THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF PHYSICS

THE EFFECT OF EXTERNAL SIGNALS AND COMBINATORIAL INTERVENTIONS IN A NETWORK MODEL OF THE EPITHELIAL-TO-MESENCHYMAL TRANSITION

DANYAS SARATHY
SPRING 2017

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Physics
with honors in Physics

Reviewed and approved* by the following:

Dr. Reka Albert
Distinguished Professor of Physics and Biology
Thesis Supervisor

Dr. Richard Robinett
Professor of Physics
Honors Adviser

* Signatures are on file in the Schreyer Honors College.

ABSTRACT

The epithelial-mesenchymal transition (EMT) is a well-studied cell fate that appears in both physiological processes (embryonic development and wound healing), and in pathologically detrimental processes like cancer. This thesis aimed to study the effect of combinatorial interventions in a network model of EMT postulated by Steinway et al. These researchers, having developed a dynamic model of the intracellular signaling network, studied the effect of internal node knockouts and interventions—the goal of this thesis, however, was to study the effect of external signals and source nodes, and to also combine constitutive activation of source nodes with internal node knockouts. It was found that every source node in EMT—HGF, PDGF, IGF1, EGF, FGR, goosecoid, and hypoxia—drove the transition when constitutively activated through an asynchronous update method using BooleanNet. These nodes varied in their rate of achieving full cell turnover, and hypoxia was the fastest state that induced EMT. When Hypoxia was combined with three internal node knockouts—SMAD, TGFβ, and MiR200—three different patterns emerged. The node knockouts either prevented the transition from occurring altogether, did not affect the transition, or slowed down the rate at which the system reached the mesenchymal attractor. These results could suggest that a discrete dynamic approach to studying biological processes could potentially elucidate experimental underpinnings

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would like to sincerely thank both Dr. Reka Albert and Dr. Richard Robinett for their kind support in encouraging this research. Without a push toward this direction, I would have never dreamt of the innovative research that this team does, and for that, I am very grateful. I am also thankful to my family, for supporting my academic career (and not just financially!) by encouraging me in every step of the way.

Chapter 1

Biological Basis of the Epithelial-to-Mesenchymal Transition (EMT)

Before engaging in the discussion of how the EMT can be modeled, it is first important to describe the biological underpinnings of this phenomenon. This chapter presents a precursory biological overview of the EMT cell fate change, and uses concepts in biochemistry and cell biology to describe the effects of the transformation in cancer signaling. Once a biological basis has been established, a simplification by virtue of graph theory and discrete dynamic modeling will allow a predictive network model (that is, one that can be experimentally manipulated) to flourish.

The Cell Biology Picture of the Epithelial-to-Mesenchymal Transition

Epithelial cells form the basis of the lining of organs and vessels—for example, blood vessels are encapsulated by multiple layers of epithelium. By lining vessels and organs, these cells are therefore "bound— "they reside in a specific location and perform a specific function. The functional part of the organ (the *parenchyma*) is a type of epithelium. Just as the different tissues in the body serve a multitude of functions, the epithelia must necessarily have a variety of structures, and thus these cells can be highly *specialized*. This key point of specialization is what makes the EMT conducive to cancer, as will be discussed later.

Mesenchymal cells, on the other hand, are *not* specialized cells—in broad terms, mesenchymal cells have undefined cell lineage: they are multipotent. While epithelial cells have specific structures that accommodate specific functions (e. g. their cell fate has been decided), mesenchymal cells can continue to differentiate into a number of different types of epithelium (myocytes, chondrocytes, etc.). Because these cells are not constrained to a particular region of the body (unlike epithelia), these cells are then "motile," and can enter the bloodstream to translocate in the systemic circulation [1]. Thus, as will be discussed later, metastasis (in cancer) is inevitable once these cells lose adhesion—cancer spreads rapidly to other locales as cancer cells are free to roam. Self-programmed destruction of mesenchymal cells—apoptosis—is made even harder by an innate resistance of mesenchymal cells to the standard cell cycle regulators: p53 is suppressed when the cell becomes mesenchymal [1].

Fundamentally, the mesenchymal cells do *not* have to just be cancerous—the EMT is useful for other biological functions as well. For example, in fetal growth, the ability for a cell to have undefined lineage is useful to signal development and differentiation of new cells—a function that epithelial cells cannot do. Therefore, there are three major classifications of the EMT: differentiation of new cells for development, clotting and menstruation, and as mentioned previously, cancer mobilization. The EMT process is not irreversible: just as epithelial cells can undergo a transition to mesenchymal cells, the reverse process can also occur, although this is rarer [2]. As expected, this process is initiated by driving epithelial inducer expression.

The key to this transition is highlighted by the mobility of the mesenchymal cells, as noted before. Motility is prevented by the class of cadherin proteins—these proteins allow cells to "stick" to each other, and in the case of epithelial cells, E-cadherin is the key protein that prevents the cell from becoming motile. Thus, EMT is characterized by the loss of E-cadherin expression, and consequently the upregulation of other molecules that would down-regulate E-cadherin. For example, Zinc finger protein SNAI1, or "snail" is a member of a set of seven transcription factors that repress E-cadherin expression [3]. Thus, knockout of SNAI1 would negatively influence EMT. There are many pathways that influence the activity of the seven transcription factors , and sometimes they even feedback regulate each other (i.e. one pathway leads to the production of a molecule that serves as the signal of the other). Thus a simplification is needed to integrate and visualize how all these different pathways contribute to the overarching end goal: loss of E-cadherin to force EMT.

Simplification: A Network Approach to Studying EMT

As the biological basis of EMT is complicated, rooted in biochemical underpinnings and complex pathways, a network interpretation can incorporate the essence of this complex system by representing the proteins, mRNAs and miRNAs that participate in this process as nodes of a network and representing the interactions and regulatory effects between them as edges of a network. For example, the central dogma of biochemistry is represented in the classical network model shown below.

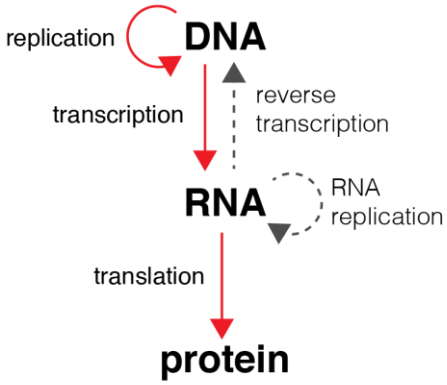

Figure 1: The Central Dogma, represented as a network model [4]

Of course, the transition between DNA to mRNA is extremely complicated, requiring no less than 20 helper proteins, and intricate pathways and repair mechanisms—however in this model, a causal relationship between DNA and mRNA is expressed as a directed edge, and conclusions can be made from this diagram alone, without the need for complex chemistry. It is evident, for example, that proteins cannot be reconstructed into RNA, and thus the flow of information is exclusively from DNA to protein. While the above example was simple—only three nodes in the network—EMT is characterized by much more complicated pathways, culminating in the regulation of the expression of the target protein E-cadherin. Since epithelial cells require E-cadherin to remain burrowed, if E-cadherin expression were to be suppressed, these cells would lose their adhesion capability and become motile—and *mesenchymal*.

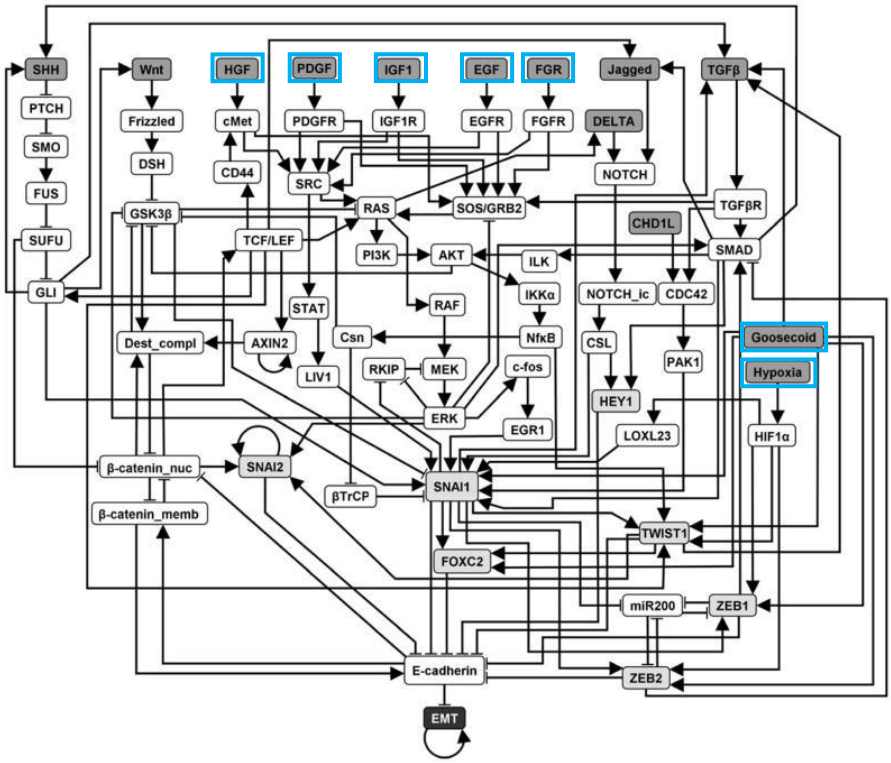

Figure 2: EMT Network Model. Source nodes (those that do not receive feedback from other nodes) are highlighted in blue. Adapted from Steinway et al. [13]

