A Systems-Level Analysis of an Epithelial to Mesenchymal Transition

by

Lindsay Rose Saunders

Department of Biology
Duke University

Date:_______________________

Approved:

_________________________
David R. McClay, Supervisor

_________________________
Xinnian Dong

_________________________
Daniel Kiehart

_________________________
H. Frederik Nijhout

_________________________
Gregory Wray

Dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology in the Graduate School of Duke University

2012
ABSTRACT

A Systems-Level Analysis of an Epithelial to Mesenchymal Transition

by

Lindsay Rose Saunders

Department of Biology
Duke University

Date:_______________________

Approved:

________________________________
David R. McClay, Supervisor

________________________________
Xinnian Dong

________________________________
Daniel Kiehart

________________________________
H. Frederik Nijhout

________________________________
Gregory Wray

An abstract of a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Biology in the Graduate School of Duke University

2012
Abstract

Embryonic development occurs with precisely timed morphogenetic cell movements directed by complex gene regulation. In this orchestrated series of events, some epithelial cells undergo extensive changes to become free moving mesenchymal cells. The transformation resulting in an epithelial cell becoming mesenchymal is called an epithelial to mesenchymal transition (EMT), a dramatic cell biological change that occurs throughout development, tissue repair, and disease. Extensive in vitro research has identified many EMT regulators. However, most in vitro studies often reduce the complicated phenotypic change to a binary choice between successful and failed EMT. Research utilizing models has generally been limited to a single aspect of EMT without considering the total transformation. Fully understanding EMT requires experiments that perturb the system via multiple channels and observe several individual components from the series of cellular changes, which together make a successful EMT.

In this study, we have taken a novel approach to understand how the sea urchin embryo coordinates an EMT. We use systems level methods to describe the dynamics of EMT by directly observing phenotypic changes created by shifting transcriptional network states over the course of primary mesenchyme cell (PMC) ingestion, a classic example of developmental EMT. We systematically knocked down each transcription factor in the sea urchin’s PMC gene regulatory network (GRN). In the first assay, one fluorescently labeled knockdown PMC precursor was transplanted onto an unperturbed
host embryo and we observed the resulting phenotype in vivo from before ingestion until two hours post ingestion using time-lapse fluorescent microscopy. Movies were projected for computational analyses of several phenotypic changes relevant to EMT: apical constriction, apical basal polarity, motility, and de-adhesion.

A separate assay scored each transcription factor for its requirement in basement membrane invasion during EMT. Again, each transcription factor was knocked down one by one and embryos were immuno-stained for laminin, a major component of basement membrane, and scored on the presence or absence of a laminin hole at the presumptive entry site of ingestion.

The measured results of both assays were subjected to rigorous unsupervised data analyses: principal component analysis, emergent self-organizing map data mining, and hierarchical clustering. This analytical approach objectively compared the various phenotypes that resulted from each knockdown. In most cases, perturbation of any one transcription factor resulted in a unique phenotype that shared characteristics with its upstream regulators and downstream targets. For example, Erg is a known regulator of both Hex and FoxN2/3 and all three shared a motility phenotype; additionally, Hex and Erg both regulated apical constriction but Hex additionally affected invasion and FoxN2/3 was the lone regulator of cell polarity. Measured phenotypic changes in conjunction with known GRN relationships were used to construct five unique subcircuits of the GRN that described how dynamic regulatory network states control
five individual components of EMT: apical constriction, apical basal polarity, motility, de-adhesion, and invasion. The five subcircuits were built on top of the GRN and integrated existing fate specification control with the morphogenetic EMT control.

Early in the EMT study, we discovered one PMC gene, Erg, was alternatively spliced. We identified 22 splice variants of Erg that are expressed during ingression. Our Erg knockdown targeted the 5’UTR, present in all spliceoforms; therefore, the knockdown uniformly perturbed all native Erg transcripts (∑Erg). Specific function was demonstrated for the two most abundant spliceoforms, Erg-0 and Erg-4, by knockdown of ∑Erg and mRNA rescue with a single spliceoform; the mRNA expression constructs contained no 5’UTR and were not affected by the knockdown. Different molecular phenotypes were observed, and both spliceoforms targeted Tbr, Tel, and FoxO, only Erg-0 targeted FoxN2/3 and only Erg-4 targeted Hex. Neither targeted Tgif, which was regulated by ∑Erg knockdown sans rescue. Our results suggest the embryo employs a minimum of three unique roles in the GRN for alternative splicing of Erg.

Overall, these experiments increase the completeness and descriptive power of the GRN with two additional levels of complexity. We uncovered five sub-circuits of EMT control, which integrated into the GRN provide a novel view of how a complex morphogenetic movement is controlled by the embryo. We also described a new functional role for alternative splicing in the GRN where the transcriptional targets for two splice variants of Erg are unique subsets of the total set of ∑Erg targets.
Dedication

I would like to dedicate this dissertation in memory of my grandfather, who passed away last year. He was always proud of me and believed in me even when I didn’t believe in myself. I’m sure my “I can do it myself” attitude comes from him and I’m surely a better scientist for it… and also a better electrician.
# Contents

Abstract ........................................................................................................................................... iv

List of Figures .................................................................................................................................. xii

List of Abbreviations ......................................................................................................................... xv

Acknowledgements ............................................................................................................................... xvi

1. Introduction: EMTs and GRNs ......................................................................................................... 1

   1.1 EMT in Embryonic Development and Cancer Metastasis ......................................................... 1

   1.2 Ingression: EMT in the Sea Urchin Embryo ............................................................................. 9

   1.3 Understanding Development on a Systems Level Using GRNs .............................................. 16

2. In Vivo Phenotypic Analysis of an EMT ......................................................................................... 24

   2.1 A Brief History of the State of EMT Research ........................................................................ 24

      2.1.1 Apical Constriction ........................................................................................................... 25

         2.1.1.1 Apical Constriction Drives Cell Shape Change ......................................................... 25

         2.1.1.2 Apical Constriction in Sea Urchin Embryos ............................................................. 28

         2.1.1.3 EMT and Apical Constriction .................................................................................... 30

      2.1.2 Cell Polarity ....................................................................................................................... 32

         2.1.2.1 Epithelial vs. Mesenchymal Polarization of Cells .................................................... 32

         2.1.2.2 Cell Polarity in Sea Urchin Embryos ......................................................................... 34

         2.1.2.3 EMT and Cell Polarity ............................................................................................... 35

      2.1.3 Cell Motility ....................................................................................................................... 36

         2.1.3.1 Mesenchymal Cells Acquire Motility ......................................................................... 36

         2.1.3.2 Motility in Sea Urchin Embryos ............................................................................... 38
<table>
<thead>
<tr>
<th>Section</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>2.5.1.3</td>
<td>Results</td>
<td>64</td>
</tr>
<tr>
<td>2.5.2</td>
<td>Analyzing Distribution of Cells Across Basal Lamina</td>
<td>70</td>
</tr>
<tr>
<td>2.5.2.1</td>
<td>Inferring Motility in a Tethered System</td>
<td>70</td>
</tr>
<tr>
<td>2.5.2.2</td>
<td>Methods</td>
<td>72</td>
</tr>
<tr>
<td>2.5.3</td>
<td>Cell Distribution Results</td>
<td>73</td>
</tr>
<tr>
<td>2.5.4</td>
<td>A Relationship Between De-Adhesion and Latent Motility</td>
<td>77</td>
</tr>
<tr>
<td>2.6</td>
<td>Discussion of In Vivo Time-Lapse Results</td>
<td>78</td>
</tr>
<tr>
<td>3.</td>
<td>Network Control of Invasion in EM</td>
<td>84</td>
</tr>
<tr>
<td>3.1</td>
<td>Experimental Design</td>
<td>84</td>
</tr>
<tr>
<td>3.2</td>
<td>Methods</td>
<td>85</td>
</tr>
<tr>
<td>3.2.1</td>
<td>Animals</td>
<td>85</td>
</tr>
<tr>
<td>3.2.2</td>
<td>Injections</td>
<td>85</td>
</tr>
<tr>
<td>3.2.3</td>
<td>Immunostaining</td>
<td>85</td>
</tr>
<tr>
<td>3.2.4</td>
<td>Image Processing and Statistical Analysis</td>
<td>86</td>
</tr>
<tr>
<td>3.3</td>
<td>Results</td>
<td>87</td>
</tr>
<tr>
<td>3.4</td>
<td>Discussion</td>
<td>92</td>
</tr>
<tr>
<td>4.</td>
<td>Alternative Splicing of Erg in the GRN</td>
<td>95</td>
</tr>
<tr>
<td>4.1</td>
<td>Introduction</td>
<td>95</td>
</tr>
<tr>
<td>4.1.1</td>
<td>Mechanism of Alternative Splicing</td>
<td>95</td>
</tr>
<tr>
<td>4.1.2</td>
<td>Alternative Splicing in the Sea Urchin Embryo</td>
<td>96</td>
</tr>
<tr>
<td>4.1.3</td>
<td>Alternative Splicing in EMT</td>
<td>97</td>
</tr>
<tr>
<td>4.2</td>
<td>Experimental Design</td>
<td>98</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1: The First Schematic to Depict Organization of EMT Events............................... 2
Figure 1.2: Five Requisite Cellular Events Coordinate to Direct a Complete EMT............. 4
Figure 1.3: Snail Transcriptionally Represses and Endocytosis Cadherin......................... 6
Figure 1.4: Apical Constriction Creates a Bottle Cell Shape. ............................................ 7
Figure 1.5: A 16-Cell Sea Urchin Embryo has 4 Micromeres............................................ 10
Figure 1.6: Ingression of PMCs is the Sea Urchin’s First EMT. .......................................... 12
Figure 1.7: PMCs form a ring around the blastocoel.......................................................... 13
Figure 1.8: PMC GRN Transcription Factor Tel is Required for Larval Skeleton............... 14
Figure 1.9: Replacement of skeleton by transfating............................................................. 14
Figure 1.10: Detail of Sea Urchin PMC Ingression.............................................................. 15
Figure 1.11: The Sea Urchin EndoMesodermal Gene Regulatory Network..................... 20
Figure 1.12: Screenshot of Biotapestry: A Dynamic Multi-Layered GRN View................. 21
Figure 2.1 Apical Constriction Precedes Ingression of C. Elegans Endoderm.................... 26
Figure 2.2: Bottle Cell Formation in Sea Urchin Primary Invagination.............................. 29
Figure 2.3: EMT is Directed by Mechanical Stresses......................................................... 31
Figure 2.4: Epithelial Cell Polarity...................................................................................... 32
Figure 2.5: Mesenchymal Cell Polarity.............................................................................. 33
Figure 2.6: Polarity of the Sea Urchin is Established by 2-cell Stage................................. 35
Figure 2.7: Random Motility vs. Directed Motility............................................................. 37
Figure 2.8: Urchin Lamellipodia at the Tip of the Archenteron........................................ 39
Figure 2.9: Neural Crest Motility During EMT ................................................................. 40
Figure 2.10: Basic Structure of an Adherens Junction ....................................................... 41
Figure 2.11: Snail and Twist are Required for EMT ............................................................ 43
Figure 2.12: Many EMT Regulators Converge onto Snail .................................................. 44
Figure 2.13: Laminin Stains the Basement Membrane of Urchins Embryos ....................... 46
Figure 2.14: Anchor Cell Invasion Clears a Path Through Laminin ............................... 47
Figure 2.15: Micromere Transplant Assay ........................................................................ 51
Figure 2.16: Still Frames from Time-Lapse Movies .............................................................. 60
Figure 2.17: Examples of Cell Tracings .............................................................................. 62
Figure 2.18: Shapes Descriptors Determined by Principal Component Analysis ............. 65
Figure 2.19: Principal Component One ................................................................................ 66
Figure 2.20: Principal Component Two ............................................................................... 68
Figure 2.21: Principal Component Three ............................................................................ 69
Figure 2.22: Control PMCs Ingress Completely into the Blastocoel ................................. 71
Figure 2.23: Motility Inferred by Cell Area Distributions .................................................. 76
Figure 2.24: Alx1, Twist, and Snail Regulatory Mechanism ............................................... 77
Figure 2.25: A Subcircuit for Apical Constriction .............................................................. 80
Figure 2.26: A Subcircuit for Apical-Basal Polarity ............................................................ 81
Figure 2.27: A Subcircuit for Motility ................................................................................ 82
Figure 2.28: A Subcircuit for De-Adhesion ...................................................................... 83
Figure 3.1: Invading PMCs Leave a Laminin Hole Behind .................................................. 88
Figure 3.2: Laminin Staining in Control Time Course ......................................................... 88
Figure 3.3: Laminin Staining in Knockdowns............................................................................ 89
Figure 3.4: Invasion Analysis..................................................................................................... 91
Figure 3.5: A Subcircuit for Invasion....................................................................................... 94
Figure 4.1: Mechanism of Alternative Splicing........................................................................ 96
Figure 4.2: Erg Spliceoform Variations and Relative Abundance............................................ 104
Figure 4.3 In Situ Hybridization of ΣErg throughout Development.......................... 105
Figure 4.4: Two Erg Spliceoforms Regulate Different PMC GRN Targets.................. 106
Figure 4.5: A Subcircuit of Erg Spliceoforms. .............................................................. 108
Figure 5.1: Integrated View of GRN. ....................................................................................... 110
Figure 5.2: Emergent Self Organization Map of Time-Lapse Data. ............................ 115
Figure 5.3: Clustering Analysis ............................................................................................. 116
List of Abbreviations

EMT .......................................................... Epithelial to Mesenchymal Transition
GRN .......................................................... Gene Regulatory Network
PMC .......................................................... Primary Mesenchyme Cell
MASO ...................................................... Morpholino Anti-Sense Oligonucleotide
hpf .............................................................. hours post fertilization
AJ ............................................................... Adherens Junction
UTR .......................................................... Untranslated Region
PCA .......................................................... Principal Component Analysis
PCN .......................................................... Principal Component N
ΣLvErg ...................................................... Total Pool of LvErg Spliceoforms
LvErg0 ....................................................... LvErg Spliceoform Full Length
LvErg4 ....................................................... LvErg Spliceoform Exon 4 Skipped
MB ............................................................. Mesenchyme Blastula
EG ............................................................. Early Gastrula
LG ............................................................. Late Gastrula
SMRT® .......................................................... Single Molecule Real-Time Sequencing Technology
ESOM .......................................................... Emergent Self Organization Map
Acknowledgements

There are so many people I would like to thank I’m not sure I can possibly cover everyone! Let me begin by saying that not a day goes by that I’m not completely amazed at how blessed I am to have so many wonderful people in every aspect of my life. There are so many people who have helped me become the person I am today, but the first person I need to thank is my allergist, Dr. Patricia Lugar. She saved my sanity and gave me my life back. I don’t even want to consider how different my life would be different if I hadn’t finally found a physician who actually listened and helped me.

Of course I also need to acknowledge my family. They have been loving and supportive through all the craziness. To my mother, you didn’t just give birth to me, you gave me everything I am today. They say you turn into your mother, and if that’s true I can only say I wouldn’t have it any other way. To my father, I want to thank you for always being the voice of reason. Dad, you are my rock. Please stop drinking orange juice every day, I want you stick around a long long time. To my sister Alison, I just want to say, “It’s the lady fit dance, the lady fit dance.” I wish we lived closer together, I really cherished those months we spent together before you got lawyerized. Ali, wherever you go and whatever you do, you’re going to kick butt at it… you just have to stick with it.

Of course, I also need to thank all of the teachers and mentors who shaped me intellectually. My high school Biology teacher Mrs. Hipp gave me my first taste for real
hands on dirty science. I still remember our first molecular biology lab when none of my lab partners wanted to micropipette our sample onto the gel... from that moment on, I knew whatever I did in life it had to involve a micropipette. I was destined to move tiny accurate amounts of small liquids into slightly less tiny amounts of liquid. If it weren’t for her, I might have ended up a chemist... or worse yet, a computer scientist!

