineering," *PNAS*, vol. 109, no. 46, pp. 18 633–18 634, Nov. 2012. [Online]. Available: http://dx.doi.org/10.1073/pnas.1215227109

[119] W. H. DePas, D. A. Hufnagel, J. S. Lee, L. P. Blanco, H. C. Bernstein, S. T. Fisher, G. A. James, P. S. Stewart, and M. R. Chapman, "Iron induces bimodal population development by Escherichia coli," *PNAS*, vol. 110, no. 7, pp. 2629–2634, Feb. 2013. [Online]. Available: http://dx.doi.org/10.1073/pnas.1218703110

[120] L. A. Lardon, B. V. Merkey, S. Martins, A. Dötsch, C. Picioreanu, J.-U. U. Kreft, and B. F. Smets, "iDynoMiCS: next-generation individual-based modelling of biofilms." *Environ. Microbiol.*, vol. 13, no. 9, pp. 2416–2434, Sep. 2011. [Online]. Available: http://dx.doi.org/10.1111/j.1462-2920.2011.02414.x

[121] J. A. Lerman, D. R. Hyduke, H. Latif, V. A. Portnoy, N. E. Lewis, J. D. Orth, A. C. Schrimpe-Rutledge, R. D. Smith, J. N. Adkins, K. Zengler, and B. O. Palsson, "In silico method for modelling metabolism and gene product expression at genome scale," *Nat. Commun.*, vol. 3, pp. 929+, Jul. 2012. [Online]. Available: http://dx.doi.org/10.1038/ncomms1928

[122] A. V. Ratushny, S. A. Ramsey, O. Roda, Y. Wan, J. J. Smith, and J. D. Aitchison, "Control of transcriptional variability by overlapping feed-forward regulatory motifs." *Biophys. J.*, vol. 95, no. 8, pp. 3715–3723, Oct. 2008. [Online]. Available: http://dx.doi.org/10.1529/biophysj.108.134064

[123] T. Kim, K. A. Afonin, M. Viard, A. Y. Koyfman, S. Sparks, E. Heldman, S. Grinberg, C. Linder, R. P. Blumenthal, and B. A. Shapiro, "In silico, in vitro, and in vivo studies indicate the potential use of bolaamphiphiles for therapeutic siRNAs delivery," *Mol. Ther. — Nucleic Acids*, vol. 2, no. 3, pp. e80+, Mar. 2013. [Online]. Available: http://dx.doi.org/10.1038/mtna.2013.5

[124] D. Murray, P. Doran, P. MacMathuna, and A. C. Moss, "In silico gene expression analysis–an overview." *Mol. Cancer*, vol. 6, no. 1, pp. 50+, Aug. 2007. [Online]. Available: http://dx.doi.org/10.1186/1476-4598-6-50

[125] N. S. Flann, A. Ghaffarizadeh, G. J. Podgorski, S. Kang, S. Kahan, J. E. Mcdermott, and I. Shmulevich, "BioCellion: A high-performance computing framework for multi-scale modeling and simulation of multicellular biological systems," in *The Society for Mathematical Biology Annual Meeting and Conference*, Jul. 2012.

[126] T. J. Newman, "Modeling multicellular systems using subcellular elements." *Math. Biosci. Engr: MBE*, vol. 2, no. 3, pp. 613–624, Jul. 2005. [Online]. Available: http://view.ncbi.nlm.nih.gov/pubmed/20369943

APPENDIX

Hindawi Publishing Corporation
Computational and Mathematical Methods in Medicine
Volume 2014, Article ID 293980, 10 pages
http://dx.doi.org/10.1155/2014/293980

Research Article

Modeling and Visualizing Cell Type Switching

Ahmadreza Ghaffarizadeh,[1] Gregory J. Podgorski,[2,3] and Nicholas S. Flann[1,4,5]

[1] *Computer Science Department, Utah State University, Logan, UT 84322, USA*
[2] *Biology Department, Utah State University, Logan, UT 84322, USA*
[3] *Center for Integrated BioSystems, Utah State University, Logan, UT 84322, USA*
[4] *Institute for Systems Biology, Seattle, WA 98109, USA*
[5] *Synthetic Biomanufacturing Institute, Logan, UT 84322, USA*