While several different signals have been studied in the induction of EMT, the model and analysis of Steinway et al. focused on a key compound that drives the transition forward if over-expressed: transforming growth factor beta (TGFβ) [5]. Indeed, they have found that by forcing expression of TGFβ, EMT is driven forward in a manner consistent with experimental observations, One follow-up question that arises in analyzing this network model is: how do the other external signals (shown in blue in Figure 2) affect EMT? TGFβ is not a *source* node, as it can be produced by the cell. So the question addressed in the following discussion will be focused on those signals—how do the HGF, PDGF, IGF1, EGF, FGR, goosecoid, and hypoxia nodes affect EMT?

Using the Biological Underpinnings of Source Nodes to Theorize the Impact on EMT

Before diving into the network analysis approach to resolve how source nodes drive EMT, a preliminary biological overview of these nodes can be used to predict whether or not their constitutive activation can force the loss of E-cadherin expression. The five source nodes in the model all correspond to protein expression. With the exception of FGR, goosecoid and hypoxia each of the source nodes are "growth factors," and as the name implies, these proteins promote the growth and differentiation of cells [6]. While growth factors are essential for early development, and the proliferation of required biomolecules (e. g. hematopoiesis which creates blood, when in an oxygen deprived environment), over-expression of these proteins is dangerous. If the proteins are not regulated, growth of the cell cannot be controlled, and cancer can develop, as suggested by the network model.

HGF, or hepatocyte growth factor, is released by already "turned" mesenchymal cells. A hepatocyte is a cell of the liver, and these cells are marked by their extraordinary ability to regenerate themselves (a function that most cells cannot perform—instead of attempting to repair themselves, ordinary damaged cells will undergo apoptosis). Thus, hepatocyte growth factor plays a key role in moderating the ability of the organ to repair itself--naturally, this involves mobilizing cells to the target organ, and since epithelial cells are frozen in place, HGF acts on epithelial cells to release them from. Therefore, based on the function of HGF, it is hypothesized that overexpression should always drive EMT, because a downstream function of the protein is to downregulate E-cadherin [7].

PDGF, platelet derived growth factor, performs a different role than HGF. While HGF is secreted by mesenchymal cells in a "positive feedback" like method to "turn" epithelial cells, the function of PDGF is in cell cycle regulation. PDGF acts upon cells to drive mitotic behavior via mitogenesis. Overexpression of mitotic behavior, as expected, is coincidental with cancer—therefore, regulation of PDGF is essential for controlling unregulated cell division. Because PDGF is a potent mitogen, constitutive activation of this node may induce cancer, however it may not necessarily trigger the EMT due to its avoidance in regulation of E-cadherin. Since PDGF influences cells that are already mesenchymal, it is hypothesized that this node may force epithelial cells to become mesenchymal [8].

IGF1, or insulin-like growth factor 1, unsurprisingly acts similarly to insulin. Unlike the previous two growth factors delineated above, IGF1 is a special type of protein—a hormone, which acts a signaling molecule in the endocrine system which can transverse the entire circulation to affect distant cells. This unique feature allows IGF1 to communicate with most cells in the body, and prompt a surge of systemic growth when needed. While IGF1 has been shown to interact in certain cancer signaling pathways, its role as a hormone do not overlap much with EMT like transformations—namely, because IGF1 is far removed from control of E-cadherin, it is hypothesized that it does not likely influence EMT [9].

EGF, or epidermal growth factor, is an important growth factor that promotes the proliferation and differentiation of many cell types. By binding to its receptor, EGFR, it initiates a secondary messenger cascade through a tyrosine kinase pathway which undergoes numerous

biochemical reactions. The end result is more available sources of nutrition for cells to undergo mitogenic behaviors—ultimately driving growth [10].

FGR, or Gardner-Rasheed feline sarcoma viral oncogene homolog, acts by modifying protein-protein interactions to down-regulate CAMs (cell adhesion molecules). This process is initiated by the beta-2 integrin pathway, and can be overexpressed through the Epstein-Barr virus (EBV). More commonly known as Herpes, EBV unsurprisingly infects epithelial cells to mobilize them for wart formation [11].

Goosecoid protein, a homeobox (HOX) protein, is a key molecule involved in neural crest differentiation. Goosecoid directly promotes the mobilization of cells in gastrulation (in which three germ layers are derived from one, a process that occurs extremely early during development) and organogenesis (the creation of organ tissues, a process that occurs very late during development). Thus, goosecoid is one of the main contributors to mobilizing cells in the body, and is strongly expected to drive EMT [12].

Finally, hypoxia is not a molecule but rather a *state* in which a cell can exist in. Hypoxia is characterized by a deficiency in the amount of oxygen a cell receives—because oxygen is required for aerobic respiration, processes like the tricarboxylic acid (TCA) or "Krebs" cycle are halted. These processes generate the most amount of energy per unit "food" for the cell, and thus disruption of these processes can cause potential damage. The cell responds to hypoxia by upregulating Hypoxia-inducible factor 1-alpha (HIF1A, immediately downstream of hypoxia), wherein this "master" protein will allow the cell to adapt to the environmental state and prevent

damage. Continued over-expression of HIF1A has been shown to promote regeneration of tissue, implicating the mobilization of cells—thus, it is predicted that hypoxia will force EMT.

Chapter 2

Application of Discrete Dynamic Modeling to the EMT Network

Building on from the network presented in the previous chapter, the rules and regulations of graph theory can be applied to the EMT network. In this chapter, the EMT network will be closely inspected, and an elaboration of how the network was manipulated through discrete dynamic modeling will be discussed.

Discrete Dynamic Modeling

The network in Fig. 2, created by Steinway et al., is composed of a multitude of "nodes" (e.g.TGFβ, Wnt, SHH) and "edges," (connections between nodes) as these are the component of any network [5]. In the context of the network, each node represents some element of biology—for example, TGFβ represents the *expression* of the growth factor. This is a binary set of states—either the molecule can be expressed in the DNA of the cell, and the growth factor can be coded for in the ribosomes, or it can be absent, and no molecule will be created. Every node in the network follows the same logic—it is either on, or off.

Once the network has been created, dynamics can follow suit. In a discrete dynamic model each node is characterized by a regulatory function (or rule) that describes how the future state of the node is determined by the current state of its regulators (i.e. the nodes that have

edges pointed toward it). Each node's new state is determined following its regulatory functions and the states of the whole network change accordingly. Steinway et al. chose to update the nodes' state in a randomly selected order as there was insufficient information on the timescales of each interaction [5]. One round of node state update is one "time step," and the process can be repeated again—the rules are now employed on the *new* network state, and the states are changed again. In order for this process to start, however, an initial state of the network is needed first. While the network's edges represent the rules of the system, the network itself must first be initialized with a set of "initial conditions," which for biological networks represents the entire state of the cell. In the EMT network, there are 69 total nodes, and each node can be on or off. This corresponds to a net total of 2^{69} initial states—this astronomically high number of states makes analysis of each one an impossibility. Of course, only a few of these states represent the state of a "normal" (e.g. epithelial) cell. Steinway et al found that a state that shows a good agreement with what is known about the protein expression in epithelial cells is an attractor (stable state that does not change) in the absence of external drivers (constitutive activation of a source node). They defined this state as the "epithelial" state of the cell. This was achieved by simplifying the network down to a 19-node network that had the same traits as the larger network [13].