My college advisor, Dr. Brittingham, I must thank for giving me such amazing teaching opportunities and fostering a love for embryos I never knew I had. I love all embryos in all shapes and sizes, I just want to eat them up. Next I need to thank Dr. Igor Dawid and his technician, my mentor, Martha Rebberg. This was first taste ‘real’ taste for bench work and I think I may have been spoiled a little. There were so many people in Igor’s lab from so many countries and they functioned so harmoniously together. What a wonderful place to learn science. I think NIH is the research communities equivalent to Disney Land. Finally, Uncle Dave, Captain Dave, Dave... I couldn’t have done this without you (and I don’t just mean your grant money). When I’ve stumbled you’ve let me find my own way back and I think you knew I had something I didn’t even know I had. There was a time I didn’t think I could do this. I didn’t think I’d ever be myself again, but you were patient and when I was finally back (again, thank you Dr. Lugar!) I realized you’d been there beside me all along and I couldn’t see it. During my time in graduate school I’ve been through some rough times, physically, emotionally, scientifically... but Dave’s been there for me through it all, just waiting patiently for the
data and knowing all along I could do it. And when the data came, boy did it come!!!!

Lastly, Hal W. Canary III, I think I wanna marry you. I hope you still want to marry me after what I’ve put you through! Thank you for the backrubs. Thank you for doing the dishes. Thank you for mowing the lawn. Thank you for feeding the dogs. Thank you for making me coffee every morning. Thank you for putting up with me. Thank you for loving me. I can’t believe the wedding is less than a month away… we better get our marriage license pronto!
1. Introduction: EMTs and GRNs

1.1 EMT in Embryonic Development and Cancer Metastasis

An epithelial to mesenchymal transition (EMT) is a dramatic biological process in which epithelial cells undergo adhesive changes, allowing them to break free from their neighbors and subsequently acquire an independently moving mesenchymal phenotype. The EMT phenomenon was familiar to embryologists for more than a century, but it wasn’t until 1968 when Elizabeth Hay first formally describe the EMT as a series of coordinated events that the developmental community had a template with which to study the mechanisms of EMT (See Figure 1.1) (Hay 1968). EMT research has progressed a long way into this modern age of genomes. EMT research remains an exciting field with many questions waiting to be answered and many researchers actively pursuing mechanistic understanding.
Figure 1.1: The First Schematic to Depict Organization of EMT Events. The chick primitive streak model shows cells detaching form the epithelium and becoming mesenchymal in an orchestrated fashion. This was the first template to methodically describe the separate cellular events leading to EMT (Hay 1968).

Three distinct types of EMT have been categorized based on their different biological functions. Developmental EMT (Type 1) encompasses events that occur during embryonic development, including implantation, gastrulation, and neural crest formation. Type 2 EMT occurs during the course of regular maintenance in adult tissues, such as wound healing, regeneration, and fibrosis. Disease or cancer type EMT (Type 3) is an abnormal EMT event, which defines a crucial step in the path from malignancy to metastasis. Most researchers agree that the Type 3 EMT is “likely to be a reactivation of developmental paradigms, albeit with fatal consequences” (Kalluri 2009). In gaining an understanding of developmental EMT mechanics, researchers may also
gain insights into how a tumor metastasizes, thereby addressing important health concerns in areas of both birth defects and cancers.

Pathological EMT events may be loosely coordinated and difficult to characterize, but in embryonic development, EMT combines several individual cellular behaviors into one strictly regulated and precisely timed series of events whose onset can be accurately predicted. The individual components that together comprise a complete EMT can be sorted into five broad categories: cell shape change beginning with apical constriction, de-adhesion from the epithelium, acquisition of directed motility, basement membrane invasion, and loss of apical-basal polarity (See Figure 1.2). Disruption of any one of these components causes system wide failure. All five cellular behaviors are vital to the EMT process as a whole.
Figure 1.2: Five Requisite Cellular Events Coordinate to Direct a Complete EMT. (1) Invasion: The cell body bulges into basal lamina. (2) Apical Constriction: The apical end contracts creating a bottle cell shape. (3) De-Adhesion: Adherens junctions are broken down, epithelial adhesion molecules are endocytosed, and adhesive affinity for extracellular matrix is gained. (4) Polarity Change: Apical-basal polarity is lost and mesenchymal polarity is acquired. (5) Directional Motility: The nascent cell moves freely toward external migratory cues. The cell is now fully mesenchymal in phenotype.
Of all the requisite EMT events, the best documented is de-adhesion: the cell’s ability to detach from neighboring epithelial cells. The transitioning cell must disassemble its tightly bound adherens junctions (AJ) by both transcriptionally down-regulating and endocytosing cadherins and catenins. Posttranslational modifications of cadherins and catenins also facilitate AJ removal (D’Souza-Schorey 2005). In widely varying cancer types and most developmental models, the major regulators of de-adhesion are the Snail family of transcription factors. Snail genes are direct transcriptional repressors of cadherins and also regulate the endocytosis of cadherin-catenin complexes from the cell membrane facilitating their degradation (See Figure 1.3) (Wu and McClay 2007).
Figure 1.3: Snail Transcriptionally Represses and Endocytosis Cadherin. (D) The co-injection of SnaMASO rescues the reduction of cadherin expression in Pmar1-overexpressed embryos. (E) SnaMASO (red) shows the apical localization of cadherin observed in adherens junctions with CadTM-GFP. (F) Same embryo as E; the intracellular punctate GFP signals indicate endocytosed cadherins (arrowheads), which do not overlap with SnaMASO-injected cells (with rhodamine-dextran). (G) The punctate GFP signals (arrowheads) can be seen in PMCs. (H) Schematic of the experimental design in E-G. Adapted from (Wu and McClay 2007)

The first observable morphological phenotypic change that indicates an EMT is underway is a change in shape of the epithelial cell. The apical surface of cells is covered in a dense actin-myosin meshwork. Coordinated contractions of actin and myosin cinch the apical side of the cell, and this causes elongation of the cell along its apical-basal axis. The cell’s nucleus and cytoplasm are pushed basally and an apical stalk is stretched thin from its attachment to the adherens junctions. The redistribution of cytoplasm causes bulging at the cell’s basal surface and this is first force generated during EMT to push the cell away from the epithelial sheet where it originated. The
resulting shape change from apical constriction may be described as a bottle cell, tear drop, or pear (See Figure 1.4).

**Figure 1.4: Apical Constriction Creates a Bottle Cell Shape.** The cell depicted in purple is apically constricting. The actin-myosin network at the apical end of the cell contracts. This contraction forces the nucleus and most of the cell’s body toward the basal end. The cell bulges into the extracellular space and the long apical stalk creates a bottle cell shape.
As apical constriction is under way the basal side of the cell actively remolds the basement membrane immediately above so that it is possible for the cell to squeeze out of the epithelial layer and into new territory. The basement membrane is comprised of over 50 different macromolecules but the bulk of the composition is a stable crosslinking of collagens and laminins (Rowe and Weiss 2008). To overcome this physical barrier, the basement membrane is degraded or softened by either secretion of proteases, including matrix metalloproteases, which are the best documented, or by protease-independent mechanical remodeling events which are not yet well understood. (Levayer and Lecuit 2008).

After the cell completes these first three EMT processes, the basal surface protrudes across the hole it has created and the bulk of the cell body invades into the extracellular space leaving behind a long narrow “tail” that is still tethered to the epithelium by the adherens junction (See Figure 1.10) (Katow and Solursh 1981). Shortly following protrusion, the cell acquires motility, loses epithelial polarity, and migrates away as a nascent mesenchymal cell (Ahlstrom and Erickson 2009). After completion of EMT the mesenchymal cell changes much of its gene expression profile; it expresses new adhesion molecules like N-cadherin that have a high affinity for extracellular matrix, and its regulatory state will cements the cell’s new fate with as a mesenchymal phenotype (Baum, Settleman, and Quinlan 2008).
1.2 Ingression: EMT in the Sea Urchin Embryo

The sea urchin embryo is relatively simplistic and develops quickly, making it a classic model for studying morphogenesis. Sea Urchin embryos are easy to manipulate by both physically swapping cells and introducing molecular perturbations. They develop externally in room temperature cultures and are optically clear which makes them an ideal system for direct observation of morphological change \textit{in vivo} using time-lapse microscopy. The PMC EMT event begins at about 9.0 hours post fertilization (hpf) at the vegetal pole with the ingestion of 32-cells and is completed in less than an hour, making real-time phenotypic analysis of EMT experimentally tractable.

Primary mesenchyme cell (PMC) ingestion at the onset of gastrulation in the sea urchin embryo is a textbook example of developmental EMT. The primary mesenchyme cells are derived from micromere precursors born at fourth cleavage about 2.5 hpf. Micromeres result from an asymmetric fourth cleavage, which produces four large cells, the macromeres, and four small cells, the micromeres (See Figure 1.5). The micromeres are located at the vegetal pole and the first micromere cleavage is also an unequal division, which results in four large micromeres and four small micromeres at the 32-cell stage.
Figure 1.5: A 16-Cell Sea Urchin Embryo has 4 Micromeres. The asymmetric fourth cleavage results in 8 equally sized mesomeres in the animal half. The vegetal half result is 4 large macromeres and 4 small micromeres, the PMC precursors.

At both the 16 cell and 32 cell stage, the micromeres are easily isolated via microsurgical manipulation and can be swapped onto different host embryos which allows for accurate lineage tracing and in this study, offers a way to isolate experimental variables and focus only on cells that will ingress. One micromere from a molecularly perturbed fluorescently labeled donor embryo can be put in place of one micromere from an unperturbed host embryo. Later, as EMT progresses, the descendants of the fluorescent donor micromere can be directly observed alongside the host’s control micromere descendants that provide an internal control for the timing of EMT events.

The descendants of the large micromere lineage initiate ingress into the blastocoel at about 9.0 hpf (See Figure 1.6). Ingression is an example of Type I EMT and four of the five requisite EMT components have been previously studied in the sea urchin model. PMCs were shown to repress the transcription of cadherin and up-
regulate cadherin endocytosis through a Snail/Twist mediated pathway. When Snail or Twist was knocked down in the embryo, PMCs did not de-adhere and failed to ingress (Wu and McClay 2007; Wu, Yang, and McClay 2008). PMCs also demonstrated a typical shape change in response to apical constriction before ingress (See Figure 1.10) (Katow and Solursh 1980). Following ingress, PMCs lose their apical-basal polarization and appear uniformly round. The directed motility and subsequent migration toward the ectoderm (Peterson and McClay 2003; Ettensohn and McClay 1986) of the newly emerged mesenchymal cell has been described and can be directly observed in vivo (Katow and Solursh 1980). The only key EMT feature that has not been experimentally observed in the urchin model is invasion and it was not known whether the invasion process happened via proteolytic degradation or if the basal lamina was loose enough that PMCs could force their way through (Wu, Ferkowicz, McClay 2007). Thus, an important goal of this current research was to directly demonstrate how the invasion component of PMC EMT works.
**Figure 1.6: Ingression of PMCs is the Sea Urchin's First EMT.** (Left) The four small red cells born at 4th cleavage are the micromeres, precursors to the PMC lineage. (Middle) The 32 cells descended from the micromeres at the “thickened” vegetal plate stage 9.0 hpf. Cells are elongated and bulging into the blastocoelar cavity, although they remain attached to the surrounding epithelium. (Right) By 10.0 hpf, the PMCs have detached completely from the epithelium and the hole from which they ingressed is now sealed. The PMCs are now freely moving mesenchymal cells inside of the blastocoel.

Immediately upon completion of EMT, the nascent PMCs divide once and migrate away from the ingression site toward ectodermal cues (Rottinger et al. 2008; Peterson and McClay 2003). Ultimately, PMCs will end their migration and form a single cell ring around the vegetal half of the blastocoel with two dense clusters on the left and right sides (Peterson and McClay 2003) (See Figure 1.7). They fuse to form a syncytium and secretion of the larval skeleton begins where the PMCs are densely clustered (Wu, Ferkowicz, and McClay 2007).
Figure 1.7: PMCs form a ring around the blastocoel. When PMCs reach their migratory destination they are organized in an ordered fashion around the blastocoel. The clusters of PMCs to the left and right are the sites where the larval skeleton will begin to be secreted. PMCs are immunostained with 1D5 mAB, a PMC specific antibody whose antigen is a cell surface protein.

The larval skeleton is birefringent making it very simple to score as a phenotype and in most cases when PMCs fail to ingress the embryo fails to secrete the skeleton (See Figure 1.8). There is an exception to this general rule and occasionally when PMCs do not complete ingression other non-skeletogenic mesodermal cells compensate by changing their fate and becoming PMCs in a transition termed transfating (See Figure 1.9). When this special transfating recovery mechanism is activated, the embryo produces a fully intact larval skeleton, albeit several hours behind the production of a control embryos larval skeleton (Ettensohn et al. 2007). It is important to keep this regulative phenomenon in mind when interpreting results of PMC perturbations.
Figure 1.8: PMC GRN Transcription Factor Tel is Required for Larval Skeleton. Control embryos 24.0 hpf (top) have perfectly formed skeletons are seen under crossed polarized light. 24.0 hpf Tel MASO embryos (bottom) have little or no skeleton.

Figure 1.9: Replacement of skeleton by transfating. Control Embryo 20 hpf (left), Micromere-Deficient Embryo 24.0 hpf (Right). The micromere deficient embryo replaced PMCs by transfating. (Ettensohn et al. 2007)
Figure 1.10: Detail of Sea Urchin PMC Ingression. (4) View from inside the blastocoel looking down toward the vegetal pole, future PMCs have begun to bulge inside. (5) Cells fated to become PMCs make a bottle cell or tear drop shape before they ingress. (Katow and Solursh 1981)
1.3 Understanding Development on a Systems Level Using GRNs

Biological research is in the midst of a data boom; as all means of data collection become faster, easier and cheaper, the amount of data collected increases exponentially and systems level approaches to biological problems become more feasible. Rapid technology advancements led to more accurate and more powerful strategies for answering a wide array of biological questions. Almost every conceivable method for data collection has the capacity to be used for high-throughput approaches, from single molecule expression sequencing to high-resolution time-lapse imaging data acquisition has been scaled up. For example, microscopes and their accompanying acquisition software are now capable of producing tremendous amounts of high-resolution image data by capturing multiple x,y positions, multiple fluorescent channels, and multiple z positions, all at regularly timed intervals (Ahlstrom and Erickson 2009). Biologists are collecting more informative higher quality data and at an exponentially increasing rate; however, this advance in data acquisition would prove meaningless if it were not accompanied by equal advance in analytical tools for interpreting these mammoth data sets. Systems biology techniques provide an opportunity to address these challenges and answer big picture biological questions.

The growth in the field of systems biology has been a response to challenges faced by scientists in all areas of research as they struggle to extract meaning from their data (Westerhoff and Palsson 2004). One new approach to emerge from the field of
systems biology is to create a map of relationships among transcription factors in a given situation. Biologists knew that relatively few transcription factors were responsible for the control of expression of many thousands of transcripts, which are translated into many thousands of proteins, the functional unit of a cell (Nowick and Stubbs 2010). Thus, the study of gene regulation by perturbing transcription factor states in a cell is an elegantly simple approach to gain insight into regulatory function. By causally linking regulatory relationships, a gene regulatory network (GRN) emerges (Davidson 2010). Networks models provide a unique template for understanding how large-scale change occurs within a developing system: germ layer organization, cell fate specification, embryonic tissue morphogenesis, etc.