Correspondence should be addressed to Nicholas S. Flann; nick.flann@usu.edu

Received 30 September 2013; Revised 20 December 2013; Accepted 10 January 2014; Published 14 April 2014

Academic Editor: Marco Villani

Understanding cellular differentiation is critical in explaining development and for taming diseases such as cancer. Differentiation is conventionally represented using bifurcating lineage trees. However, these lineage trees cannot readily capture or quantify all the types of transitions now known to occur between cell types, including transdifferentiation or differentiation off standard paths. This work introduces a new analysis and visualization technique that is capable of representing all possible transitions between cell states compactly, quantitatively, and intuitively. This method considers the regulatory network of transcription factors that control cell type determination and then performs an analysis of network dynamics to identify stable expression profiles and the potential cell types that they represent. A visualization tool called *CellDiff3D* creates an intuitive three-dimensional graph that shows the overall direction and probability of transitions between all pairs of cell types within a lineage. In this study, the influence of gene expression noise and mutational changes during myeloid cell differentiation are presented as a demonstration of the *CellDiff3D* technique, a new approach to quantify and envision all possible cell state transitions in any lineage network.

1. Introduction

During development, a complex system of tissues and organs emerges from a single cell by the coordination of cell division, morphogenesis, and differentiation. Understanding the differentiation of cell types is necessary to understanding development and its associated defects, for improved control of stem cell differentiation in therapeutic use and for taming diseases such as cancer. Cellular differentiation occurs when a less specialized cell or its progeny becomes increasingly specialized by acquiring properties that allow specific functions. In animals, differentiation typically results in a terminally differentiated state in which a specialized cell can no longer acquire the properties of other specialized adult cells. Recent discoveries, however, have shown that terminally differentiated cells can be reprogrammed to revert back to multipotent and pluripotent stem cells which have the potential to differentiate into other cell types [1, 2] or to transdifferentiate into other specialized cell types [3].

Differentiating cells normally follow well defined paths to mature cell types. Taken together, these paths are referred to as a lineage tree. Pluripotent stem cells give rise to progeny that specialize into more constrained multipotent cells. In turn, multipotent cells produce a variety of stable, terminally differentiated cells. This process is usually depicted as a tree with a pluripotent cell at its root, multipotent cells as intermediate nodes, and the mature cell types as branch tips. As an example, a simplified portion of the myeloid cell lineage tree is illustrated in Figure 1. This figure shows that common myeloid progenitor stem cells produce two pluripotent cell types, a megakaryocyte-erythrocyte progenitor and a granulocyte-monocyte progenitor, that in turn produce terminally differentiated erythrocytes, megakaryocytes, monocytes, and granulocytes.

Ghaffarizadeh *et al. BMC Bioinformatics* 2014, **15**(Suppl 7):S7
http://www.biomedcentral.com/1471-2105/15/S7/S7

BMC Bioinformatics

RESEARCH Open Access

Multistable switches and their role in cellular differentiation networks

Ahmadreza Ghaffarizadeh[1], Nicholas S Flann[1,3,4*], Gregory J Podgorski[2,5]

From The 10th Annual Biotechnology and Bioinformatics Symposium (BIOT 2013)
Provo, UT, USA. 5-6 December 2013

Abstract

Background: Cellular differentiation during development is controlled by gene regulatory networks (GRNs). This complex process is always subject to gene expression noise. There is evidence suggesting that commonly seen patterns in GRNs, referred to as biological multistable switches, play an important role in creating the structure of lineage trees by providing stability to cell types.