Before diving into the network analysis of the complicated EMT, a precursory glance must be given to simpler networks, to highlight what it is we are looking for in the larger model. Shown below are common network patterns (known as *motifs)* that can exhibit interesting behavior, even with few nodes.

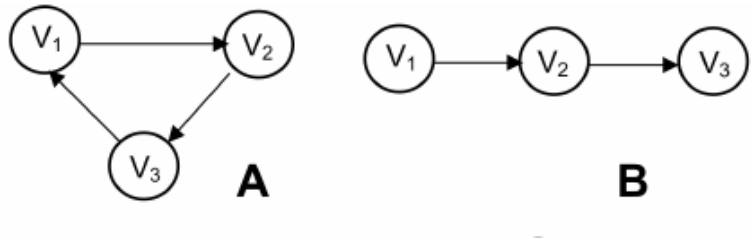

Figure 3: Two common network motifs. Source: (Pavlopoulos et al. 2011) [19]

Consider the network in Fig. 3A with the regulatory functions $V_1^*=V_3$, $V_2^*=V_1$ and $V_3^*=V_2$, where the * indicates that we refer to the future state of the marked node. Starting from the initial state of $V_1 = 1$, $V_2 = 0$, $V_3 = 0$, we observe a unique pattern known as cyclic behavior. In the first time step, when the rules of the network are applied simultaneously to all nodes, the network moves from the initial state to the following state: $V_1 = 0$, $V_2 = 1$, $V_3 = 0$. In the second time step, the rules are reapplied, and the following state is established: $V_1 = 0$, $V_2 = 0$, $V_3 = 1$. Finally, when the rules are applied again, the state re-enters the initial condition that it started at: $V_1 = 1$, $V_2 = 0$, $V_3 = 0$. This process continues forever, and so the network (with this initial condition) is characterized as periodic with a period of 3 time steps. 3B, on the other hand, is acyclic. When the rules of the network ($V_1^*=V_1$, $V_2^* = V_1$, $V_3^* = V_2$) are applied starting form the initial state $V_1 = 1$, $V_2 = 0$, $V_3 = 0$, the network will always end in the "attractor" state $V_1 = 1$, $V_2 = 1$, $V_3 = 1$.

A key aspect of the dynamics behind those two networks, however, is the update method. In the previous two examples, a "synchronous" update was applied—the network's rules were applied simultaneously to all nodes. This reflects the implicit assumption that all the synthesis and decay processes in the network have the same duration. While is this certainly an acceptable

way to model a network, it unfortunately fails under the purview of biological systems: in a biological framework, not every process has the same duration. Instead, *stochastic* behavior is seen experimentally—for example, a gene's size may determine the time it takes to express a protein, and so not all genes would express their respective protein at the same time. To account for this, a stochastic model is needed, and thus "asynchronous" update more aptly mimics the biology of the cell [13]. In the asynchronous update, random nodes are selected at each time-step to be updated, as opposed to every node at once.

While this is a rather rudimentary example, the result is characteristic of what we look for in the EMT network—we must first find an initial condition that is stable (an attractor), and consistent with the epithelial state of the cell. Once this state has been established, we can manipulate the network by driving certain signals (and in this case, we choose to drive source nodes by constitutive activation)—simply put, we force these nodes to be in the "on" state, no matter what. Then, we analyze if the network has been "pushed out" of the stable attractor it was in initially, and to see if eventually the state of the network will trigger EMT by loss of E-cadherin. Given the rules of the network in Fig.2, once E-cadherin expression has been lost, EMT will be turned on, and since EMT forces itself on, the cell fate will forever be stuck in this state.

Methodology: Discrete Dynamic Modeling Using BooleanNet

To simulate and manipulate the EMT network, BooleanNet was used as the principle software. BooleanNet takes as an input the rules of the network in a text file (see Appendix A), and updates the states of individual nodes by applying the rules specified. By instantiating the

model in the text file (i.e., giving an initial state and supplying the rules between nodes), Python commands can be issued to analyze the state of the network after different time-steps. Various helper functions allow for averages, and plots to be generated. To drive the network, the "on […]" and "off […]" functions were employed to constitutively activate, and knockout, nodes respectively.

The initial state chosen for this experiment was the "Normal-2" state from Steinway et al. [13]. This initial state was chosen as it represents the prototypical epithelial attractor—that is, when the nodes of the network are in this state, they stay fixed in such a way that is experimentally consistent with the biological interpretation (molecular composition, expression of proteins, etc.) of an epithelial cell. The states for the epithelial attractor were determined by finding the pathways that induced EMT experimentally, and setting the nodes within those pathways to be off. This initial state is outlined in Appendix A.

Chapter 3

E-Cadherin and EMT Expression in the Presence of Single Node Knockout Interventions

In using BooleanNet, the percentage of simulations which achieved the states of EMT and retained the state of E-Cadherin were analyzed. 50 simulations were ran for each source node over-expression, and each simulation consisted of 10 steps using an asynchronous update method. The state of the target nodes (E-Cadherin, EMT) were recorded for each simulation, and averaged after 50 simulations had been ran. This appeared to be the point of diminishing returns on accuracy—use of 100 simulations nearly doubled the computation time, but was accurate to ±1% of the 50-simulation method.

A sample data run is shown below—when EGF is overexpressed, these are the average number of simulations in which EMT had been achieved for each time-step.

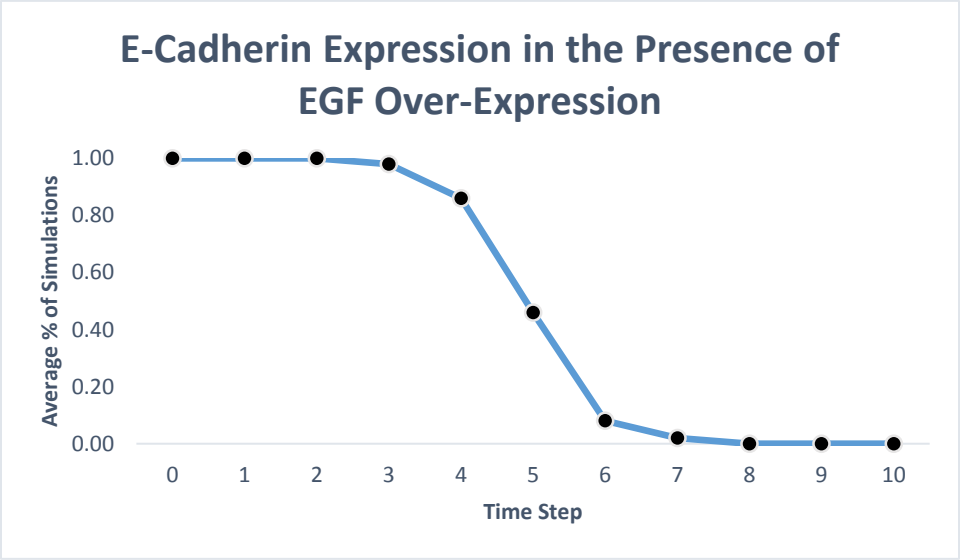

Figure 4: Average E-Cadherin expression (50 simulations) in the presence of EGF over-expression.

As seen in Fig. 4, the network remains stable until time-step 4 (as the signal from EGF has not yet propagated through down to the target node), in which some simulations lose E-Cadherin

expression. Due to asynchronous update, the transition is not a step function—the randomness of the application of the rules to the network allow some simulations to retain the state of E-Cadherin while others do not. Eventually, however, as time progresses, the signal gradually reaches E-Cadherin in every simulation, and they all lose expression of the protein.