GRNs have been constructed in several different ways. Some were computational models built from large data sets and these provide a prediction of how a biological system might behave in a given situation. Predictive GRNs direct scientists toward questions in a biological system. Rather than randomly testing all possible variables, these enable biological experiments to authenticate predicted nodes and thereby narrow the number of experimental variables needed.

A handful of GRNs were painstakingly built from experimental data so that every node represents an authenticated biologically relevant connection. Experimentally based GRNs have an advantage over mathematically modeled GRNs or GRNs drawn from large data sets simply because more nodes in the GRN are
authenticated. These GRNs are ready for next-level systems analysis since the scientist can bypass the tedious authentication step. The sea urchin GRN is easily the most advanced experimentally confirmed GRN and has been continually improved upon by the collective sea urchin development community (See Figure 1.11 and 1.12) (Davidson et al. 2002). It is both a powerful predictive and descriptive model of sea urchin development. At the time of this study, the sea urchin GRN contains 120 authenticated nodes that are confirmed by thousands of experiments. However, in spite of the complexity, breadth, and depth of the current GRN model, a complete understanding is still years away. Uncovering the dynamics of an entire developing embryo, even a simple embryo such as the sea urchin, is an enormous challenge.
Figure 1.11: The Sea Urchin EndoMesodermal Gene Regulatory Network.
A view of the GRN with Bio Tapestry; the graphical user interface used to view the urchin GRN; it is freely distributed under the General Public License and is accessible to anyone interested in learning about the network dynamics of urchin development. The GRN is continually updated with the most current experimental data and Bio Tapestry keeps the entire sea urchin community simultaneously up to date on the most recent network build. The centralization of community wide experimental results also fosters a spirit of collaboration and keeps the urchin model on the cutting edge of GRN based research. Bio Tapestry also offers an interactive view of the network. The GRN itself is essentially a map of all the known transcription factors and signaling molecules present during development and all predicted and confirmed interactions as they mold cell fates with changing network states. The static view of the total GRN shows temporal information from top to bottom; the earliest embryonic network states launched from maternal transcripts are at top and the bottom shows the final cell fate determinates. The GRN is subdivided into embryonic territories; the leftmost territory represents the network state of PMCs. The GRN was first described in (Davidson et al. 2002) and the most current view of the network can be found at http://sugp.caltech.edu/endomes/
Figure 1.12: Screenshot of Biotapestry: A Dynamic Multi-Layered GRN View. BioTapestry provides users with a dynamic view network state changes over time; Screenshots show how using a time slider one can view the exact network state running at one hour intervals through late gastrulation. Clicking any gene will display only nodes relevant to that gene and the experimental data used to determine the regulation at each node. BioTapestry depicts solid lines in the GRN to indicate experimentally confirmed cis-regulatory interactions and dashed lines refer to cis-regulatory elements that have yet to be functionally authenticated. A target pad with a bent arrow represents each gene in the network; temporal expression of genes can be viewed with the time slider, genes are expressed only if they are colored. Links to a gene’s transcriptional targets begin from the bent arrow and terminates in an arrowhead wherever that gene acts as an activator; the links terminate in a bar to when the relationship is repressive. (Longabaugh, Davidson, and Bolouri 2009)
The Primary Mesenchyme Cell GRN is of critical importance for this study because it is the best characterized GRN in the embryo and the transcription factors in that territory must govern their signature morphogenetic event as well as the specification of PMCs. We hypothesized that the GRN could be used to gain a systems level understanding of EMT events in the PMC lineage along their path to differentiation; this will be a novel use of the GRN. Connecting the GRN's transcriptional regulation to the cell biology of ingress and morphogenesis was a lofty undertaking, but an extensive and integrative approach was, we believed, the only way to reach a comprehensive understanding of how a morphogenetic movement molds the embryo.

We decided the first phase of understanding developmental EMT was to figure out how the PMC GRN provided a control function governing each step of EMT. In this study, we find that the PMC GRN contains five discrete subcircuits of control over EMT. Each subcircuit viewed in isolation describes the transcriptional control of a single component from the entire orchestration of EMT events; when all five subcircuits are viewed within the context of the current developmental GRN alongside cell fate specification, the result is a holistic picture of how a micromere becomes a PMC through both dramatic cell biological changes and molecular differentiation. We also add a new level of gene regulation to the network by identifying a role for alternative splicing of the Erg transcription factor; we place two splice variants of Erg in the PMC GRN within
overlapping but unique subcircuits of $\Sigma$Erg’s current targets which have been described by the GRN before splice variation of Erg was discovered. The GRN enabled us to find different molecular targets for alternatively spliced transcription factors acting in the same cell type; a result that would have been very difficult to discover \textit{de novo} without the GRN to point us towards potential targets. The total result of this study is a deeper understanding of how EMT is organized by gene regulatory networks and two novel uses for gene regulatory network logic in biological discovery.
2. *In Vivo* Phenotypic Analysis of an EMT

2.1 A Brief History of the State of EMT Research

An epithelial to mesenchymal transition (EMT) is a complex cell biological event enacted during embryonic development with complex timing and great precision in order to shape and develop a complex body plan. By definition an EMT is characterized by an epithelial cell’s adoption of a new mesenchymal fate; it detaches from the surrounding epithelium, invades through the basal lamina, and freely moves about the extracellular space. This EMT process continues to be important throughout the entire lifespan of an organism.

In adults, the EMT is activated in response to inflammation aiding in wound healing and when developmental EMT programs are inappropriately reactivated in a cancer cell the results can be lethal (Acloque et al. 2009). Cancerous EMT phenotypes are beginning to be used in clinical therapies to predict drug resistance of malignant tumors(Krasnapolski, Todaro, and de Kier Joffe 2011). Unfortunately, cancer activated EMTs are difficult to study experimentally because cancer cells do not execute the timed series of EMT events with the same level of predictability and synchrony that is characteristic of developmental EMTs (Singh and Settleman 2010). The mechanisms of EMT are more easily studied in embryos where the timing of EMT has been meticulously parsed and the morphogenetic change occurs *in vivo* during that established time window with reasonable dependability.
Today’s understanding of the cellular mechanisms responsible for an EMT is extensive in both breadth and depth. Creating a new mesenchymal phenotype from an existing epithelial cell requires a complex set of mechanistic changes. The total body of work on EMT mechanics was built primarily using reductionist approaches. Biologists have been dissecting individual EMT events into a minutia of details for years and the result is an impressive knowledge base of EMT mechanics. Transcending traditionally compartmentalized fields of study, the systems biology toolkit allows scientists from any field to find answers to complex morphological questions, such as how EMT is orchestrated, from an integrated systematic approach rather than continue down the well traveled path of field specific reductionist research. The study we present here approaches EMT from this integrating perspective; we designed our experiments using knowledge of cell biology combined with classic embryological manipulations and with a systems level analysis of the results.

2.1.1 Apical Constriction

2.1.1.1 Apical Constriction Drives Cell Shape Change

The scope of this study spans five cell biological events comprising morphological change through an EMT. The first observable phenotypic change in the embryo to indicate an EMT is underway is apical constriction. Apical constriction is a driving force of epithelial morphogenesis across widely studied developmental systems and in many cases it is responsible for the very first gastrulation events. In *Xenopus*
*laevis* embryos, gastrulation begins by forming the dorsal lip of the blastopore from bottle cells created via apical constriction and entire sheets of epithelial cells are internalized forming germ layers (Keller 1981). Ventral furrow formation, the first gastrulation even in *Drosophila melanogaster*, is driven by apical constriction (Leptin and Grunewald 1990) and so is the first gastrulation event of *C. elegans* development with the internalization of just two endodermal cells (Putzke and Rothman 2003) (See Figure 2.1).

**Figure 2.1 Apical Constriction Precedes Ingression of *C. Elegans* Endoderm.** The endodermal precursors (green) constrict at the apical surface before they are internalized. (Roh-Johnson et al. 2012)
The purse string model has been the dominant model used to explain the mechanism of apical constriction. In this model, force is generated when an actin-myosin belt that links a group of cells together around their circumference uniformly contracts and cinches the cells together all at once. More recent research in the field of apical constriction suggests a mechanism for uncoupled pulsation of actin-myosin networks creating constriction forces in a ratcheting model (Martin et al. 2009). Still another study proposes a clutch model for generating constriction forces; this model is based on the observation that pulsed actin-myosin contractions are constitutively active well before shape changes occur and do not by themselves constitute a mechanism for shape change (Roh-Johnson, et al. 2012). In other words, some unknown force is putting the “brakes” on continuous cytoskeletal contractions, which tightens the contractions at the time of shape change. There may not be yet be an accepted mechanism for understanding how forces are generated during apical constriction mediated shape change, but there is no doubt that these forces are an important component in remodeling of epithelial sheets and EMT alike.
2.1.1.2 Apical Constriction in Sea Urchin Embryos

In sea urchin development, primary invagination of the archenteron (See Figure 2.2) is driven by forces generated during apical constriction (Ettensohn 1985). When bottle cells are laser ablated in the midst of invagination, the archenteron’s ability to continue with elongation is impaired. Additionally, apical constriction is dependent upon the cell’s attachment to the outer hyaline layer and necessary for initiating invagination. When this attachment is disrupted, bottle cells do not form and primary invagination fails to begin (Kimberly and Hardin 1998).

Apical constriction is an important shape change during PMC ingression, however it has not been well studied on a mechanistic level in urchin development. One study suggests apical constriction is not an absolute requirement for ingression. Localized of F-actin staining was shown concentrated on the apical surface of PMCs during EMT and ultra-structure showed ingressing PMCs made the tear drop shape typical of an apically constricted cell. When constriction was inhibited by cytochalasin, an f-actin polymerization inhibitor, PMCs ingressed abnormally and very slowly, but ingestion was not entirely blocked (Anstrom 1992). However, global introduction of cytochalasin into the embryo may not offer the most accurate picture of endogenous mechanism within one cell type.
Figure 2.2: Bottle Cell Formation in Sea Urchin Primary Invagination. Cells that will eventually be the tip of the archenteron begin primary invagination by changing shape by way of apical constriction (B). F-actin staining (A) localizes to the apical end of bottle cells, indicating that a similar cytoskeletal mechanism drives shape change in urchin as in other species. (Kimberly and Hardin 1998)
2.1.1.3 EMT and Apical Constriction

Apical constriction during EMT is required for more than shape change and force generation. The constriction of many cells brought closely together at the apical side of the epithelium is a convenient mechanism for “sealing” the hole in epithelium left behind when nascent mesenchymal cells eventually make their exit (Baum, Settleman, and Quinlan 2008). Another unique role for apical constriction in EMT is patterning. Mechanical stresses determine the position and direction that EMT will occur. In tissue culture, epithelial cells forced to grow in different shapes were transformed with TGFβ, a potent inducer of EMT in cultured cells. As seen in Figure 2.3, the cells displayed increased mesenchymal-like molecular phenotypes in the regions of shapes where mechanical stress was the highest (Gomez et al. 2010).
Figure 2.3: EMT is Directed by Mechanical Stresses. (B,E) Cells treated with TGFβ show the frequency pattern of αSMA, a mesenchymal marker. (C,F) Levels of endogenous stress points within two shapes the epithelial sheets were plated on. Mesenchymal markers are expressed higher in regions where mechanical stress is higher. Adapted from (Gomez et al. 2010).
2.1.2 Cell Polarity

2.1.2.1 Epithelial vs. Mesenchymal Polarization of Cells

Epithelial cells establish polarity via specialized polarity complexes (See Figure 2.4). The Par complex and Crumbs complex localize and define the apical identity of the epithelium. The Scribble complex localizes baso-laterally and defines the basal end. These complexes are well conserved throughout the animal kingdom and antagonize one another to maintain polarization (Assemat et al. 2008)

![Figure 2.4: Epithelial Cell Polarity](image)

**Figure 2.4: Epithelial Cell Polarity.** The apical side of the cell is functionally determined by the position of neighboring cells and localization of PAR and Crumbs complexes to the apical membrane at the adherens junctions. PAR and Crumbs promote apical identity and inhibit the Scribbles complex to suppress basal identity. The Scribbles complex antagonizes PAR and Crumbs to promote basal identity at the opposite end of the cell.
Mesenchymal cell polarity is drastically different (See Figure 2.5). The apical and basal complexes are absent. Polarity in mesenchymal cells is not pre-defined the way it is for epithelial cells. Instead, mesenchymal polarity is generally marked by the leading edge of a cell that is migrating by small GTPases, like CDC42, that actively remodel the cytoskeleton (Huang and Muthuswamy 2010). In epithelial cells, polarity is determined by their relative position within the tissue and the mechanism of establishing polarity is cued by their immediate surroundings. Mesenchymal polarity is defined internally, though often the direction of motility is in response to external signals.

![Figure 2.5: Mesenchymal Cell Polarity](image)

**Figure 2.5: Mesenchymal Cell Polarity.** The polarity of a mesenchymal cell is determined internally by its direction of motility. The leading edge is defined by lamellipodia and other protrusions where the actin-cytoskeleton is being polymerized. The trailing edge is characterized by actin stress fibers which will retract as the cell moves forward.
2.1.2.2 Cell Polarity in Sea Urchin Embryos

Functional apical-basal polarity of the urchin embryo is established very early in development, immediately following the first cleavage when Par6 can be seen localized to the apical membrane (Alford, Ng, and Burgess 2009). Apical-basal polarity persists through later stages of development. During swimming blastula stages the monoclonal antibody 1c10, which recognizes an apical antigen, is localized to the apical ends of epithelial cells. By mesenchyme blastula stage, 1c10 continues to localize to the apical side of the epithelium but is cleared from PMCs (Nelson and McClay 1988). Taken together these data show that epithelial polarity is established prior to micromere birth and is maintained in the epithelium throughout development but is lost by mesenchyme. Therefore, in the context of PMC ingression, any polarity phenotypes controlled by the PMC GRN discovered by the time-lapse assay must be described as a loss of previously established polarity since polarity is established before the start of zygotic transcription.
Figure 2.6: Polarity of the Sea Urchin is Established by 2-cell Stage. Par 6 (a’) and Cdc42 (b’) are localized to the apical membrane following the first cleavage. (Alford, Ng, and Burgess 2009)

2.1.2.3 EMT and Cell Polarity

Apical-basal polarity is a key feature in coordination of EMT events. If a cell fails to establish the proper polarity relative to its position within the epithelium, then it will be unable to execute all subsequent EMT events which must occur in specific regions of the cell that are defined by their apical or basal positions. On the other hand, loss of previously established basal identity can directly lead to EMT. When the Crumbs complex is repressed in the embryo D. melanogaster cadherin is removed from adherens junctions circumventing the downregulation of cadherin transcription (Campbell et al. 2011).
Shape change in response to apical constriction is dependent upon an established apical-basal polarity. Basement membrane invasion is a polarization dependent event because it requires remodeling of the basal lamina, which is restricted to basal side of the cell. De-adhesion from the epithelium requires disassembly of adherens junctions which are apically distributed, although if polarization was never established in the first place then adherens junctions cannot be properly formed in the first place and so there would be very little de-adhesion required. Finally, if a cell is lacking all polarity or is unable to switch from an epithelial to mesenchymal polarity, then it will not be able to establish directed motility, may only be able to move with random motion, and would not be capable of migrating to its final destination in the embryo (Micalizzi, Farabaugh, and Ford 2010).

2.1.3 Cell Motility

2.1.3.1 Mesenchymal Cells Acquire Motility

The cell biology of motility is highly conserved across all metazoans and has been well studied. The leading edge of the cell protrudes in the direction of cell movement, the cell “grabs” on to the substrate and retracts the trailing edge. The cycle repeats ad infinitum for as long as the cell is in motion. However, the regulation of directed motility in response to migratory cues is still a long way from being sorted out (Kurosaka and Kashina 2008).
Figure 2.7: Random Motility vs. Directed Motility. The direction and displacement of movement over time tells whether a cell is moving randomly or directionally down a signaling gradient.