Results: To explore this question a new methodology is developed and applied to study (a) the multistable switch-containing GRN for hematopoiesis and (b) a large set of random boolean networks (RBNs) in which multistable switches were embedded systematically. In this work, each network attractor is taken to represent a distinct cell type. The GRNs were seeded with one or two identical copies of each multistable switch and the effect of these additions on two key aspects of network dynamics was assessed. These properties are the barrier to movement between pairs of attractors (separation) and the degree to which one direction of movement between attractor pairs is favored over another (directionality). Both of these properties are instrumental in shaping the structure of lineage trees. We found that adding one multistable switch of any type had a modest effect on increasing the proportion of well-separated attractor pairs. Adding two identical switches of any type had a much stronger effect in increasing the proportion of well-separated attractors. Similarly, there was an increase in the frequency of directional transitions between attractor pairs when two identical multistable switches were added to GRNs. This effect on directionality was not observed when only one multistable switch was added.

Conclusions: This work provides evidence that the occurrence of multistable switches in networks that control cellular differentiation contributes to the structure of lineage trees and to the stabilization of cell types.

Introduction

Understanding differentiation is critical to knowing how normal development unfolds and for taming diseases, such as cancer, that are associated with defects or reversals in differentiation. In animals, the process of differentiation typically results in cells reaching a terminally differentiated state. However, recent discoveries have shown that "terminal differentiation" may be a misnomer as fully differentiated cells can be reprogrammed to revert back to a pluripotent state, with these pluripotent cells having the potential to differentiate into other cell types.

Transitions between cell types can be mapped as a directed tree of cell types, known as a lineage tree, with embryonic stem cells at the root, various classes of precursor cells as internal nodes, and terminally differentiated cells as branch tips. Gene regulatory networks (GRNs) that respond to both external stimuli and to gene expression noise control transitions between cell types and determine the structure of lineage trees [1]. Given that differentiation is driven by the output of dynamic gene regulatory networks, a useful, network-based perspective for envisioning different stable cell types is as basins in an attractor landscape [2,3]. In this dynamical systems view, differentiation is the process of moving between the different attractor basins that are generated by the dynamics of the gene regulatory network.

* Correspondence: nick.flann@usu.edu
[1]Computer Science Department, Utah State University, Logan, UT 84322, USA
Full list of author information is available at the end of the article

VITA

Ahmadreza Ghaffarizadeh

Computer Science Department, Utah State University, Logan, UT
Institute for Systems Biology, Seattle, WA
ghaffarizadeh@aggiemail.usu.edu, (435) 512-6651

(a) Education

Utah State University, Logan, UT; **Computer Science; Ph.D., 2014**
Dissertation: Computational models of intracellular and intercellular processes in developmental biology.
Arak Azad University, Arak, Iran; **Software Engineering; M.S., 2010**
Thesis: Accurate quantitative trait loci mapping using evolutionary algorithms and partial least regression.
Shiraz University , Shiraz, Iran; **Software Engineering; B.S., 2007**
Project: Investigation of evolutionary algorithms, emphasizing mass extinction model.

(b) Professional Experience

2014–2014: **Visiting Scholar**, Institute for Systems Biology, Seattle, WA

2011–2014: **Research Assistant**, Utah State University, Logan, UT

2010–2011: **Teaching Assistant**, Utah State University, Logan, UT

2006–2010: **Software Developer & Project Manager**, Pardazeshgarane Pardis, Iran

(c) Technical Skills

1. 10 years of programming experience in C/C++, C#, Java and Python.
2. Mathematical modeling of biological systems.
3. Numerical and data analysis in Matlab, R, and CPLEX.
4. Machine learning (neural networks, random forests and regression models).
5. Distributed computing using Hadoop (MapReduce).
6. Natural language and image processing.

(d) Research Projects

1. Multiscale modeling of glioma development and related immune responses.
2. Agent-based modeling of ductal carcinoma in situ.
3. Cell fate prediction using gene expression data.
4. Multiscale modeling of spatial and temporal dynamics of yeast colony development.
5. CellDiff3D: a tool for modeling and visualization of cell type switching.（www.celldiff3d.org).
6. Regulation type identification in reconstructing gene regulatory networks.
7. Modeling of peroxisome biogenesis.