To demonstrate the application of synchronous update (i.e. not biologically consistent), the following figure is presented of the same experiment shown previously (expression of E-Cadherin in the presence of EGF over-expression):

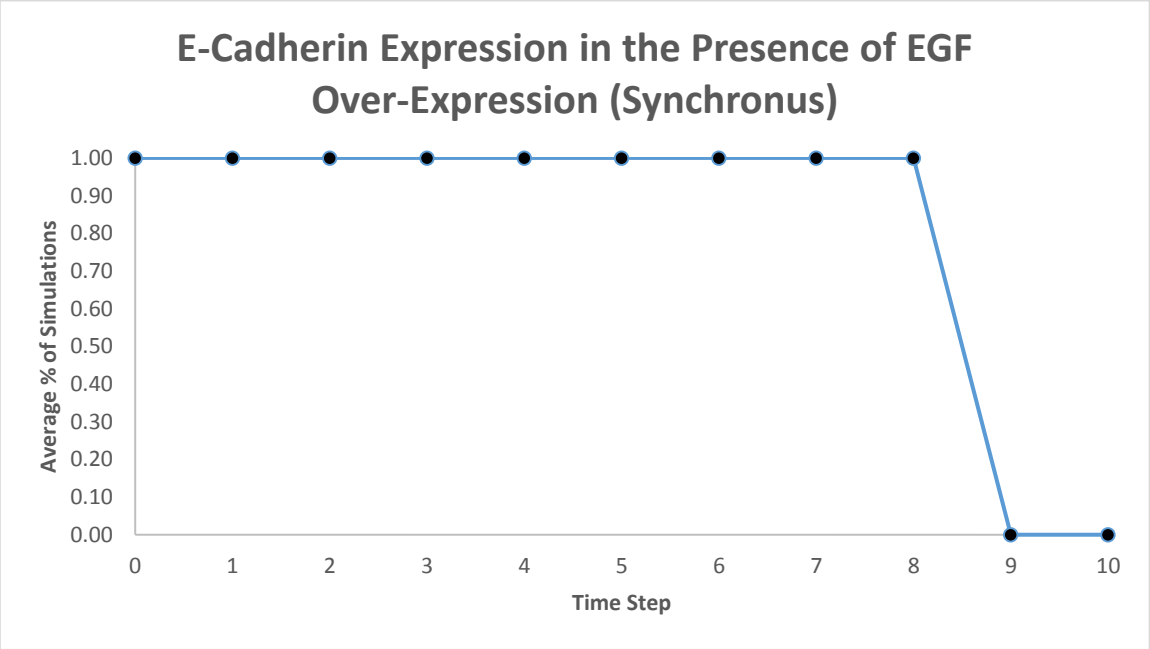

Figure 5: Synchronous update model of the impact of EGF over-expression to E-Cadherin

As shown, if all the updates were to happen simultaneously, the signal would reach E-Cadherin precisely at time-step 9. Again, because this is not biologically relevant, only the asynchronous update method was utilized and thus is implied for the rest of the data.

When compared to EMT expression, the E-Cadherin has an inverse relationship, as expected. Loss of E-Cadherin *implies* EMT, and so as simulations lose E-Cadherin expression, Fig. 6 demonstrates that EMT occurs, as shown below:

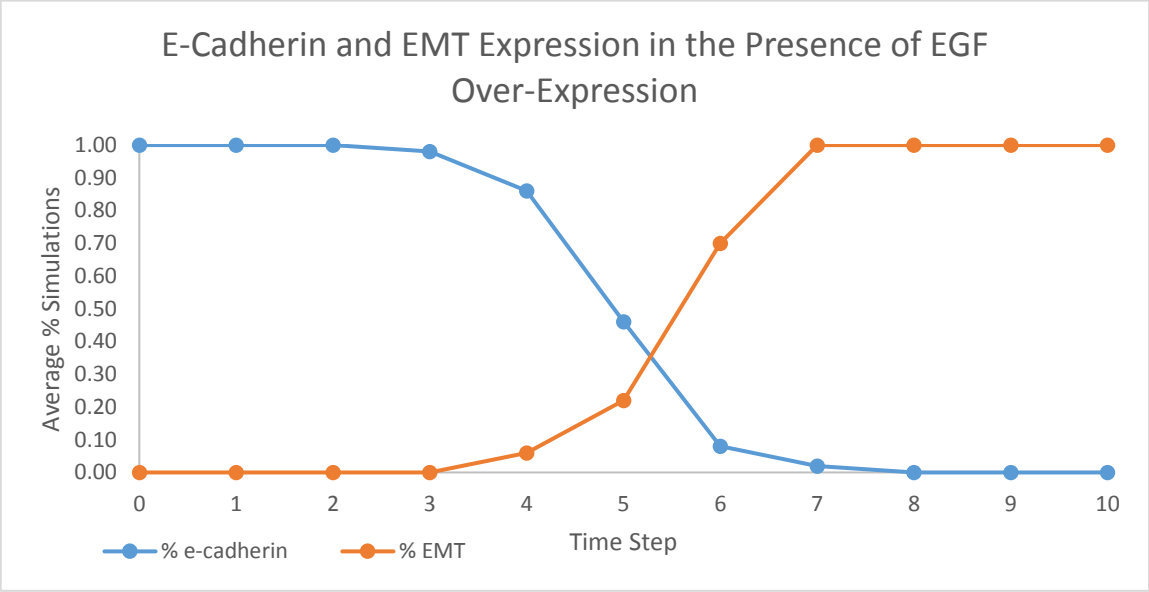

Figure 6: EMT and E-Cadherin expression in the presence of EGF constitutive expression

In analyzing Fig. 6, it is evident that the percentages of E-Cadherin and EMT do not add to 100%. This is because of the discrete time-steps following asynchronous: there is a lag in the propagation of E-Cadherin to EMT, and thus E-Cadherin falls behind. This is merely an artifact of the update, and does not represent anything physical. The data for all source node over-expression is tabulated below for both E-Cadherin and EMT expression.

Table 1: Average E-Cadherin Expression over multiple source node over-expressions

% E-Cadherin Expression	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
egf	1.00	1.00	1.00	0.98	0.86	0.46	0.08	0.02	0.00	0.00	0.00
fgf	1.00	1.00	1.00	1.00	0.88	0.42	0.14	0.02	0.00	0.00	0.00
hgf	1.00	1.00	1.00	1.00	0.94	0.54	0.08	0.02	0.00	0.00	0.00
goosecoid	1.00	1.00	0.90	0.44	0.12	0.00	0.00	0.00	0.00	0.00	0.00
hypoxia	1.00	1.00	1.00	0.90	0.50	0.10	0.00	0.00	0.00	0.00	0.00
igf1	1.00	1.00	1.00	0.96	0.88	0.58	0.10	0.02	0.00	0.00	0.00
pdgf	1.00	1.00	1.00	1.00	0.88	0.48	0.16	0.02	0.00	0.00	0.00
chd1l	1.00	1.00	1.00	0.90	0.68	0.36	0.12	0.02	0.00	0.00	0.00

Table 2: Average EMT Expression over multiple source node over-expressions

% expression EMT	**0**	**1**	**2**	**3**	**4**	**5**	**6**	**7**	**8**	**9**	**10**
egf	0.00	0.00	0.00	0.00	0.06	0.22	0.70	1.00	1.00	1.00	1.00
fgf	0.00	0.00	0.00	0.00	0.00	0.34	0.64	0.94	1.00	1.00	1.00
hgf	0.00	0.00	0.00	0.00	0.06	0.30	0.70	0.98	1.00	1.00	1.00
goosecoid	0.00	0.00	0.00	0.26	0.72	0.92	1.00	1.00	1.00	1.00	1.00
hypoxia	0.00	0.00	0.00	0.02	0.22	0.48	0.86	1.00	1.00	1.00	1.00
igf1	0.00	0.00	0.00	0.00	0.10	0.46	0.82	1.00	1.00	1.00	1.00
pdgf	0.00	0.00	0.00	0.00	0.02	0.30	0.74	0.98	1.00	1.00	1.00
chd1l	0.00	0.00	0.00	0.00	0.12	0.40	0.82	0.94	1.00	1.00	1.00

When the EMT expression data is visualized for all eight source nodes on a single graph, an interesting node is found: goosecoid in Fig. 7 appears to drive the network to the mesenchymal state the fastest out of all the source nodes. While there is one distinct cluster, to which most source nodes fall within a few time steps of, goosecoid instead drives the network to the mesenchymal state well before the other cluster.