When considering the case of motility the distinction between directional and random movement is also important. A key family of molecules that govern motility in a cell are members of the Rho family of GTPases, which are responsible for remodeling cytoskeletal elements. The leading edge of the motile cell is defined mainly by lamellipodia protrusions. Whether a cell’s motion is persistent in one direction or random is primarily dependent upon the stability or transience of the leading edge lamellipodia (Petrie, Doyle, and Yamada 2009).
2.1.3.2 Motility in Sea Urchin Embryos

When the *S. purpuratus* genome was published in 2006, the annotation of genes encoding cytoskeletal and motility proteins did not turn up any surprises. Overall, the phylogenetic comparisons show a high similarity between urchins and vertebrates with the exception of a few vertebrate specific genes missing urchin homologues (Morris et al. 2006). The underlying cell biological mechanism of motility in PMCs has not been characterized well in urchin development (if one agrees to not include the study of sperm motility under the category of development). However, one study does describe lamellipodia at the leading edge of the archenteron as it elongates toward the animal pole (Hardin 1989). Considering that the urchin has the expected array of cytoskeletal and motility elements and high conservation of motility mechanisms exists across phyla and there is evidence of directional lamellipodia driven movements, a conventional mechanism for motility in the urchin embryo is a likely bet.
Figure 2.8: Urchin Lamellipodia at the Tip of the Archenteron. Basal lamellipodia form at the tip of the elongating archenteron as it moves toward the animal pole. (Hardin 1989)

2.1.3.3 EMT and Motility

The best-characterized example of a developmental EMT event concurrent with motility is the case of neural crest cells. The neural crest undergo EMT, delaminating from the neural plate border as the neural tube closes, and migrate long distances within the embryo to their destinations. The acquisition of motility has been observed even before full de-adhesion is completed, leaving behind cell pieces in the epithelium. Neural crest also exhibits signs of motility in the form of lamellipodia and blebs prior to exiting the epithelium (Ahlstrom and Erickson 2009; Berndt et al. 2008).
Figure 2.9: Neural Crest Motility During EMT. Chick neural crest in vivo time-lapse shows a neural crest cell as it stretches away from the neural tube, becomes motile, the apical ‘tail’ retracts, and the neural crest is free to migrate to its final destination. (Ahlstrom and Erickson 2009)

2.1.4 Adhesion and De-Adhesion

2.1.4.1 Mechanics of Cellular Adhesion

The physical interactions responsible for maintaining epithelial sheet integrity by connecting individual cells together are called cell-cell adhesions. The majority of contact between cells can be explained in terms of cadherins and their association at apical adherens junctions (See Figure 2.10). Cadherins are the primary building blocks of the adherens junctions. They are a transmembrane proteins with the ability to form homodimers in between neighboring cells “sticking” them together. Catenin molecules stabilize the adherens junction by linking the cadherin molecules to the actin cytoskeleton. Adherens junctions are constitutively maintained by trafficking Cadherin molecules to and from the cell membrane. (Yap, Crampton, and Hardin 2007).
Figure 2.10: Basic Structure of an Adherens Junction. Cell membranes adhere closely together at the adherens junctions. Cadherin is a transmembrane adhesion molecule that homodimerizes with cadherin of the neighboring cell. The junction is formed when cadherins are anchored to the actin cytoskeleton by catenins.

2.1.4.2 De-Adhesion in Sea Urchin Embryos

In the sea urchins, cadherin, β-catenin and α-catenin accumulate together and co-localizing at the apical side of the cell to the site of adherens junctions. At the time of primary cell ingression, adherens junctions are disassembled at the vegetal plate and cadherin/catenin expression is cleared (Miller and McClay 1997b, 1997a).

The sea urchin uses a conserved mechanism for de-adhesion and transcriptional regulation of the mechanism has also been at least partially solved. Snail repressed the transcription of cadherin and up-regulated the rate at which cadherin was removed from the cell membrane by way of endocytic trafficking. The removal of
cadherin from the cell surface and mechanisms to de-stabilize the adherens junctions are the causal event of de-adhesion in PMC ingestion. Transcription of Snail and its repression of cadherin begins too late to initiate de-adhesion, as there is still cadherin on the surface of PMCs as they ingress (See Figure 1.3). Instead, the transcriptional repression appears to be a mechanism to prevent adherens junctions from re-assembling. The transcription factor Twist was also essential to ingestion. Studies showed that Twist and Snail share an upstream regulator, Alx1 (Wu and McClay 2007; Wu, Yang, and McClay 2008) and they appear to regulate de-adhesion in parallel to one another (See Figure 2.11).
Figure 2.11: **Snail and Twist are Required for EMT.** (A-C) Control embryos are full of ingressed PMCs at MB stage and development continues normally. (D-F) Snail MASO injected embryos are stalled with no PMC ingestion and empty blastocoels. (G-I) Twist MASO causes a similar phenotype to Snail, but embryos eventually begin to recover, though development lags behind controls. (Wu and McClay 2007; Wu, Yang, and McClay 2008)

### 2.1.4.3 EMT and De-Adhesion

In the case of EMT, de-adhesion mediated by endocytosis and transcriptional repression of cadherin has been exhaustively studied, especially in cancer type III EMT. The body of research encompassing all EMT knowledge to date is overwhelmed by the numerous detailed descriptions of cadherin regulation (Delva and Kowalczyk 2009).
In particular, the regulation of de-adhesion leading to EMT is of great significance in various different cancers. Different cancers progress differently and cadherin regulation can be an indicator of how well a patient will respond to drug therapy (Krasnapolski, Todaro, and de Kier Joffe 2011). However, the central dogma of de-adhesion from urchin to cancer always uses the same motif: Snail (See Figure 2.12) and its family members repress cadherin and promote its degradation through an endocytic pathway, thereby promoting EMT (Georgiou et al. 2008; Barrallo-Gimeno and Nieto 2005; Baranwal and Alahari 2009).

Figure 2.12: Many EMT Regulators Converge onto Snail. Many different signaling mechanisms across many different EMT examples ultimately result in the transcription of Snail, an important transcription factor for de-adhesion during EMT. (Barrallo-Gimeno and Nieto 2005)
2.1.5 Invasion

2.1.5.1 Basement Membrane Structure and Function

The basement membrane is an extracellular matrix primarily composed of laminin and collagen (Yurchenco, Cheng, and Colognato 1992). This dense meshwork of macromolecules is held together by self cross-linkages and provides a barrier between the epithelium and the mesenchymal tissues (Yurchenco and Patton 2009). The adhesion between the basal side of cells and the basement membrane is primarily facilitated by integrins. The basement membrane is actively modified and reconstructed during development, so it should not be thought of as merely a passive cell substrate or simple barrier (Rowe and Weiss 2008). In development, the basement membrane promotes polarity in the cells that attach to it and aid in re-organization of tissue layers in the developing embryo. As a barrier it prevents cells from leaving the epithelium by accident (Yurchenco and Patton 2009; Yurchenco, Amenta, and Patton 2004). The intentional breach of the basement membrane when a cell undergoes EMT is referred to as invasion. It should be noted that passage through the basal lamina is a normal function for many cell types not involved in an EMT, especially immune cells that actively move through the basal lamina, but most likely by a distinct mechanism. Immune cells have not been observed to remodel the basement membrane and leave no visible migratory path as they navigate through the meshwork (Huber and Weiss 1989; Rowe and Weiss 2008).
2.1.5.2 Invasion in Sea Urchin Embryos

A very limited amount of research on the basement membrane of sea urchin shows that laminin, collagen, and integrin are all present. However, the timing and distribution of laminin staining during EMT has not yet been determined laminin (Benson et al. 1999; McCarthy, Beck, and Burger 1987; Hertzler and McClay 1999).

Figure 2.13: Laminin Stains the Basement Membrane of Urchins Embryos. Staining with an antibody specific to laminin (right) is a thin layer coating the inside of the blastocoel localized to the basal lamina. (Benson et al. 1999)

PMCs successfully cross the basal lamina during invasion, although the mechanism for how this occurs in urchin had not been studied. The prevailing assumption was that the matrix was loose enough that PMCs could physically push it out of the way or squeeze through it (Wu, Ferkowicz, and McClay 2007). We have addressed this gap in knowledge within the scope of this study (See Chapter 3.3 Results).
2.1.5.3 EMT and Invasion

Basement membrane invasion is the fifth and final feature of EMT addressed by this study. In a developmental EMT, invasion is a required step for exiting the epithelium and in cancer EMT the remodeling of basement membrane components can actually induce invasion resulting in metastasis (Larsen et al. 2006).

**Figure 2.14: Anchor Cell Invasion Clears a Path Through Laminin.** The anchor cell (green) of the *C. elegans* embryo invades the basement membrane leaving a large hole in laminin staining (purple) demonstrated using *in vivo* time-lapse microscopy in two-dimensional (top) and three-dimensional (bottom) views. (Hagedorn and Sherwood 2011)

Invasion can proceed in three ways, proteolytic dependent degradation of the basement membrane, proteolytic-independent invasion by squeezing or pushing through pores in non-cross linked meshwork, or a combination of both. The best understood method for invasion is by proteolytic degradation by matrix
metalloproteases. While there is evidence that cells can invade in absence of proteases during an EMT, it is a phenomenon that has only been observed in vitro (Rowe and Weiss 2008).

2.2 Experimental Design

In this study, we engaged a combination of systems level analysis techniques alongside classical embryology. Embryo manipulations were used to trace the progeny of the micromeres, the PMC precursors, during development through the EMT. High-resolution fluorescent microscopy detailed phenotypic cellular changes, to create a fully integrated picture of how five separate cell biological events are coordinated with precision and stringency by the embryo to complete a single fluid act of EMT. The data were analyzed quantitatively and objectively using unsupervised statistical analyses. We use the sea urchin embryo as the model system in our approach because it is perfectly set up for answering this challenging question of how complex EMT change is regulated. The regular sequences of events that underlie urchin PMC ingression, the developmental EMT of interest to this study, have been well documented and in tandem with the established sea urchin GRN, the urchin model system provides a solid foundation of knowledge upon which an integrated explanation of how a individual cell biological behaviors are orchestrated in a developmental EMT from a novel system wide vantage point.
We developed an assay to capture real-time phenotypic change during PMC ingression \textit{in vivo} using time-lapse imaging caused by altering network states in a systematic way to approach an integrated view of EMT event coordination. That assay required us to overcome many technical challenges including the cilia on the embryo that were necessary to stop for the time-lapse imaging, and design of a chamber in which the embryos would thrive for hours while imaging. We then designed an approach that allowed us to alter GRN states in a systematic way. We knocked down each PMC GRN transcription factor, transplanted one fluorescently labeled micromere (PMC precursor) with the knocked down network component, and transplanted it to a control host. The transplanted cell was followed with fluorescence and its progeny were monitored over the course of ingression.

The output of the assay was image data recording real-time cellular phenotype changes of individual EMT behaviors in cells running different network states. We excluded pmar1 from the study because it is a global specifier of PMC fate and phenotypic analysis of EMT failure from perturbing this gene would be misleading, as a cell that is not fated to become a PMC would never begin the cellular EMT events during ingression that we wished to observe. In total, we performed this assay using 12 knockdown perturbations in addition to unperturbed controls. Figure 2.15 visually summarizes the assay and a detailed description of the techniques employed follows.

Fertilized eggs were triple injected with two fluorescent markers (RFP and GFP) and one
experimental variable. The first fluorescent marker, membrane-bound GFP (mem-GFP), outlined the cell’s shape by incorporating into all phosho-lipid membranes and the second fluorescent marker, histone-H2B RFP (H2B-RFP), localized to the nucleosomes which package DNA together to highlighted each nucleus. The experimental component of the injectate was either the standard control morpholino antisense oligonucleotide (MASO) or a knockdown to one of the transcription factors in the PMC gene regulatory network (GRN): LvAlx1MASO, Dominant-Negative LvEts1 mRNA (DN-LvEts1), LvTbr MASO, LvTel MASO, LvErg MASO, LvHex MASO, LvTgif MASO, LvSnail MASO, LvTwist MASO, LvFoxN2/3 MASO, LvFoxB MASO, or LvFoxO MASO.

At 2.5 hours post fertilization (hpf) injected embryos, after the first asymmetric cleavage had given rise to the four micromeres fated to become PMCs, one micromere from the injected embryo was transplanted via microsurgery onto an un-injected host embryo replacing one removed host micromere. This provided a non-fluorescent background in which to visualize the fluorescently labeled micromere progeny as development progressed. The three control micromeres that remained in the host embryo served as internal controls for PMC ingresson.

At 8.5 hpf, about a half hour prior to the beginning of ingestion, hatched blastula stage embryos were de-mobilized and mounted on slides in a manner which ensured sufficient oxygen exchange but not liquid evaporation over several hours. Embryos were imaged using on one of two live-imaging microscope systems at regular
intervals beginning at 9.0 hpf, the very first sign of control cell ingestion, and up to 15.0 hpf, long after control cells had completed ingestion. Fluorescent and DIC images were collected at each time-point and for experiments where embryos remained in frame for at least one-hour past the completion of control cell ingestion, or about two hours, images were processed using image processing software and assembled into movie projections.

**Figure 2.15: Micromere Transplant Assay.** At 2.5 hpf, one micromere is removed from the donor embryo (lower left) which has been injected with a transcription factor perturbation and two fluorescent markers. One micromere is removed from the unperturbed host (upper left) embryo and discarded. The fluorescent experimental micromere is transplanted onto the host (middle), which provides a controlled environment for viewing the EMT phenotype 6 hours post surgery with fluorescent microscopy (right).
2.3 Methods

2.3.1 Animals

Sexually mature *Lytechinus variegatus* adult sea urchins were obtained from Reeftopia (Key West, FL), Sea Life Inc. (Tavernier, FL), the McClay lab and/or Dr. Maria Wise of Duke University Marine Lab (Beaufort, NC), or Anne Leaser (Wilmington, NC). Gametes were collected by injecting 0.5M KCl directly into the coelomic cavity. Embryos were cultured at 23°C or room temperature in artificial seawater.

2.3.2 Cloning *Lv* PMC Transcription Factors

Nested primer sets were designed to the annotated *S. purpuratus* coding sequences for cloning the following transcription factors in the species *L. variegatus*: *LvDri*, *LvErg*, *LvFoxB*, *LvFoxO*, *LvHex*, *LvTel*, and *LvTgif*. Low annealing temperatures (~50°C) were used with the outer primer sets to generate PCR amplicons from *L. variegatus* mesenchyme blastula cDNA preparations; the resulting PCR products were used as templates for subsequent rounds of amplification using inner primer sets and a more stringent annealing temperature (58°C-62°C). PCR products were either gel extracted or PCR purified and ligated into the pGEMT-Easy Vector and sequenced using T7 and SP6 primers. Full-length coding sequences with 5’ and 3’ UTRs were obtained for each transcription factor using the FirstChoice RLM-RACE kit by Ambion. Full-length clones were isolated from mesenchyme blastula stage cDNA using PCR primer sets designed with restriction enzyme sites against the *Lv* UTR sequences to span only
the ORF and purified PCR produced were cloned into the pGEMT-Easy Vector for synthesis of anti-sense RNA probes and ORFs were confirmed by sequencing; restriction enzyme digests of clones at the restriction sites added during PCR amplification were used to directionally clone ORFs into the expression vector PCS2+. When the \textit{L. variegatus} genome assembly became available in 2011, all 7 cloned transcription factor sequences were authenticated.