(e) Professional Services and Accomplishments

1. Received 1st rank in the First Student Olympiad of Innovative Applied Design in route optimization for railway engineering, 2009.
2. **Teaching assistant:** Outstanding teaching assistant of Computer Science Department, Spring 2011.

3. **ACM programming contest:** First place in Shiraz internal ACM programming contests and the eighth place in international ACM Asia programming contest (2006).
4. Awarded 8th place in Master entrance exam in Mechatronics field between 2417 applicants and accepted in Sharif University of Technology (the most prestigious university in Iran), 2009.

(f) Publications

(f).1 Peer-reviewed Journals

1. **A. Ghaffarizadeh**, N.S. Flann, and G. J. Podgorski, "Multistable Switches and their Role in Cellular Differentiation Networks," *BMC Bioinformatics* 2014, 15 (Suppl 7), S7.
2. **A. Ghaffarizadeh**, G. J. Podgorski, and N.S. Flann, "Modeling and Visualizing Cell Type Switching," *Computational and Mathematical Methods in Medicine*, vol. 2014, Article ID 293980, 10 pages, 2014. doi:10.1155/2014/293980.
3. **A. Ghaffarizadeh**, A. K. Esmailizadeh, M. Eftekhari, and N.S. Flann, "Quantitative Trait Loci Mapping Problem: An Extinction-Based Multi-Objective Evolutionary Algorithm Approach," *Algorithms*, vol. 6, no. 3, pp. 546-564, Sep. 2013.
4. **A. Ghaffarizadeh** and V. H. Allan, "History based Coalition Formation in Hedonic Context using Trust," *International Journal of Artificial Intelligence & Applications*, vol. 4, no. 4, pp. 1-8, Aug. 2013.
5. H. Barzegar Awal, P. Ahmadi and **A. Ghaffarizadeh**, "Thermal- Economic Multi-Objective Optimization of a Gas Turbine Power Plant with Preheater Using Evolutionary Algorithm," *International Journal of Energy Research (IJER)*, 2010, DOI: 10.1002/er.1696.

(f).2 Conferences

1. Q. BaniBaker, **A. Ghaffarizadeh**, S. Kwon, G. J. Podgorski, and N.S. Flann, "An Agent-based Model of Ductal Carcinoma in Situ (DCIS) and its Validation in a Tissue-engineered Model of DCIS," in *Institute of Biological Engineering 2013 Annual Conference*, March 2013.
2. **A. Ghaffarizadeh**, G. J. Podgorski, and N.S. Flann, "Epigenetic Landscape: a New Look," in *The Society for Mathematical Biology 2013 Annual Meeting and Conference*, June 2013.
3. **A. Ghaffarizadeh**, G. J. Podgorski, and N.S. Flann, "Analysis and visualization of cell state transitions during differentiation" in *5th International Conference on Bioinformatics and Computational Biology 2013,* BICoB 2013, PP. 231 - 236.
4. N.S. Flann, **A. Ghaffarizadeh**, S. Kang, S. Kahan, J. E. Mcdermott, and I. Shmulevich, "BioCellion: A High-performance Computing Framework for Multiscale Modeling and Simulation of Multicellular Biological Systems," in *The Society for Mathematical Biology 2012 Annual Meeting and Conference*, July 2012.
5. **A. Ghaffarizadeh**, K. Ahmadi, and N. S. Flann, "Sorting unsigned permutations by reversals using multi-objective evolutionary algorithms with variable size individuals," in *Evolutionary Computation (CEC), 2011 IEEE Congress on*. IEEE, Jun. 2011, pp. 292-295.
6. **A. Ghaffarizadeh**, K. Ahmadi, and M. Eftekhari, "Adding crossover to Extinction-Based evolutionary algorithms," in *2009 Second International Conference on Computer and Electrical Engineering*, ser. ICCEE '09. Washington, DC, USA: IEEE, 2009, pp. 43-48.
7. M. Meigounpoory, P. Ahmadi, and **A. Ghaffarizadeh**, "Optimization of Combined Cycle Power Plant Using Sequential Quadratic Programming," in *ASME 2008, Heat transfer summer conference*, Florida, USA, 2008.