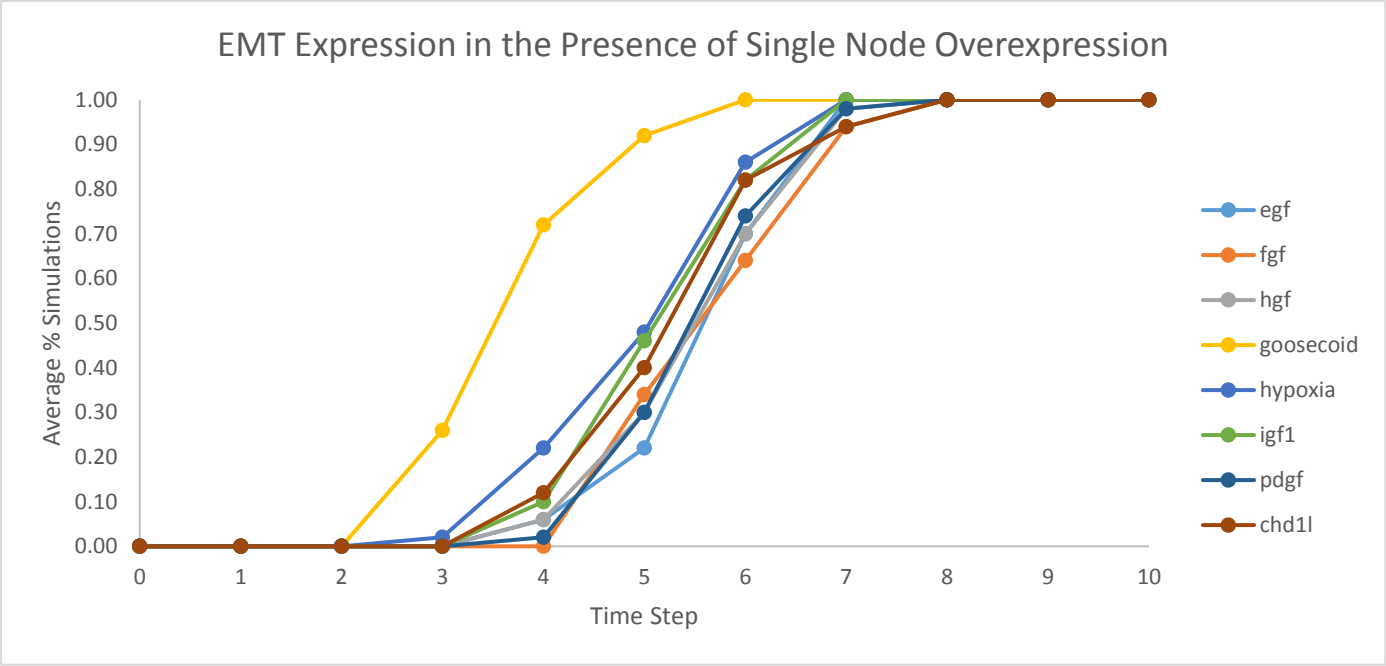

Figure 7: EMT expression when eight source nodes are constitutively activated.

Fig. 7 reveals that most source nodes in the EMT network model, when over-expressed, drive EMT at a similar rate. With the exception of the goosecoid protein, over-expression of every other source node follows a sigmoidal function in which about half of all simulations achieve the mesenchymal state between five and six time steps, and all simulations achieve the mesenchymal state by the eighth time-step. Within the cluster of source nodes, there appears to be a slight degree of variation at the rate to which the system achieves the mesenchymal state: for example, at the fourth, fifth, and sixth time-steps, the variation in the percent of simulations

achieving EMT varies over a range of 30%. The data suggest that egf, pdgf, fgf, and hgf drive the network at the slowest rate (these data points are shifted toward the right), while goosecoid and hypoxia drive the network at the fastest rate (these data points are shifted toward the left). This is apparently a consequence of the degree to which these nodes are connected to E-Cadherin: the "embedded" source nodes—hypoxia and goosecoid—found deep within the network are fewer nodes away from E-cadherin than their counter parts. Goosecoid, for example, only needs to propagate its signal through two nodes (Fig. 2, FOXC2, E-Cadherin) to change the network to the mesenchymal attractor. On the other hand, source nodes like egf and hgf are found at the very "start" of the network, and thus have a longer journey to travel to reach E-cadherin, and ultimately affect EMT.

Chapter 4

E-Cadherin Expression in the Presence of Multifactorial Node Knockout Interventions

While every source node manipulation drove the entire network to the mesenchymal attractor, another interesting question posits itself: what occurs when embedded nodes (i.e., non-source nodes) are combined alongside source nodes? That is, if a certain embedded node alone cannot drive EMT through constitutive activation, would it prevent or hinder a source node from driving EMT? Also, would a knockout of embedded nodes prevent source nodes from driving EMT? To answer these questions, this chapter discusses a few representative double node knockout/activation combinations to assess the types of responses the network can have.

By combining the overexpression of a source node that is known to drive EMT—something extensively discussed in the last chapter—alongside an embedded node knockout that *does not* drive EMT, three interesting patterns appear: the network is prevented from achieving the mesenchymal attractor due to the knockout, the network ignores the knockout and achieves the mesenchymal attractor at the same rate, or the network achieves the mesenchymal attractor at a slower rate (compared to if the knockout was absent). To demonstrate these three patterns, the hypoxia source node was chosen as the representative node to overexpress (as it is known to drive EMT from Fig. 7), and the three following embedded nodes were knocked out: SMAD, TGFβ, and MiR200.

To begin, we will examine the case where EMT is prevented from being driven: if the node TGFβ is knocked out, the epithelial attractor stays stagnant, and if the node is overexpressed, EMT is driven as demonstrated below:

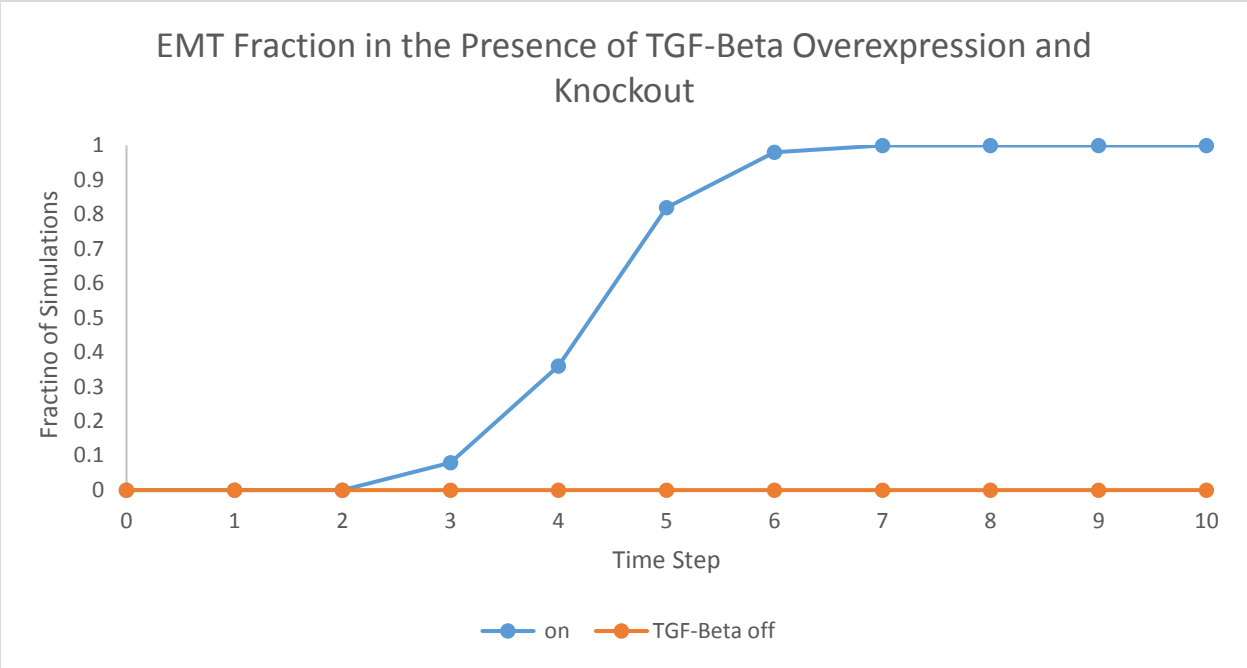

Figure 8: EMT when TGFß is constitutively activated (blue), or knocked out (orange)

Thus, the reasonable conclusion is that TGFβ is capable of driving EMT, however when the node is knocked out, it is unclear whether it hinders EMT, or leaves it unaffected. Thus, is it necessary to pair this knockout with a known driver of EMT, and to evaluate the combinatorial result. When TGFβ is knocked out, and Hypoxia is overexpressed, the result is shown below:

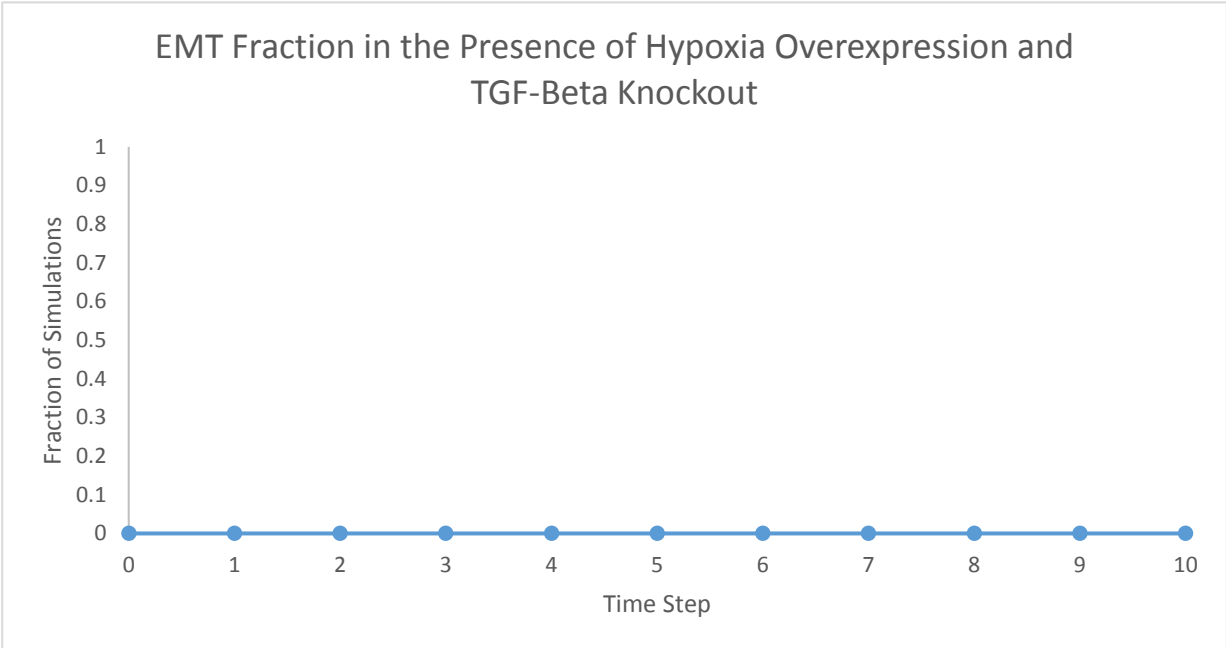

Figure 9: EMT expression in a TGFß knockout paired with Hypoxia overexpression.

In this instance, EMT is never driven under this combinatorial intervention. It is clear, then, that TGFβ is required for EMT—if the node is knocked out, it is impossible for the transition to be driven.

Next, we examine the case where the knockout has no effect—EMT is driven regardless of whether the node is knocked out. In this case, we examine the node MiR200; just like TGFβ, if the node is knocked out, EMT is not driven, and if the node is overexpressed, EMT is driven. When the knockout of MiR200 is paired with Hypoxia, the resultant EMT expression is as follows:

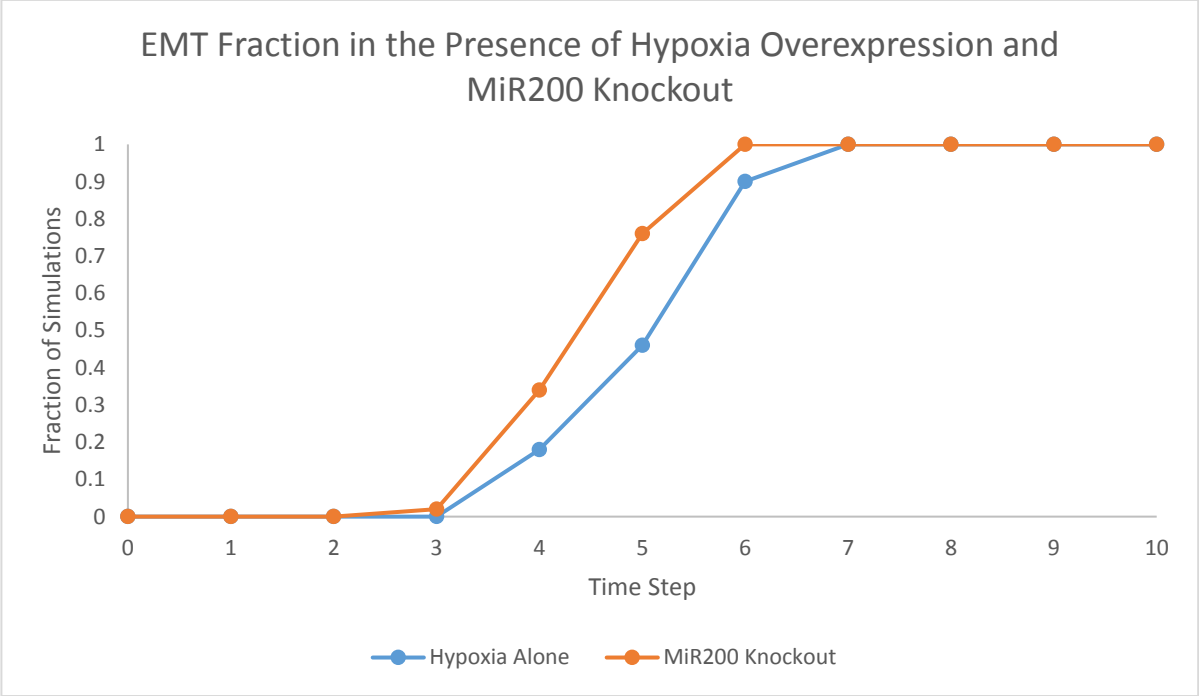

Figure 10: EMT expression in a MiR200 knockout paired with Hypoxia overexpression. Normal (single node overexpression) shown in Blue, combinatorial intervention shown in Orange.

about 7 time steps for both cases. In this scenario, EMT is driven at the same rate (or perhaps even faster with these simulations) regardless of the node knockout. Therefore, MiR200 knockout had no effect on driving EMT relative to source node overexpression.

Finally, the last case to examine is where the route to EMT is slowed down by a node knockout. Like the two other node knockouts, SMAD when turned off does not drive EMT, but when turned on does. When a SMAD knockout is paired with Hypoxia overexpression, the result is as follows:

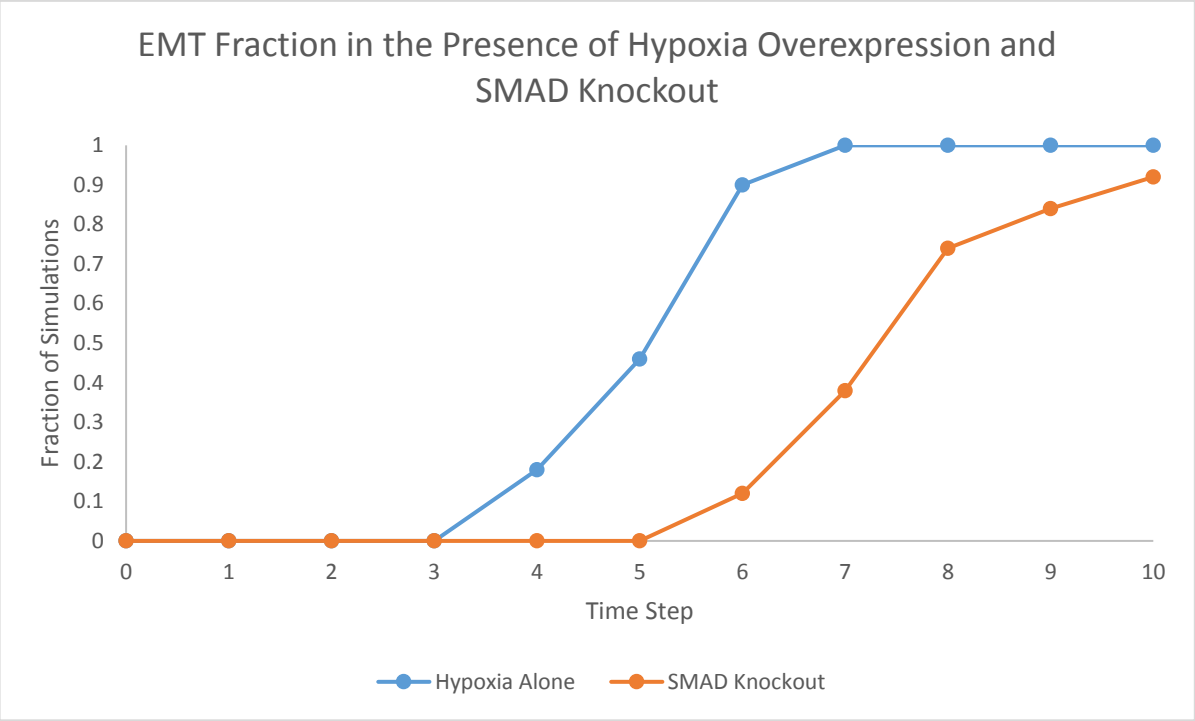

Figure 11: EMT expression in a SMAD knockout paired with Hypoxia overexpression. Normal (single node overexpression) shown in Blue, combinatorial intervention shown in Orange.