### 2.3.3 Transcription Factor Knock-Downs and Fluorescent Markers

Two non-overlapping morpholino antisense oligonucleotides (MASOs) targeted to each one of the 7 cloned PMC transcription factors were designed and synthesized by Gene-Tools. The effective concentrations were determined:

- \(2.0\text{mM} \text{LvDri1:} \quad (5' \rightarrow GTAAAGTCTACAGACATTCGTTTGC \rightarrow 3')\)
- \(1.5\text{mM} \text{LvDri2:} \quad (5' \rightarrow CGCGGTGGTTCAACCCGAAAACCGAA \rightarrow 3')\)
- \(0.75\text{mM} \text{LvErg1:} \quad (5' \rightarrow AATAATCACAATCGATACGACGT \rightarrow 3')\)
- \(0.5\text{mM} \text{LvErg2:} \quad (5' \rightarrow GCTTTACTCAAAGGCACGGATA \rightarrow 3')\)
- \(0.3\text{mM} \text{LvFoxB1:} \quad (5' \rightarrow CTGGTATTTTCACAGAAAGTCATGC \rightarrow 3')\)
- \(0.5\text{mM} \text{LvFoxB2:} \quad (5' \rightarrow CCCCTCTAAATAGATACATGGTC \rightarrow 3')\)
- \(1.0\text{mM} \text{LvFoxO1:} \quad (5' \rightarrow GGTCGTTATCAACCATTGGATGA \rightarrow 3')\)
- \(0.75\text{mM} \text{LvFoxO2:} \quad (5' \rightarrow ATCCTAAATGGTCACAAGTACAC \rightarrow 3')\)
- \(2.0\text{mM} \text{LvHex1:} \quad (5' \rightarrow TGCACGAAAGATGATCCAGAAATGC \rightarrow 3')\)
- \(1.5\text{mM} \text{LvHex2:} \quad (5' \rightarrow GGTGCTGAACCTTACACAACACTCT \rightarrow 3')\)
0.75mM LvTel1 _____ (5’ – ATCGTTCTTGTCCTGGGCTACGTTCC – 3’)
1.0mM LvTel2 _____ (5’ – ATGGTTCCTGTCTTGTAACCTGA – 3’)
0.5mM LvTgif1 _____ (5’ – ATCTTTCTTTGATAATCCGCATC – 3’)
1.0mM LvTgif2 _____ (5’ – CGTATGGTGGTTTTTGCAGTGTT – 3’)

The following concentrations were used for previously published Lv PMC transcription factor MASOs utilized in this study: 1.0mM LvAlx1 (Ettensohn et al. 2003), 1.0mM LvSnail2 (Wu and McClay 2007), 1.5mM LvTwist (Wu, Yang, and McClay 2008), 0.5mM LvTbr (Croce et al. 2001), 0.7mM FoxN2/3-1 (Rho and McClay 2011). mRNA for injection was transcribed in vitro using mMessage mMachne SP6 or T7 kits by Ambion. The following final concentrations were used for mRNA injections: 750 ng/μL Dominant-Negative LvEts1 (DNEts1) (Sharma and Ettensohn 2010), 250ng/μL Histone2B-GFP (H2B-GFP), 500ng/μL H2B-RFP, 750ng/μL membrane-RFP (mem-RFP), 500ng/μL mem-GFP.

2.3.4 Injections and Micromere-Transplantations

Embryos were fertilized in 0.5% paramino-benzoic acid in artificial seawater and triple injected with either 500 ng/ul of mem-GFP and 500 ng/μL of H2B-RFP or 750 ng/μL of mem-RFP and 250 ng/μL of H2B-GFP and either 0.5mM of standard control MASO or one of the previously described concentrations of PMC transcription factor knockdowns. Embryos were cultured at 23°C for 2.5 hours. Micromeres were transplanted at the 16-cell stage; one micromere from an injected donor embryo was inserted onto an un-
injected 16-cell host embryo from the same parental cross in the place of one discarded host micromere. Detailed methods of injections and transplants were followed as previously described (Sherwood and McClay, 1999, Logan et al. 1999).

2.3.5 In Vivo Time-Lapse Imaging

At 8.5 hpf embryos were de-ciliated in a 10.0 second wash with 2x hypertonic artificial seawater. Immobilized embryos were mounted on a slide coated in 2% protamine sulfate under a clay footed coverslip coated in 2% protamine sulfate in 30uL of artificial seawater containing a 10uM final concentration of synthetic parsley seed derived compound, which inhibited ciliary motion for over 12 hours with no delay in development (Semenova, 2007) and coverslips were pressed down until fluid covered 75% of the coverslip area and left air over 25% of the area and were subsequently sealed with V.A.L.A.P.

Embryos were imaged using one of two live imaging microscopy systems both with motorized stages: either the Zeiss Axio Observer Z1 or the DeltaVision Elite Deconvolution microscope. Images were taken on the Zeiss microscope using a 63x/1.4oil Plan Apochromat DIC objective with the Coolsnap ES2 high-resolution CCD camera. RFP channel, GFP channel, and DIC images were each collected for up to five embryos per experiment at 60-second intervals beginning at 9.0 hpf and ending between 12.0 hpf and 15 hpf. Images were taken on the DeltaVision microscope using either the 40x/0.65-1.35 oil UAPO40X0I3/340 DIC objective or the 60x/1.42 oil PLAPON60XO DIC
objective with the Coolsnap HQ2 high-resolution CCD camera. The Dual em pass GFP/mcherry high speed dichroic was used to simultaneously capture both RFP and GFP channels along with a reference DIC image for up to 10 embryos at 60-second intervals through 5 z-sections 5µm apart beginning at 9.0 hpf and ending no earlier than 12.0 hpf. Time-lapse images collected were made into movie projections using Metamorph image processing software for Zeiss data or SoftWorx image processing software for DeltaVision data.

2.3.6 Data Handling

In total, 389 out of the 500+ movies that were processed contained sufficient data for further analysis. Raw data for all experiments and processed data for 389 experiments is archived in triplicate on multi-terabyte external hard drives formatted with three different file systems (NTFS, FAT-32, ext3) and stored in two buildings 5 miles apart to ensure data security.

2.4 Results

Movies where embryos hosted a standard control transplant micromere showed a typical pattern of EMT events; the intermediate shape phenotypes were seen as PMCs lost their epithelial shape and became an teardrop shaped when the apical end of the cell constricted and the basal end bulged out into the blastocoel. Directed motility towards the blastocoel stretched the cells away from their adherens junctions and long, thin apical tails were observed. As they entered the blastocoel the apical tails contracted into
the cell body; cells became more rounded, showed no discernable apical-basal polarity and they migrated inside the blastocoel. After entry into the blastocoel PMCs divided once and proceeded to move about freely.

Movies of embryos containing a transplant with perturbed network state demonstrated a variety of phenotypes that differed from the typical EMT phenotype reported in control transplants. Some knockdown groups exhibited a phenotype that appeared halted at juncture where they should have switched their adhesion profile. Alx1, Twist, and Snail knockdowns each made an exaggerated upside teardrop shape and appeared to be tethered to the epithelium via AJs by abnormally long, thin, apical stalks. Ets1, Tbr, Hex, Tgif, Erg, Tel, FoxB, FoxO, and FoxN2/3 knockdowns all failed to make the teardrop shape at all; the FoxN2/3 knockdown looked the most epithelial while others made less descript shapes. The knockdowns Alx1, Twist, Snail, FoxB, Tel, and Tgif showed transplants that bulged into the blastocoel over time but never made it fully inside; Ets1, FoxN2/3, FoxO, and Hex were examples of transplants that remained almost entirely within the epithelium for the duration of the movie and Tbr and Erg represented an intermediate phenotype. Figure 2.16 shows still frames from movie examples for each gene and describes the observed phenotype as determined by simple observation and visual classifications. Our human visual system is highly sophisticated at pattern recognition more so than any current computer based technology and discrete phenotypes were clearly discernable: the problem with this phenotypic classification
method however, is that while it provided strong anecdotal evidence in support of discrete EMT behavior phenotypes, it was subjective and not quantifiable. We therefore sought, rigorous analytical approaches using systems biology methods to objectively and quantitatively determine phenotypic classifications of each transcription factor using the perturbation data before any transcription factor could be assigned with accuracy to a subcircuit regulating a singular cellular EMT behavior.
Figure 2.16: Still Frames from Time-Lapse Movies Show Distinct Phenotypes.
Still frames over a 90 minute time-lapse show EMT phenotypes for two control embryos and 12 PMC transcription factor perturbations. PMCs containing the perturbation are labeled with mem-GFP and H2B-RFP, embryo manipulations were performed as previously described in Figure 2.15. Only control transplants have PMCs in the blastocoel with no attachments to the epithelium. Different knockdowns show clearly distinct phenotypes where EMT has failed in different ways. The most extreme examples are shown by Alx1 and FoxN2/3. Alx1 appears to make a long apical stalk like a control cell and the cell body is pushed into the blastocoel. The Alx1 movie shows cells actively ‘pulling’ away from their attachments to the epithelium which demonstrates a latent motility. FoxN2/3 also fails to ingress but appears to not complete any of the required cellular EMT behaviors that can be seen using this assay: no apical constriction, irregular shape changes, and no motility or active pulling. Most other transcription factor knockdowns resulted in phenotypes that fell somewhere in between these two extreme examples. While the time-lapse movies themselves told us something about why EMT failed under an individual perturbed PMC GRN state, a more objective statistical analysis of these phenotypes was used to accurately classify each cellular EMT behavior. Refer to the Appendix for the original time-lapse movies summarized in this figure.
2.5 Statistical Analyses

2.5.1 Analyzing Shapes

Every EMT event is characterized by a striking change in cell shape as epithelial cells become mesenchymal. The intermediate inverted teardrop shape distinguishes those cells that are still in the process of transitioning from those that have not yet started or have already finished an EMT. The time-lapse data collected contained cell boundary information delineated by membrane-bound GFP fluorescence. Individual cell boundaries of the transplanted cells were traced and 2D outline data was subjected to a principal component analysis using SHAPE ver.1.3. Outlines were normalized along the apical-basal axis and elliptical Fourier transforms principal component analysis was performed on the data consisting of 389 movies with 4 cells traced per movie. The 4 cells traced for shape analysis were chosen based on two criteria: (1) to remove the possibility of scoring small micromere descendants, which do not undergo EMT, the two smallest cells were never chosen for tracing and (2) each chosen cell was in focus throughout the entirety of the movie and represented the 4 least obscured i.e. fewest overlapping neighbors.

2.5.1.1 Principal Component Analysis

Statistical analysis of the cell shapes contained within our data set was necessary in order to empirically determine which transcription factors were involved in regulating the network states responsible for normal shape changes created during EMT. By
defining phenotypes based on resultant shapes, apical constriction and apical-basal polarity defects could both be explained using a single analytical method.

Unfortunately, images collected by time-lapse in vivo fluorescent microscopy contained cell shape information, as determined by localization of mem-GFP fluorescence, which was not suited for automated shape analyses. Even with nuclei staining available for watershed methods, automated software could not accurately predict cell boundaries because too many boundaries crossed over one another. See Figure 2.17. Traditional biological methods of shape analysis used in the field of morphometrics use landmark data points, such as anatomical features, contained within the shapes to infer both size and shape variation. The cell shape data contains only one landmark, the nuclei, and so it was not suited for landmark data analysis methods.

**Figure 2.17: Examples of Cell Tracings.** Cell shapes could not be analyzed by automated methods with any accuracy because of crossovers and occasional out of focus issues, and therefore were traced manually.
The statistical shape analysis that was best suited to describe variable shape phenotypes in the dataset was Principal Component Analysis (PCA). PCA determines shape variability by reducing outline data into low dimensional vector space and determining which component of that space describes the highest variability within the shape data. PCA is performed without regard to the size or orientation of the objects analyzed. Removing the size and orientation from the data analysis was a benefit in this case. Variation in both size and orientation of cells was caused by the relative size of the eggs shed from the mother and the rotation of the embryo on the slide and was independent of network perturbations. Only pure shape variation would be provide useful information about the phenotypic control of cell shapes by PMC network transcription factors. PCA provided an unbiased statistical analysis for understanding phenotypic changes resulting purely from shape differences.

2.5.1.2 Methods

Cell membranes were traced using Adobe Photoshop CS5.1 and the resulting edited outline data was analyzed using the program SHAPE ver. 1.3 (Iwata and Ukai, 2002). Outlines were elliptical Fourier transformed to the first 20 harmonics. Transforms were normalized along the apical-basal axis of each cell and subjected to a principal component analysis.
2.5.1.3 Results

The first three principal components explain 86% of the total shape variation and the first five principal components explain 92% of the total shape variation. Each principal component corresponds to one shape descriptor and those with the highest variation represent the most descriptive and informative characteristics to consider when describing phenotypic shape changes.

The mean shape of the principal component analysis was the upside down teardrop that is characteristic of a cell in the midst of transformation. The first three principal components showed meaningful shape differences. For each 12-minute interval in a two-hour time window, principal component values were averaged within each experimental group and the control group.

The first principal component, the principal component that showed the highest variability in the data set, separated cells elongated by apical constriction from rounded mesenchymal cells. Three groups sorted toward the end of the first principal component’s continuum that represented apical elongation: control, LvAlx1 MASO, and LvSnail MASO. The Alx1 and Snail cell shapes remained the same over time. The control cell shapes were dynamic; in the first 30 minutes following the start of ingression, the controls rapidly rose toward the opposite end of the first principal component’s continuum, which represented a rounded shape. The control group averaged a rounded value for the duration of the time-lapse.
The first 4 principal components explain 86% of the total shape variation. The mean shape is a teardrop. The odd numbered components explain shape variation along the apical-basal axis and the even numbered components explain shape variation along the lateral axis. Only the apical-basal axis was informative for this study and so PC1 and PC3 were considered when classifying phenotypes. PC1 describes shapes that are strongly apically constricted through round mesenchymal shapes. PC3 describes the polarity of shapes where -2 StdDev has cytoplasm distributed more apically and +2 StdDev has cytoplasm distributed more apically.
Figure 2.19: Principal Component One. PC1 distinguishes between cells that are apically constricted with a long apical stalk from round mesenchymal like cells. Alx1 and Snail are significantly more apically constricted than other perturbations while control cells become more mesenchymal as time progresses.

The second and third principal components separated cells by lateral and apical-basal asymmetry respectively. At the time of ingression, the control PMCs are asymmetrical from animal to vegetal pole but left-right asymmetry is not morphologically distinguishable. The absolute values of the second principal component were averaged to represent a spectrum from laterally symmetrical shapes to laterally asymmetrical shapes; taking the absolute value removed left-right directionality of the asymmetry leaving only measures of symmetry versus asymmetry. Only the control group was distinguishable using the second principal component; the controls
sorted to the laterally symmetric end of the spectrum and the experimental groups sorted to the laterally asymmetric end (See Figure 2.20). The third principal component, apical-basal polarity, separated the groups between cells with mass distributed more apically versus more basally (See Figure 2.21). Control group values initially sort to the basally heavy polarity end of the spectrum and shift during the course of the time-lapse to a value indicating equal apical-basal mass distribution. The Twist and Alx1 groups were sorted as basally heavy toward the initial values of control cells and changed little over the time-course. The Snail group began in the middle of the spectrum at the mean shape where slightly more mass is distributed basally than apically; over time, Snail sorted more basally heavy with Twist and Alx1. The FoxN2/3 group sorted toward the opposite end of the spectrum from Twist, Alx1, and Snail where values correspond to a slightly more apical distribution of mass. Over the time-course FoxN2/3 gradually sorted toward the control group’s ending values with equal apical-basal mass distribution. Other experimental groups showed no significant apical-basal polarity changes.
Figure 2.20: Principal Component Two. The absolute value of PC2 was taken to remove left v. right and visualize only symmetrical or asymmetrical. The control cells are the most laterally symmetric, however, this was not a meaningful shape descriptor for classifying transcription factor phenotypes.
Figure 2.21: Principal Component Three. PC3 shows asymmetry of cytoplasm distribution along the apical-basal axis. Twist is significantly higher than others, indicating a slight amount of apical constriction with too short a stalk to be significant in PC1. FoxN2/3 stands out as having more cytoplasm at the apical end and less basally, the exact opposite of the mean, indicative of an AB-Polarity problem.
The results describe two EMT subcircuits, each regulating a different aspect of shape change. Knockdown of a large subset of the transcription factors results in loss of apical constriction: Tbr, Ets1, Erg, Hex, Tgif, Tel, FoxB, and FoxO. However, additional experiments would be needed to determine whether each transcription factor participates directly in control of apical constriction or whether the upstream transcription factors are only involved via epistatic interactions. For these reasons, Tel, FoxB, and FoxO are drawn in the subcircuit as the most distal, and likely most direct inputs into apical constriction. Fox N2/3 alone appears to regulate the subcircuit driving apical-basal polarity.

2.5.2 Analyzing Distribution of Cells Across Basal Lamina

2.5.2.1 Inferring Motility in a Tethered System

Nascent mesenchymal cells must be motile in order to completely leave the epithelium and enter the blastocoel; once fully inside, PMCs migrate outward from their entry point at the center of the vegetal plate towards an epithelial ectodermal signal. Ultimately they arrange themselves into a ring around the inside of the blastocoel before fusing into a syncytium and becoming immotile again.

The time-lapse assay confirms that all control PMCs move in a straight line across the basal lamina toward the center of the blastocoel over the course of ingression to complete an EMT (See Figure 2.22). Upon entry into the blastocoelar space, the control cells were observed to move in various directions inside the blastocoel. Earlier
experiments showed that until PMCs locate their targets the net direction of movement is outward and in all directions away from the site of ingression. Measuring the motility of control cells is relatively simple. Using the fluorescently labeled nuclei, there are a number of software packages that are able to track each nucleus and its displacement over a time course. However, measuring the motility of knockdowns was more challenging.

Figure 2.22: Control PMCs Ingress Completely into the Blastocoel. (Left) Control PMCs at 9.0 hpf prepare for ingression and bulge into the blastocoel (dotted line). (Right) 10.0 hpf the control PMCs are entirely inside the blastocoel. Note: two cells left behind within the epithelium after ingression are descendants of the small micromere lineage and not PMCs.

The question remained open as to why each knockdown was unable to move. By eye, the knockdowns Alx1, Twist, and Snail appeared to be actively pulling away from their adherens junctions towards the blastocoel; they appear to be unable to move any
distance because they are tightly tethered to the epithelium. On the other hand, knockdowns of FoxN2/3, Hex, and FoxO appear to “sit” in place and rather than stretch outwards the blastocoel. A new measure was needed apart from nuclear displacement in order to determine whether movement in each knockdown was being restricted by some other factor or whether cells were genuinely immotile.

In control cells, apical constriction caused a shape change and cytoplasm was redistributed to the basal side of the cell, which subsequently bulged into the blastocoel. In the experimental situation, if the cell were inherently motile but unable to disassemble its adherens junction, the cell stretched further into the blastocoel and the portion of the cell remaining in the epithelium was stretched more than controls. With this phenotype in mind, motility of a tethered cell versus absence of motility was a measure of cell contents distribution and not of relative nuclear movement. Thus the analytical tools were designed to characterize the type of movement by cells.

**2.5.2.2 Methods**

Time-lapse data was analyzed for latent motility with Fiji (Schindelin et al. 2012), an open source image analysis platform built as a more functional distribution of NIH’s ImageJ software. First, the mem-GFP channel was made into a thresholded image. The threshold was automatically calculated by Fiji and adjusted up or down only under circumstances where the threshold included areas outside of cell area (in the case of background noise) or if the threshold level didn’t leave a hole in the middle of an
area filled by cells (in the case of embryos with lower than usual mem-GFP expression). An imageJ macro was written so that thresholded images were automatically bunched into binary and fed into the Analyze Particles tool, which measured the total cell area for each cell in each movie at all time-points. Another macro was written which batched the separation of individual nuclei using Find Maxima (to correct for overlapping nuclei) and subsequently counted the nuclei at every time frame using Analyze Particles. In order to find the distribution of cell area and nuclear count contained inside the epithelium only, another macro was written. The boundary between the epithelium and the blastocoel in the DIC reference images for each movie was defined manually for each embryo over the course of each time-lapse movie and another macro was used to project the traced boundary from the DIC image onto the corresponding fluorescent images for both nucleus and cell area and the portion of the fluorescent images contained inside the boundary was deleted. Both images with blastocoelar fluorescence removed were measured using Analyze Particles.

2.5.3 Cell Distribution Results

Each knockdown transplant remained more or less in the same space over the time-course as measured using nucleus tracking software (Supplemental Figure) and a novel approach to determining latent motility in a tethered cell was required. We calculated each total cell area of fluorescence inside the epithelium and inside the blastocoel of each embryo. Additionally, we determined the position of each nucleus
relative to the epithelium-blastocoel boundary. In general, the distribution of cell area closely mimicked the distribution of nuclei over time (See Figure 2.23). Absolute measurements of cell area were highly variable and were dependent on outside variables, in particular the size of the eggs shed by the adult. The variation of egg size was not an informative measure of differential EMT control and therefore we normalized the area measurements by using the percentage of fluorescent cell area remaining within the epithelium beneath the basal lamina relative to the whole fluorescent cell area of the embryo. As the percentage of cell area and number of nuclei within the epithelium goes down, the likelihood that the cell has a latent motility increases. Knockdowns with an increasingly lower percentage of cell area and nuclei contained inside the epithelium over time include Alx1, Snail, Twist, FoxB, Tgif and Tel; these genes are likely to possess latent motility, i.e. if they did not remain attached to their neighbors in some way, if an external force were to cut them free, they would move out of the epithelium and migrate away from the ingression site the same as any control cell. The opposite phenotype to latent motility is a very high percentage of cell area and nuclei inside the epithelium with little or no change in the distribution across the basal lamina. The latter phenotype was the case for Ets1, FoxN2/3, Hex and FoxO, which strongly suggested those knockdowns did not possess a latent ability to mobilize. The remaining two knockdowns, Tbr and Erg, don’t sort easily into either category and represent an intermediate motility phenotype. This analysis shows that the subcircuit
driving motility during EMT is most likely controlled by Ets1, FoxN2/3, Hex and FoxO; once again, because Ets1 is an upstream regulator of FoxN2/3 its connection to motility may only be its epistatic relationship to FoxN2/3, therefore, FoxN2/3, Hex and FoxO are placed into the subcircuit as the direct inputs into EMT motility.
Figure 2.23: Motility Inferred by Cell Area Distributions. The average percentage of cell area below the basal lamina over time within epithelium are blue, within blastocoel are green, standard error is red, and nuclei positioned directly on the blastocoel/epithelial boundary are black. The percentages are stacked on top of each other to represent the cell percentage distribution across the boundary over 2 hours. Controls have 60% average cell area within the epithelium at time 0 and end with ~30% cell area after 2 hours. Those that follow the decreasing trend ending with less than 60% cell are in the epithelium have latent motility over time. Those that have a very high percentage of cell area in the epithelium or an increasing rather than decreasing trend are immotile: Tbr, Erg, Ets1, FoxN2/3, FoxO, and Hex.
2.5.4 A Relationship Between De-Adhesion and Latent Motility

When we considered all of the genes that had at least one significant phenotypic difference from typical control ingression behavior so far, three genes remained with no explanation for how they differed from controls: Alx1, Snail, and Twist. Alx1, Snail and Twist all successfully apically constricted, showed no polarity defects, and had a decreasing cell area percentage within the epithelium over time indicating latent motility.

Figure 2.24: Alx1, Twist, and Snail Regulatory Mechanism. Previous studies showed that Alx1 is an upstream regulator of both Snail and Twist and that Twist and Alx1 create a feed-forward loop (Wu and McClay 2007).
Snail and Twist have both been previously shown to regulate cell adhesion in the sea urchin; specifically, Snail suppresses cadherin transcription and upregulates cadherin endocytosis (Wu, Yang, and McClay 2008). Data in virtually every other system where Snail and Twist have been studied, from fly to human, all support the same conclusion and implicate them in some form adhesion control (Medici, Hay, and Goodenough 2006; Morales et al. 2007; De Renzis et al. 2006). Alx1, Twist and Snail define the smallest possible de-adhesion subcircuit. The possibility remains that other genes may participate in this important EMT event. A separate assay to directly observe adhesive change could be used at a later date to expand the de-adhesion subcircuit to include all of the transcription factors with a role. Other studies directly indicated and considering no contradictory evidence was found in this study, the de-adhesion EMT subcircuit is made up of three members, Alx1, Twist, and Snail, with Twist and Snail being the direct inputs into adhesive changes. The caveat of this subcircuit decision is that knockdowns where there was no latent motility meant the cells did not move away from the epithelial layer, so there was no way to score for loss of adhesion via this approach.

2.6 Discussion of In Vivo Time-Lapse Results

The results of the time-lapse provided enough information to build four new subcircuits of control for EMT components within the context of the GRN. Those transcription factors that showed a defect for a particular behavior when they were
knocked down were upstream regulators of that EMT behavior. The data implied that some knockdowns might have operated in an epistatic manner to control EMT. At the level of analysis, the upstream gene may or may not have any direct inputs into cell biological components transcriptionally regulated by the network state change. Future experiments will be necessary to rule in or out a direct function for upstream transcription factors.

We were still interested in determining the subcircuitry for all five EMT characteristics mentioned previously. In order to score each GRN transcription factor for the ability to invade the basement membrane we needed to develop a separate assay from the original time-lapse experiment. The time-lapse assay included control micromeres that successfully invaded the basement membrane, proteolytically, mechanically, or in combination, leaving a path through which the transplanted cells could follow; therefore, we would never have seen an impaired invasion phenotype using the time-lapse assay.
Figure 2.25: A Subcircuit for Apical Constriction. Transcription factors that are involved in apical constriction were determined by shape analysis. Those necessary for proper apical constriction were Ets1, Tbr, Hex, Tgif, Erg, Tel, FoxB, and FoxO.
Figure 2.26: A Subcircuit for Apical-Basal Polarity. The results of shape analysis showed FoxN2/3 knockdown was the only perturbation that resulted in an apical-basal polarity defect.
Figure 2.27: A Subcircuit for Motility. Loss of motility was inferred by a knockdowns ability to pull past the basal-lamin/epithelial boundary and progress further into blastocoelar space as time proceeded. Transcription factors determined to be required for latent motility were Ets1, Erg, Hex, FoxN2/3, and FoxO.
Figure 2.28: A Subcircuit for De-Adhesion. Alx1, Twist, and Snail constitute a de-adhesion subcircuit based on previous data (Wu, Yang, and McClay 2008; Wu and McClay 2007) along with a shared phenotype: motility, shape change, polarity and invasion all occur normally yet ingression fails. These three transcription factors represent the minimum de-adhesion subcircuit.
3. Network Control of Invasion in EM

3.1 Experimental Design

We developed a second assay, separate from the *in vivo* time-lapse assay containing invading control PMCs, in order to build a subcircuit for basement membrane invasion. We first determined the precise timing of invasion during the normal course of ingression during sea urchin development. Embryos injected with a standard control morpholino were fixed at half-hour intervals between 7.0 hpf and 10.0 hpf. Fixed embryos were double immuno-stained with 1D5, a monoclonal antibody specific to a glycoprotein on the cell surface specific to PMCs, and anti-laminin (Abcam ab11575).

To determine the transcriptional subcircuit responsible for clearing a hole in the basement membrane during ingression, we measured the ability to invade for each gene of interest. Embryos were injected with knockdowns to each of the transcription factors downstream of pmar1 in the PMC gene regulatory network. All embryos were fixed at 10.0 hpf, after controls had completed ingression, and immuno-stained with 1D5 and anti-laminin. Because the laminin antibody used stains a meshwork of basal lamina, there were many small “holes” seen in the staining pattern but they were quantitatively distinguished from the large hole created by invasion. We measured the pixel intensity in three areas of each embryo: lateral left edge, lateral right edge, and vegetal pole. The difference between the mean of the lateral measurements and the vegetal measurement...
was averaged for each knockdown and their respective control embryos.

3.2 Methods

3.2.1 Animals

Animals were used as previously described in this document, See Chapter 2.3.1.

3.2.2 Injections

Embryos were injected as previously described with one of the following knockdowns in 10% glycerol: 1.5mM Standard Control MASO, 1.0mM Alx1 MASO, 0.75 µg/µL DNEts1 mRNA, 1.0mM Tbr MASO, 1.5mM Snail (2) MASO, 1.5mM Twist, 1.0mM FoxN2/3 (2) MASO, 0.5mM Erg (2) MASO, 0.75mM Tel (2) MASO, 1.5mM Hex (2) MASO, 0.5mM Tgif (1) MASO, 0.3mM FoxB (1) MASO, 0.75mM FoxO (2) MASO, or 1.5mM Dri (1) MASO.

3.2.3 Immunostaining

Embryos were cultured for 10.0 hours post fertilization at 23°C and were fixed in 100% MeOH, washed 4 times in PBS and incubated for 4 hours at room temperature or 4°C overnight in 4% Normal Sheep Serum in PBS. Embryos were stained with anti-Laminin antibody (1:300) obtained from Abcam (catalog: ab11575) and 1D5 mAb (1:1) in PBS and incubated 4°C overnight. Embryos were washed 4 times in PBS and incubated for one hour at room temperature in Cy3-conjugated anti-Rabbit IgG (1:300) and Cy2-conjugated anti-mouse IgM (1:300) (Jackson Immunoresearch Laboratory) diluted in PBS. Embryos were stained in a 1:1000 dilution of Hoechst’s (Molecular Probes) for 5
minutes and washed 5 times in PBS. Stained embryos were imaged with a Zeiss LSM 510 upright confocal microscope with a 40x/1.4 Oil plan-Apochromat objective. Z-slices were spaced either 1.0 μm or 0.5 μm apart spanning the diameter of the embryo.

3.2.4 Image Processing and Statistical Analysis

Three-dimensional projections of control embryos were rendered from all collected confocal z-slices spaced 0.5 μm apart using Imaris v.7.1.1 (Bitplane). Fiji software (Schindelin et al. 2012) was used for all subsequent image processing and statistical analysis. Confocal slices above or below the embryo that contained only background noise were removed from the data set and images were normalized for fluorescence attenuation using Fiji > Image > Adjust > Bleach Correction module in Fiji. Two-dimensional projections were rendered from z-slice = N/2 + 5 through z-slice = N/2 − 5, where N is the sum of all z-slices and N/2 is rounded to the nearest integer, for all embryos with slices spaced 1.0 μm apart for a projection through the center of the embryo totaling 10 μm exactly using Fiji > Image > Stacks > Z Project.... Stacks were converted to hyperstacks and rotated based so that the region of nuclei stacked 2+ deep were centered at the bottom of the image. Images were cropped to bound nuclei and pixel intensity values were measured using Fiji > Analyze > Measure > Raw Int Density as defined by a rectangular section Rectangle = (x, y, width, height), for three rectangular selections with scale set to μm: Lateral Left Rectangle = (0, (height/2 − 5), width/2, 10), Lateral Right Rectangle = (width/2, (height/2 − 5), width/2, 10), and Center Vegetal
Rectangle = ((width/2 – 5), height/2, 10, height/2).