It is readily seen from the above graph that SMAD knockout impairs the ability of the Hypoxia overexpression to induce EMT. Normally, a Hypoxia overexpression drives EMT within 7 time steps, however when SMAD is knocked out, only 90% of simulations reach the mesenchymal attractor. Thus, this example of a multifactorial node knockout satisfies the last category—a situation in which a knockout slows down the rate at which a known driver of EMT reaches the attractor.

In analyzing these results, it is important to recognize whether or not these categories stem from an artifact of the model—that is, the discrete dynamic analysis and framework—or whether they represent something biological. Thus, the next chapter will be focused on analyzing the results presented by these multifactorial node knockout combinations, and address the significance behind the results.

Chapter 5

Biological Analysis of Knockout Interventions

The results shown in Fig. 7 posit that different nodes, when overexpressed, take variable amounts of time to eventually drive EMT. While this is appreciable through the network—the interconnectedness of the nodes—this implies a further biological context. Moreover, when node knockouts are combined with constitutive activation, three distinct "patterns" of behavior are seen in the network; this is also biologically relevant. This chapter will address the implications of the results presented in the last two experimental chapters in a biological context, and assess how the network model can be applied to a real cell.

Analysis of the Results of Single Node Constitutive Activation

As highlighted in the first chapter, constitutive activation of a node is part of larger biological framework where a gene is overexpressed, causing a protein (which, in most of these source nodes end up being growth factors) to be unregulated and constantly produced in the ribosomes. Not all proteins are the same, however—different growth factors necessarily affect the cell in different ways. There is, however, one key similarity: no matter which source node was selected, every single one drove EMT. This observation allows for a simple conclusion (which is corroborated by the literature): if a cell is in a cancerous state (that is, if there is unregulated protein production), then the mesenchymal phenotype is almost guaranteed to be reached. And if the mesenchymal phenotype is reached, this cancerous cell can become motile

and transverse the systemic circulation. Thus, this provides an understanding of how metastasis may occur—cancer may spread throughout the body due to an epithelial cell becoming motile (mesenchymal); this necessarily occurs if protein production is unregulated, as demonstrated in Fig. 7. While this conclusion is not unique [1], the discrete dynamic approach provides insight into looking at metastasis in a different way: the "signal" from the mutated DNA which overexpresses proteins will always propagate downstream to uproot the cell.

Another key observation made from Fig. 7 is the fact that different source nodes drive EMT at variable rates. This, of course, must do with how far away the source node is from E-Cadherin, but is there a biological way of looking at this conclusion? Let us look exclusively at the goosecoid protein, which was capable of driving EMT the fastest. Goosecoid is almost exclusively a protein expressed for the intent of neural crest development, a process in embryogenesis where it is necessary to "split" different germ layers to specialize cells [12]. Thus, to accomplish this, goosecoid plays a role in inducing cells to move about and form the neural crest. This highly specialized protein basically has only one role—to cause an epithelial cell to move, and locate to another region of the body where it can further be specialized—and because this role aligns so closely with EMT, it is no wonder that when the cell "sees" goosecoid being released, it wants to immediately become mesenchymal. In contrast, let us examine chd1l, which took the longest amount of time to induce EMT. CHD1L is relatively less studied compared to the other proteins, but it has been suggested that this is an enzyme that assists in DNA repair. Because the process of DNA repair, biologically speaking, is so far removed from EMT, it is not a surprise that overexpression of this enzyme did not immediately cause a mesenchymal phenotype. The large conclusion that can be made is that the function of the protein dictates the rate at which EMT will be achieved—should the protein be implicated in a

process which necessitates motility and pluripotency (such as development), then EMT is sure to occur quickly, as the cell recognizes the specific function of that protein. Should the protein, on the other hand, serve a more general function (like DNA repair), EMT will take a longer time as the "signal" from the DNA is not specific enough to induce cell fate change immediately.

Analysis of the Results of Multifactorial Interventions

The results presented in chapter 4 led to the conclusion that there were three different outcomes in which a network node was knocked out, and a source node was over expressed: the network drove EMT at the same (or faster) rate, the network drove EMT at a slower rate, or the network was prevented from driving EMT altogether.

In the most dramatic case, EMT is prevented from being driven if TGFβ is knocked out. This implies that TGFβ is necessary for EMT to occur, and thus the function of TGFβ must directly influence EMT. This is indeed the case as Steinway et al. demonstrated, without the expression of TGFβ, there is no EMT. TGFβ initiates a downstream signaling cascade which mobilizes the necessary proteins into place to allow the cell to become mesenchymal—without this "master protein," there can be no cell fate change. The network highlights this discovery by a motif in Fig. 2—certain nodes linked to TGFβ are necessarily required for EMT to function, because of the connectedness of TGFβ to E-Cadherin

In a similar vein, when SMAD is knocked out, the duration to a full mesenchymal phenotype is increased. This is because SMAD is intricately connected to TGFβ. While TGFβ was *required* for EMT to occur, SMAD sits just downstream of TGFβ, and is an associated protein that carries the signal of TGFβ by acting as a transcription factor of TGFβ. When SMAD

is knocked out, TGFβ's "propagation" is more difficult, as the SMAD channel has been severed—thus, the signal is sent through the other downstream proteins, albeit at a slower rate. The end result is that EMT is eventually driven, but slower, because the cell had fewer transcription factors to carry the TGFβ signal.

Finally, in the simplest case, it was shown that MiR200 apparently did not negatively influence hypoxia induced EMT (and may have even driven it faster). MiR200, a micro RNA, has been implicated in the EMT pathway, however it behaves as an inhibitor and thus expression of the micro RNA should inhibit EMT from occurring. Thus, it would make sense that when knocked out, EMT would be driven at a slightly faster rate.

Appendix A

EMT Rules and Initial States in BooleanNet

```
#INITIAL STATE
akt = False
axin2 = False
bcatenin_memb = True
bcatenin_nuc = False
btrcp = True
cd44 = False
cdc42 = False
cfos = False
chd1l = False
cmet = False
csl = False
csn = False
delta = False
dest_compl = True
dsh = False
ecadherin = True
egf = False
egfr = False
egr1 = False
emt = False
erk = False
fgf = False
fgfr = False
foxc2 = False
frizzled = False
fus = False
gli = False
goosecoid = False
gsk3b = True
hey1 = False
hgf = False
hif1a = False
hypoxia = False
igf1 = False
igf1r = False
ikka = False
ilk = False
jagged = False
liv1 = False
loxl23 = False
mek = False
mir200 = True
nfkb = False
notch = False
notch_ic = False
pak1 = False
patched = True
pdgf = False
pdgfr = False
pi3k = False
raf = False
ras = False
rkip = True
shh = False
smad = False
smo = False
snai1 = False
snai2 = False
sos_grb2 = False
src = False
stat = False
sufu = True
tcf_lef = False
tgfb = False
tgfbr = False
twist1 = False
wnt = False
zeb1 = False
zeb2 = False
#RULES
1: akt *= ilk or pi3k
1: dest_compl *= (gsk3b and axin2 and bcatenin_nuc) or (gsk3b and dest_compl)
1: axin2 *= axin2 or tcf_lef
1: bcatenin_memb *= ecadherin and not bcatenin_nuc
1: bcatenin_nuc *= not dest_compl and not bcatenin_memb and (not sufu or not ecadherin)
1: btrcp *= not csn
1: cd44 *= tcf_lef
1: cdc42 *= tgfbr or chd1l
1: cfos *= erk
1: cmet *= hgf or cd44
1: csl *= notch_ic
1: csn *= nfkb
1: delta *= ras
1: dsh *= frizzled
1: ecadherin *= bcatenin_memb and (not snai1 or not hey1 or not zeb1 or not zeb2 or not foxc2 or not twist1 or not snai2)
1: egfr *= egf
1: egr1 *= cfos
1: emt *=  not ecadherin or emt
1: erk *= mek
1: fgfr *= fgf
```

1: foxc2 *= goosecoid or snai1 or twist1
1: frizzled *= wnt
1: fus *= smo
1: gli *= tcf_lef or not sufu
1: gsk3b *= not dsh and not akt and (not csn or not erk or not dest_compl)
1: hey1 *= csl or smad
1: hif1a *= hypoxia
1: igf1r *= igf1
1: ikka *= akt
1: ilk *= smad
1: jagged *= tcf_lef or smad
1: liv1 *= stat
1: loxl23 *= hif1a
1: mek *= raf or not rkip
1: mir200 *= not snai1 and not zeb1 and not zeb2
1: nfkb *= ikka
1: notch *= delta or jagged
1: notch_ic *= notch
1: pak1 *= cdc42
1: patched *= not shh
1: pdgfr *= pdgf
1: pi3k *= ras
1: raf *= ras
1: ras *= sos_grb2 or src or not gsk3b or tcf_lef
1: rkip *= not erk or not snai1
1: shh *= smad or gli
1: snai2 *= erk or bcatenin_nuc or snai2 or twist1
1: smad *= (erk or tgfbr) and (zeb1 or not zeb2)
1: smo *= not patched
1: snai1 *= gli or loxl23 or smad or liv1 or pak1 or csl or egr1 or goosecoid or not btrcp or not gsk3b
1: sos_grb2 *= (pdgfr or cmet or tgfbr or fgfr or igf1r or egfr) and not erk
1: src *= pdgfr or egfr or fgfr or cmet or igf1r
1: stat *= src
1: sufu *= not fus
1: tcf_lef *= bcatenin_nuc
1: tgfb *= goosecoid or snai1 or twist1 or gli
1: tgfbr *= tgfb
1: twist1 *= nfkb or hif1a or tcf_lef or goosecoid or snai1
1: wnt *= gli
1: zeb1 *= (hif1a or snai1 or goosecoid) and not mir200
1: zeb2 *= (hif1a or snai1 or goosecoid) and not mir200
1: egf *= egf
1: fgf *= fgf
1: hgf *= hgf
1: goosecoid *= goosecoid
1: hypoxia *= hypoxia
1: igf1 *= igf1
1: pdgf *= pdgf
1: chd1l *= chd1l

BIBLIOGRAPHY

[1] A. Voulgari and A. Pintzas, "Epithelial-mesenchymal transition in cancer metastasis: Mechanisms, markers, and strategies to overcome drug resistance in the clinic," *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer,* pp. 75-90, 2009.

[2] L. Larue and A. Bellacosa, "Epithelial-mesenchymal transition in development and cancer: role of phosphatidylinositol 3' kinase/AKT pathways," *Oncogene,* pp. 7443-7454, 2005.

[3] K. Hajra, D. Y. S. Chen and E. Fearon, "The SLUG Zinc-Finger Protein Represses E-Cadherin in Breast Cancer," *Cancer Research,* vol. 62, no. 6, 2002.

[4] B. Farley, Artist, *Central Dogma of Biology.* [Art]. UC Berkley, 2015.

[5] S. N. Steinway, J. G. Zanudo, W. Ding, C. B. Rountree, T. Loughran and R. Albert, "Network Modeling of TGFB Signaling in Hepatocellular Carcinoma Epithelial to Mesenchymal Transition Reveals Joint Sonic Hedgehog and Wnt Pathway Activation," *The Journal of Cancer Research,* pp. 5963-5977, 2014.

[6] J. T. Gallagher and M. Lyon, "Molecular structure of Heparan Sulfate and interactions with growth factors and morphogens.," in *Proteoglycans: structure, biology, and molecular interactions*, New York, New York, Marcel Dekker Inc., 2000, pp. 27-59.

[7] O. O. Ogunwobi and C. Liu, "Hepatocyte growth factor upregulation promotes carcinogenesis and epithelial-mesenchymal transition in hepatocellular carcinoma via Akt and COX-2 pathways," *Clin Exp Metastasis,* pp. 721-731, 2011.

[8] C. H. Heldin, "Structural and functional studies on platelet-derived growth factor," *The EMBO Journal,* pp. 4251-4259, 1992.

[9] S. Julien-Grille, R. Moore, L. Denat, O. Morali and V. Delmas, "The Role of Insulin-like Growth Factors in the Epithelial to Mesenchymal Transition," in *The Madame Curie Bioscience Database*, Austin (TX), Landes Bioscience, pp. 2000-2013.

[10] J. M. Buonato, I. S. Lan and M. J. Lazzara, "EGF augments TGFB-induced epithelial-mesenchymal transition by promoting SHP2 binding to GAB1," *Journal of Cellular Science*, vol. 21, no. 128, pp. 3898-3909, 2015.

[11] G. Naharro, S. Tronick, S. Rasheed, M. Gardner, S. Aaronson and K. Robbins, "Molecular Cloning of Integrated Gardner-Rasheed Feline Sarcoma Virus: Genetic Structure of Its Cell-Derived Sequence Differs from That of Other Tyrosine Kinase-Coding onc Genes," *Journal of Virology*, pp. 611-619, 1983.

[12] X. Tong-Chun, G. Ning-Ling, Z. Lan, C. Jie-Feng, C. Rong-Xin, Y. Yang, Y. heng-Long and R. Zheng-Gang, "Goosecoid Promotes the Metastasis of Hepatocellular Carcinoma by Modulating the Epithelial-Mesenchymal Transition," *PloS One*, vol. 9, no. 10, 2014.

[13] S. N. Steinway, J. G. Zanudo, P. J. Michel, D. J. Feith, T. P. Loughran and R. Albert, "Combinatorial interventions inhibit TGFB-driven epithelial-mesenchymal transition and suppoort hybrid cellular phenotypes," *Nature Publishing Journals: Systems Biology and Applications*, 2015.

[14] I. Albert, "Github," 24 April 2014. [Online]. Available: github.com/ialbert/booleannet. [Accessed 2017].

[15] S. Lamouille, J. Xu and R. Derynck, "Molecular mechanisms of epitheilal mesenchymal transition," 2014.

[16] S. R. Wang, A. Saadatpour and R. Albert, "Boolean modeling in systems biology: an overview of methodology and applications," *Physical Biology*, 2012.

[17] J. G. Zanudo and R. Albert, "Cell Fate Reprogramming by Control of Intracellular Network Dynamics," *PLOS Computational Biology,* 2015.

[18] R. Kalluri and R. A. Weinberg, "The basics of epithelial-mesenchymal transition," *Journal of Clinical Investigation.,* pp. 1420-1428, 2009.

[19] G. Pavlopoulous, M. Secrier, C. Moschopoulous, T. Soldatos, S. Kossida, J. Aerts, R. Schneider and P. Bagos, "Using graph theory to analyze biological networks," *BioData Mining,* vol. 4, no. 10, 2011.

ACADEMIC VITA

Academic Vita of Danyas Sarathy
danyas@psu.edu

Education
Major: Physics
Minor: Mathematics
Honors: Physics

Thesis Title: The effect of external signals and combinatorial interventions in the epithelial-to-mesenchymal transition network model
Thesis Supervisor: Dr. Reka Albert

Work Experience
Summer 2015
Army Global Health Research Intern
United States Army Research Institute of Infectious Disease—Ft. Detrick, MD
MAJ Dr. Elena Kwon, MAJ Dr. Michael D'Onofrio

Summer 2016
Army Blast Induced Neurotrauma Research Intern
Walter Reed Army Institute of Infectious Disease—Ft. Glen Annex, MD
Dr. Joseph Long, Dr. James DeMar

Awards:
Elsbach Honors Scholarship in Physics
Strickler Honors Scholarship in Science
Tsui Honors Scholarship
A&K Agarwal Family Honors Scholarship

Professional Memberships:
Phi Kappa Phi Honors Society—Member
American Judo Association—Member

Publications (poster):
Defense against Especially Dangerous Pathogens in Low Infrastructure Environments: Preparedness among East African Clinicians
ASTMH Global Conference, Philadelphia PA, 2015

Post Injury Omega-3 Fatty Acid Emulsion Delivery as a Means to regulate Inflammation from bTBI, as Measured in Rat Livers
Undergraduate Research Poster Exhibition, PSU, 2016

Community Service Involvement:
Volunteer—Mount Nittany Medical Center, Fall 2016-2017

International Education:
Shadowing:
The Royal Liverpool Hospital, England—Summers 2011, 2012
Alder Hey Pediatric Hospital, England—Summers 2011, 2012
Sparsh Hospitals, India—Summer 2014
Indira Gandhi Institute of Child Health, India—Summer 2014
Manipal Hospital, India—Summer 2014
Volunteering:
Swami Vivekananda Tribal Hospital, India—Summer 2014