3.3 Results

Laminin is a major component of basement membrane in the sea urchin. A polyclonal antibody that recognized laminin was obtained commercially, and it stained a meshwork pattern in a thin layer around the inside of the blastocoel where endogenous laminin deposits were detected previously (Benson et al. 1999). At 8.0 hpf embryos just began to hatch and the laminin antibody stained uniformly around the blastocoel (See Figure 3.2). At 8.5 hpf, when the vegetal plate thickened as future PMCs prepared for ingressin, small punctate 1D5 staining was seen within the cytoplasm of cells in the vegetal plate, the presumptive PMCs, and laminin antibody staining was cleared leaving a small hole adjacent only to 1D5 expressing epithelium. At 9.0 hpf, the onset of PMC ingressin, 1D5 stained the basal cell surface of nascent PMCs as they entered the blastocoel and laminin antibody was cleared leaving a hole encircling but never overlapping the ingressing PMCs. Ingression progressed through 9.5 hpf and most PMCs emerged into the blastocoel; 1D5 staining expanded to cover the entire cell surface of each PMC and laminin clearing expanded leaving a hole approximately the same diameter as the PMC cluster inside the blastocoel. The laminin hole at 10.0 hpf remained the same size as PMCs began to migrate away from the site of ingressin.
Figure 3.1: Invading PMCs Leave a Laminin Hole Behind (same embryo). Three-dimensional projection of confocal stacks. Anti-laminin (red) stains the basal lamina, 1D5 (green) is specific to PMCs, and Hoechst’s (blue) stains nuclei.

Figure 3.2: Laminin Staining in Control Time Course. At 7.5 hpf laminin staining (red) is uniform around the blastocoel. As PMCs (green) prepare for ingestion from 8.0-9.0 hpf the laminin above them begins to clear. When PMCs have fully ingressed, 9.5-10.0 hpf the hole in laminin has increased in diameter.
**Figure 3.3: Laminin Staining in Knockdowns.** Embryos 10.0 hpf were fixed and stained for laminin (red) to show basement membrane and 1D5 (green) which recognizes PMCs. The control embryos have completed ingression, all PMCs are fully inside the blastocoel and a hole remains in the laminin of the basement membrane where the PMCs entered almost an hour prior. In spite of the fact that none of the knockdowns have ingressed PMCs, most of them do have a hole in laminin where PMCs would normally enter the blastocoel. Therefore, many perturbed PMCs have prepared for invasion by clearing a path at the time when they would have invaded. There were three knockdowns which were unable to prepare for invasion and laminin staining was uniform throughout the inside of the blastocoel: Tbr, Hex, Dri.
The presence or absence of an invasion hole through the basal lamina was easily classified by manual visualization. However, because the antibody stained in a meshwork pattern, there were small empty spaces throughout the staining pattern in addition to the larger hole where the basement membrane was remodeled. The presence of endogenous empty space in laminin staining made it important to design an objective and empirical method to distinguish between invasion holes and empty mesh space.

We determined the presence or absence of a hole using the measured intensity difference between lateral laminin staining, where empty spaces were created by the mesh pattern, and vegetal laminin staining, where empty space may either be an invasion hole or empty mesh space. We measured total intensity sums within the volume of a rectangular prism (10 µm x 10 µm x embryo-radius µm) at three pre-defined landmarks of the embryo, left lateral, right lateral, and vegetal intensities were compared. The average intensity of laminin at the equator of the embryo was then compared to the intensity at the vegetal pole. A true invasion hole had very low intensity sums at the vegetal pole compared to the intensity mean at the equator, i.e. a large intensity difference indicated vegetal holes were created by invasion. Embryos lacking the ability to invade had very similar intensity measurements compared to the equator, i.e. intensity differences near zero meant holes at the vegetal pole were equivalent only to the empty space of the mesh and were not the result of invasion.

The difference between controls with an invasion hole and knockdowns in which
no hole appeared was determined to be statistically significant using pair-wise t-tests (See Figure 3.4). The knockdowns for Tbr, Hex, and Dri were statistically significant at the levels of p < 0.01, p < 0.001, and p < 0.0001 respectively.

![Figure 3.4: Invasion Analysis](image)

**Figure 3.4: Invasion Analysis.** The graph shows average intensity value differences between equatorial laminin and vegetal laminin staining. A hole created by invasion results in a large difference in intensities. If invasion fails and no hole is present in the basal lamina, then the intensity for the equatorial and vegetal laminin intensities will be roughly the same and the difference will be close to zero. Red bars show the average intensity difference for each knockdown group. Black bars show the average intensity difference for the control group. The control group always invades the basal lamina creating a hole and therefore always has a large intensity difference. Most transcription factors create a similar hole to controls and the intensity differences are not significantly different. The transcription factors Tbr, Hex, and Dri did not invade and their equatorial and vegetal laminin was distributed evenly with almost no difference. This was significantly different from respective controls at **p<0.01, ***p<0.001, and ****p<0.0001 respectively.
3.4 Discussion

A subcircuit for EMT’s invasion ability was constructed to include Tbr, Hex, and Dri. This fifth and final subcircuit was unique from previously built subcircuits in that the transcription factors driving invasion share no upstream regulatory inputs and have no downstream targets in common. This result could be explained in several ways. The first possibility is that each transcription factor directly upregulates the transcription of a different gene or suite of genes that is/are necessary for basement membrane invasion. The alternative to that hypothesis is that they are all three directly involved in the transcription of a single gene or suite of genes that is necessary for invasion; perhaps they regulate expression in an additive dose dependent manner. Other possible explanations could combine any number of parallel inputs and individual regulations of a set of genes required for invasion.

A particularly interesting case was FoxN2/3. The cells with FoxN2/3 MASO were scored as being motility deficient and lacking in a polarity shift. Based on those first phenotypes one possibility was that FoxN2/3 knockdowns had a broken internal “clock” and were simply delayed. This hypothesis would be congruent with previous data that shows FoxN2/3 MASO cells eventually do enter the blastocoel, albeit hours behind ingress of control cells (Rho and McClay 2011). Contradictory to this hypothesis, is the new evidence that FoxN2/3 knockdown resulted in production of the basal laminar hole. Thus, the FoxN2/3 knockdown failed to ingress on time due to its inability to
migrate and lack of proper polarity switch and not because the entire EMT program was delayed. In fact, the FoxN2/3 knockdown produced an invasion hole right on time.

Finally, because FoxN2/3 neither bulged into the blastocoel via apical constriction forces nor invaded by forcing its own path open through the basal lamina by directed motion, we conclude that there must be a proteolytic degradation occurring to allow the FoxN2/3 MASO to make a hole in basal lamina. This does not exclude the possibility that sea urchin invasion utilizes both a proteolytic mechanism, to begin a hole, and a proteolytic-independent mechanism, to widen an existing hole. However, this is strong evidence for a proteolytic mechanism of invasion as perhaps the first necessary step for breaching the basement membrane. These results provide the first direct evidence for the requirement of a proteolytic mechanism of basement membrane remodeling in the sea urchin.
Figure 3.5: A Subcircuit for Invasion. The invasion subcircuit is controlled by Tbr, Hex, and Dri, none of which regulate each other.
4. Alternative Splicing of Erg in the GRN

4.1 Introduction

4.1.1 Mechanism of Alternative Splicing

Regulation of transcript expression is the major mechanism by which a cell can alter its identity; changing the transcripts expressed directly alters the expression of proteins that functionally define the cell type. The gene for a transcription factor is transcribed initially to include both exons and all introns between coding exons. That transcript is then processed: the exons are spliced together to produce a mature transcript ready for translation. For some transcription factors, some exons are skipped during splicing, allowing different mRNAs to be created from a single transcriptional event. The mechanism of skipping over exons is called alternative splicing and represents a second level of transcriptional regulation. As transcriptome sequencing has become more routine, more and more genes are being found with alternative splicing variation. While the importance of alternative splicing has been understood for decades, the ubiquity of alternative splicing as a regulatory mechanism is finally coming to light. (Kosti, Radivojac, and Mandel-Gutfreund 2012).
Figure 4.1: Mechanism of Alternative Splicing. A simplified diagram shows how alternative splicing creates different proteins with unique functions from a single gene by splicing different combinations of exons together.

4.1.2 Alternative Splicing in the Sea Urchin Embryo

A handful of genes in the sea urchin have been shown to be alternatively spliced. An even smaller subset of alternatively spliced transcripts encode regulatory genes themselves. One example of alternative splicing of a transcription factor is blimp1/kroxa and blimp1/kroxb. Each alternatively spliced variant of this gene is turned on at different times during development and their expression is regulated by separate upstream transcription factors (Livi and Davidson 2006). However, a specific role for alternative splicing of a transcription factor has yet to be placed in the urchin GRN. Thus we asked whether alternative splicing produced functionally different
transcription factors with different regulatory targets, and we identified two such instances.

4.1.3 Alternative Splicing in EMT

In the human genome, nearly all genes with two or more exons are expressed as multiple spliceoforms (Wang et al. 2008). Considering the pervasive use of alternative splicing as a means of regulating transcription and creating diversity, a role for splice variation in the complex EMT event should be expected. In fact, alternative splicing has been shown to be a key regulatory mechanism for controlling EMT. Alternative splicing events have been shown to correlate with a risk of metastatic EMT occurrence in breast cancer (Warzecha and Carstens 2012) and prostate cancer (Leshem et al. 2011). A recent broad scale study of alternative spliceoforms in EMT uncovered new genes with literature predicting specific functions for alternatively spliced forms of IRSp53, NUMB, TCF-4, and Dynamin-2; many more new splicing events were found during EMT that were not predicted by any previous literature to have a specific function (Warzecha and Carstens 2012). This study is evidence that alternative splicing may be an important mechanism for coordinating EMT, at least in the case of human cancers (Warzecha and Carstens 2012; Shapiro et al. 2011; Weise et al. 2010).
4.2 Experimental Design

During routine cloning of the PMC transcription factor LvErg for utilization in the \textit{in vivo} time-lapse assay described previously (Chapter 2) we discovered that LvErg was a target for alternative splicing and a minimum of 11 splice variations were characterized. This was a surprising discovery because the sea urchin is a basal deuterostome and has a more ‘streamlined’ genome and very few regulatory genes identified in alternatively spliced forms as compared to ‘higher’ deuterostomes in the phylum chordata, especially in mammals, where a large percentage of regulatory genes are found as splice variations (Keren, Lev-Maor, and Ast 2010); finding that a regulatory gene with a large number of alternatively spliced transcripts in urchin was unprecedented. There are only two regulatory genes in urchin that have been shown to be alternatively spliced during urchin development; blimp/krox (Livi and Davidson 2006) and otx (Li et al. 1997). In this case, we had found more spliceoforms of LvErg than have been show in the homologous human complement of Erg variants, which totals seven at the time of this writing (C.D. Sun 2008).

When Erg was cloned from reverse transcribed cDNA, each time a pool of clones was Sanger Sequenced, a new or slightly different sequence of LvErg was discovered. After we characterized eleven alternative splice forms, we hypothesized that we had not yet saturated the potential for discovering every splice variant. In order to sequence enough transcripts to attempt reaching the point of saturation, we employed the third
generation sequencer PacBio® RS which is built on Single Molecule Real-Time (SMRT®) sequencing technology. We sequenced amplicons enriched for alternatively spliced transcripts of LvErg targeted to the variable regions. This approach allowed us to mine thousands of sequence reads at once for new splice variations; our goal was to achieve saturation with a level of confidence that could not be achieved by merely tens or hundreds of sequences that could feasibly be produced using traditional Sanger Sequencing methods. PacBio® sequencing also removes the cloning vector and bacterial transformation steps that were required for isolating unique transcripts for Sanger sequencing; directly reading the PCR products is therefore only one bias removed from native transcription levels, as opposed to the total of three biases introduced during the clone isolation process of heterogeneous PCR products. However, because of the potential for bias to be introduced during the PCR amplification step, the sequences could not be used to determine native transcription levels of spliceoforms. We were able to use, however, the relative read abundances for each spliceoform to infer relative abundance of alternatively spliced transcripts expressed in the embryo.

We then sought a way to ask if the different Erg spliceoforms actually had different roles in regulating targets in the PMC GRN. An experiment was designed to determine the placement of each of the two most abundant spliceoforms in the hierarchy of the PMC GRN. Both LvErg MASOs targeted to 5’UTR only and knocked down all native transcripts, i.e. every spliceoform (ΣLvErg). We determined the full suite of
genes targeted by ∑LvErg knockdown, then asked which of those targets were rescued if we selectively expressed only one splice variant within that all-removal background. Both Erg knockdowns and control embryos were collected at MB stage when ∑LvErg levels are highest (Materna, Nam, and Davidson 2010) and transcriptional changes were determined by in situ hybridization for genes in the PMC GRN.

We then attempted to rescue with the two most abundant as inferred by PacBio® sequencing: LvErg0 (all exons present) and LvErg4 (exon 4 missing). The data show that each splice form rescues expression of a subset of target transcription factors (a different subset rescued by each), but neither rescues expression of all known transcription factors controlled by Erg.

4.3 Methods

4.3.1 Animals

Animals were collected as previously described in this document, See Chapter 2.3.1.

4.3.2 Cloning LvErg

Cloning of LvErg was performed as previously described in Chapter 2.3.2 with additional sequencing steps described here. After sequencing full-length ORF clones obtained and ligated into the cloning vector pGEM-TEasy, a total of 4 clones were chosen for sequencing with SP6 and T7 primers. The sequences obtained from each of the 4 clones were aligned using the DNA sequence analysis software Sequencher v5.0 (GeneCodes). Sequences aligned with 100% fidelity between the start codon and base
where sequences began to diverge and all 4 sequences were unique. Sequencing many more clones confirmed the sequence divergences was indeed present in the clones themselves and not a sequencing error. The pool of FL LvErg sequences obtained from over 30 individual clones responded to 8 unique sequences. Those sequences were aligned to the annotated SpErg sequence and the missing portions of sequence in each of the 8 LvErg transcripts corresponded to known exon/intron boundaries in the species \textit{S.purpuratus}. To confirm that our finding was a species difference and not an anomaly obtained from the RT Reaction used to transcribe cDNA from whole embryo mRNA preps, the protocol for cloning full length LvErg was repeated a total of 8 more times using 8 different templates for PCR that consisted of either MB stage or LG stage cDNA in 3 different genetic backgrounds from 3 separate parental crosses that were collected by 3 different researchers and a total of 80 clones were sequenced, 10 clones per experimental condition. All 8 previously sequenced transcripts of LvErg sequences were represented in the pool of 80 sequences along with 3 additional unique transcripts not previously sequenced. Of the 11 unique transcripts, 5 were chosen based on their frequency for completing the additional steps as previously described in Chapter2.

\subsection{4.3.3 PacBio\textregistered Sequencing}

The LvErg PCR protocol was optimized for annealing temperature to produce large enough quantities for PacBio\textregistered sequencing. LvErg transcripts were amplified from a MB cDNA pool using primers [F – 5’ CTGCAGATCCTAATATGTGGACTGC – 3’; R –
CTCGACTTAGCTTGTCGTAGTTCA – 3’] which flanked the variable regions and produced a pool of amplicons 912bp or smaller. The PCR product was cleaned and sequenced by the Duke DNA Sequence & Analysis Core Facility on the PacBio® RS third generation sequencer. The sequencing facility also performed the assembly and analysis of contigs from the raw sequence data that was output.

4.3.4 Naming Scheme for Spliceoforms

The sheer number of spliceoforms sequences to date makes a simple numerical or alphabetical naming scheme difficult to keep track of. One would need a reference table to keep track of them all. Therefore, a consistent and logical naming scheme was applied based on the combination of exons skipped in each transcript. For example, LvErg4 does not contain exon 4, LvErg0 has had no exons skipped and is the full-length transcript, LvErg4.6.8 has had exons 4, 6, and 8 skipped, etc.

4.3.5 In Situ Hybridization

Anti-sense mRNA probes were in vitro synthesized using DIG-labeled UTPs for the following probes: Alx1 (Ettensohn et al. 2003), Erg (this study, all exons present), Hex (this study, See Chapter2), FoxO (this study, See Chapter2), Tel (this study, See Chapter2), Tgif (this study, See Chapter2), and Tbr (Croce et al. 2001). In situ hybridization protocol was followed as previously reported (Bradham and McClay, 2006).
4.3.6 Injections

Embryos were injected as previously described with one of four injectates: [0.5mM standard control MASO + 500 ng/µL mem-GFP mRNA], [0.5mM Erg2 MASO + 500 ng/µL mem-GFP mRNA], [0.5mM Erg2 MASO + 1,000 ng/µL LvErg4 mRNA], or [0.5mM Erg2 MASO + 500 ng/µL LvErg0 mRNA]

4.4 Results

In total, 22 spliceoforms of LvErg were sequenced. The LvErg in situ hybridization pattern shows that ∑LvErg pattern is restricted to the PMC territory at the MB stage, and therefore all of the splice variation is expressed within the same cell type at the same time. LvErg4 is the most abundantly expressed transcript comprising nearly half of all sequencing reads. LvErg0 is the next most abundant variant representing 13% of reads. The majority of sequenced spliceoforms occur very infrequently; the sum of the 18 lower abundance unique spliceoforms comprises only 19% of the population relative to the total number of reads.
Figure 4.2: LvErg Spliceform Variations and Relative Abundance. LvErg contains 11 exons, 7 of which are alternatively spliced. Exons 1, 2, and 11 span conserved protein domains and all transcripts sequenced contained each of those 3 exons. PacBio® sequencing showed the relative abundance of transcript variations for 21 different splice variants of LvErg present at 9.5 hpf. A single splice variant was found exclusively by Sanger sequencing, for a total of 22 spliceoforms identified.

We also confirmed ∑LvErg’s place within the PMC GRN hierarchy through perturbation analysis. We confirmed all of the downstream targets and found two additional targets not depicted in the original GRN model; FoxN2/3 and Tel. This could simply be a species difference between L.variegatus and S.purpuratus. We separately
placed LvErg0 and LvErg4 into the network hierarchy by knocking down $\Sigma$LvErg by MASO, adding back one spliceoform, and performing in situ hybridization for all the previously confirmed targets of $\Sigma$LvErg. We found $\Sigma$LvErg was upstream of Tbr, Tel, Erg, Hex, Tgif, FoxN2/3, and FoxO. LvErg0 was upstream of Tbr, Tel, FoxN2/3, and FoxO. LvErg4 was upstream of Tbr, Tel, Hex, and FoxO.

**Figure 4.3 In Situ Hybridization of $\Sigma$Erg throughout Development.** $\Sigma$Erg expression begins at 7.5 hpf in future PMCs and remains expressed in PMC territory until 10.0 hpf. By 2 hours post ingestion $\Sigma$Erg is off in PMCs and expression begins in the NSM territory, the NSM’s will also undergo an EMT event and once NSM EMT is completed $\Sigma$Erg expression is off. (Klein 2012)
Figure 4.4: Two Erg Spliceoforms Regulate Different PMC GRN Targets.
Embryos were injected with either control mRNA, ΣErg MASO, ΣErg MASO + Erg-0 spliceoform mRNA, or ΣErg MASO + Erg-4 spliceoform mRNA. Embryos were fixed at 10.0 hpf and in situ hybridization was carried out using probes specific to FoxN2/3, Hex, Tbr, Tgif, Tel, and ΣErg. Results show ΣErg is a repressor of Tbr and has positively regulates FoxN2/3, Hex, Tgif, and Tel. Spliceoform Erg-0 suppresses Tbr and up-regulates FoxN2/3 and Tel. Spliceoform Erg-4 suppresses Tbr and up-regulates Hex and Tel. (Klein 2012)
4.5 Discussion

The two alternatively spliced Erg transcripts, LvErg0 and LvErg4, each regulate a unique subset of the genes target by ∑LvErg. These results suggest a unique role for unique spliceoforms in the developing embryo. Further study will be needed to find out how many spliceoforms have unique regulatory roles versus redundant roles. For 22 spliceoforms to regulate a total of 6 downstream targets, we hypothesize that there will be some redundancy in the pool of splice variants. However, the maximum possible combination of unique subsets for controlling 6 downstream targets is 720, so it is certainly mathematically possible for each of the 22 spliceoforms to have a unique subset of transcriptional targets; it would still be possible for spliceoforms to have unique roles even if twice as many species are ultimately found once sequencing saturation is reach and we are certain we have described the full complement of spliceoforms.
Figure 4.5: A Subcircuit of Erg Spliceoforms. Erg-0 and Erg-4 regulate unique but overlapping subcircuits of $\Sigma$Erg. FoxO is an unconfirmed target of $\Sigma$Erg in this species is drawn as a dotted line to show its tentative relationship. We have found several additional targets in *L. variegatus* that are not present in the *S. purpuratus* data, however we have yet to find a node in *Sp* that does not also exist in *Lv*. (Klein 2012)
5. Discussion

5.1 From Subcircuits to Networks and Beyond

Over the course of this study, we uncovered a total of seven new subcircuits of the PMC GRN. The first five subcircuits each describe one blueprint of regulatory control over one distinct cellular behavior; together the five subcircuits describe one completed morphological change: the ingress of PMC’s into the blastocoel via EMT. The last two subcircuits map the regulatory control for two different splice variants of the same transcription factor, LvErg. This constitutes a novel role for alternative splicing in development as a secondary level of transcriptional regulation and network state change.

The subcircuits provide a convenient way to visualize which set of transcription factors is responsible for each cellular EMT component and show how Erg spliceoforms add complexity to the regulatory logic complicating the means through which one transcription factor is able to regulate morphogenesis in multiple subcircuits of EMT. When all the subcircuits are mapped back onto the PMC GRN, the unified view provides a first glimpse of the embryo’s solution to the question of how many different cell biological changes will be coordinated and directed into one seemingly fluid morphogenetic change, the ingress EMT event.
Figure 5.1: Integrated View of GRN. All EMT subcircuits and spliceoforms integrated into the PMC GRN show a broad picture of regulatory mechanism for orchestrating both morphogenetic movements and subsequent differentiation.
Building cell biological event subcircuits onto the GRN provides an exciting and novel view of how the embryo develops new tissue types, but it is not the only way to view the complexities of developmental mechanisms. The GRN graphic primarily represents causal events using Boolean logic to explain relationships between transcription factors, and while this is arguably the most useful way to depict the data, any subtle relationships or interactions that exist in the embryo would be lost if it were the only method used to report our findings. Those findings grouped the transcription factors into subcircuits based on similar quantitative outcomes from the various methods employed. We sought an even higher dimensional way to report the findings, again seeking the most objective means possible. Two standard analytical methods for visualizing high-dimensional multi-variant data sets were chosen.

We chose to visualize our data set using both Emergent Self Organization Maps (ESOM) and cluster analysis. ESOM is a common tool used for determining patterns within high dimensional data sets and can be used in a wide array of applications ranging from “extracting interpretable muscle activation patterns” (Moerchen, Ultsch, Noecker, et al. 2005) to “visualization of music collections according to perceptual distance” (Moerchen, Ultsch, Hoos et al. 2005). This visualization tool should help us discern patterns outside of the network that connect PMC transcription factors together. This is both a useful tool for interpreting meaning from the data and also an independent test of the viability using a Boolean network model to fit the results.
Finally, we performed a cluster analysis on our data set. Cluster analyses are traditionally used to group similar items together. By objectively clustering the data obtained from the time-lapse analyses, the precise relationships between transcription factors and their phenotypes will increase confidence that our association calls were correct. This method supplements the Boolean model by showing how objects are related over a continuum.

5.2 Methods for Visualizing High-Dimensional Data

The ESOM was generated using ESOM analysis software freely distributed by the Databionics Research Group at the University of Marburg, Germany (Ultsch 2005). The dataset, comprised of measurements from both the time-lapse assay and the invasion assay, was preprocessed for training by removing outliers +/- 2 standard deviations away from the mean. The empirical probability distribution was analyzed using histograms, and while it was slightly skewed, non-linear transformation did not improve distribution. The data was re-scaled to [0,1] to normalize the mean and variance. The ESOM was trained with the preprocessed data set and the original data set was mapped.

For cluster analysis, we used the freely distributed software CLUTO v.2.1.2 (Karypis 2002). We ran the cluster analysis using over 300 different combinations of algorithm parameters from [3,20] clusters. We determined the most meaningful clustering results for our dataset were obtained using the following optional parameters
[CLMethod=RBR, CRfun=I2, SimFun=CorrCoef, #Clusters: 6]. The similarity function chosen provided the widest variance in resulting clusters.

5.3 An Integrated View of EMT: Next Level GRN

The integrated view of the all EMT subcircuits built onto the existing PMC GRN is likely the most complete description of a total in vivo developmental EMT ever described to this date. We see that the de-adhesion subcircuit is self contained while the other four events (invasion, polarity change, motility, and apical constriction) overlap by varying degrees while still each is uniquely controlled. We then built yet another layer of interpretation to our results using ESOM and clustering tools. This provided a deeper layer of understanding how transcription factors were related to each other that the GRN alone is not able to show us.

The ESOM shows the variation within the total dataset organized as if it were a topological map (See Figure 5.2). The white peaks represent objects in the data set that are far from any other neighbors and the blue at the bottom occurs wherever objects are very dense or close in to their neighbors. When we view the distribution of different transcription factor knockdown objects we see clear sorting into valleys for transcription factors that have little variability within their own phenotypes and we see the most variable phenotypes on mountains. The organization tells us something about the phenotypes that one could never tell from looking at the isolated GRN. Even transcription factors with a high dissimilarity within their own phenotypes are always
more closely related to themselves than to the phenotype of any other transcription factor. For instance, the GRN predicts several genes should have identical phenotypes since they are always together in the same subcircuits. If the GRN is correct, then transcription factors that are found in identical sets of subcircuits should be indistinguishable from each other using unsupervised ESOM and if they are distinguishable from each other it may be said that their phenotypes are similar but unique. Just as the integrated GRN predicts, Ets1 and FoxO are almost indistinguishable from each other on the ESOM. However, we see that the same is not true for Tgif and FoxB which share an apical constriction phenotype in common, the only subcircuit either one controls. Instead, the ESOM shows them distinctly apart from each other. So while FoxB and Tgif are certainly related to each other, they can hardly be called identical.

Finally, a cluster analysis of the in vivo time-lapse data shows us how phenotypes and transcription factors are related to each other in an unbiased way, independent of the GRN. Interestingly, the cluster analysis recapitulates, by itself, many of the same relationships that were established and concluded from the experimentally determined GRN topology (See Figure 5.3). We also see the relationship between phenotypes when they are clustered together, which tells us how phenotypes may depend upon each other. The clustering confirms the subcircuit prediction that the ability to invade the basal lamina is independent of other phenotypes. On the other hand, the ability to
Figure 5.2: Emergent Self Organization Map of Time-Lapse Data. The white peaks and blue valleys on the map represent data points that are highly dissimilar and similar respectively. Objects that fall in valleys together represent groups or features within the data set. Objects on peaks are the most unique. The presence of peaks and valleys illustrates that there is a high degree of variation in the dataset and that the scored EMT phenotypes are different enough to be used as a meaningful classification method.
Figure 5.3: Clustering Analysis. Each tree represents a hierarchal k-means cluster. The tree on the left represents each transcription factor in the PMC GRN and the tree on top shows how phenotypes relate to each other. The transcription factor relationship tree mimics many of the same groups that the PMC GRN predicts, therefore gene regulatory interactions are interrelated to the phenotypes they produce.

Apically constrict is related to the acquisition of motility. When the cluster is considered alongside the time-lapse observations, it appears that if the first driving force of an EMT is not apical constriction then it is more likely that EMT will additionally fail during the acquisition of motility.

The clustering results become most informative and biologically relevant in the instances where it does not recapitulate the relationships of phenotypes that would be
predicted by the regulatory relationships among transcription factors. For example, the
PMC GRN shows that Tbr is a direct regulator of FoxN2/3, yet the phenotypes for these
two transcription factors do not cluster closely together. This highlights the importance
of this integrative type of research; the transcriptional circuitry outlined in the GRN is
highly predictive of phenotypic control, but it is not the whole story and understanding
the control of morphogenesis is more complicated than the sum of GRN nodes.

The new multi-layered GRN for cell fate specification and EMT driven
morphogenesis is the first of its type and provides an objective analysis that reveals a
novel total system explanation for how the urchin embryo completes ingression and
specifies PMCs. Importantly, it demonstrates that the GRN controlling EMT does so by
deploying subcircuits of transcription factors, each subcircuit controlling a separate
component of the EMT process. In isolation, the GRN’s explanation of EMT
orchestration is very powerful, but even more insight is gained by viewing the data in
new ways that explain relationships independently based on phenotypic identities.

The phenotypic identities discovered in this study are a new way of
understanding how transcription factors are related. In terms of cancer research,
knowing which transcription factors are active in a tumor could help predict the best
route for drug therapies. When we know which cellular components of EMT
phenotypes are activated by those transcription factors, drugs specific to the cell biology
driving only those EMT behaviors could be used to prevent metastasis. For instance,
using our sea urchin model, we would predict that a matrix-metalloproteinase inhibitor as a cancer drug would be entirely ineffective for treating tumors in patients with up-regulated Twist and Snail genetic markers; however, targeting endocytic degradation pathways in such a tumor to inhibit the disassembly of adherens junctions would be more likely to prevent metastasis. Today, cancer research focuses on preventing the EMT event as a whole. By considering the intermediate phenotypes of an EMT individualized therapy for specific tumor GRN states could be targeted and would hopefully be more effective than the broad scale use of “EMT inhibiting” drugs which may or may not be relevant to any given patient’s disease. If cancer researchers begin using our integrated GRN/phenotype approach to understanding metastasis, we could expect to see more customized drug therapies and hopefully, greater life expectancies for cancer patients.

For developmental biologists our findings are important because we now understand how the embryo is able to coordinate the morphogenetic movement of an EMT at a deeper level than ever before. We know that intermediate phenotypes are orchestrated together by transcriptional regulation that is both complexly intertwined, as in the case for motility and apical constriction, as well as motifs that are not even remotely related within the GRN context, in the case of invasion. This total approach to understanding the complicated EMT process, which remodels the embryo so
dramatically, is a new stepping stone on the path to understanding how an embryo develops from a single cell into a fully formed organism.

5.4 Future Directions

The future research to follow this study will be very exciting. The goal of understanding the whole picture of EMT event orchestration is closer than ever and we have all the tools necessary to fill in those holes in knowledge still remaining. The next logical step to follow this study is to screen for the direct transcriptional targets of each PMC transcription factor. Now that we understand which cell biological behaviors each transcription factor has a hand in controlling, we can use the large knowledge base for each cell biology event for a candidate gene approach to finding direct mechanistic targets. It would be prohibitively expensive to use an RNAseq approach for finding transcriptional targets of every PMC transcription factor knockdown. However, a more informed RNAseq approach would be to sequence only the few knockdowns that we show are most likely to be directly regulating different EMT subcircuits. There are several combinations of only four knockdowns that would cover at least one of the terminal transcription factors for all five EMT subcircuits and that is a much more reasonable goal for a whole transcriptome sequencing approach.

The Erg spliceoform project has exciting future possibilities as well. The first PacBio® run found several splice variants in a very low abundance and entirely failed to sequence one spliceoform that was picked up by Sanger sequencing. We can make a
reasonable prediction that we have not reached saturation of the total complement of
splice variations possible and perhaps one or two more PacBio® runs will ensure we
have found every possible variation of Erg. We will continue to place more spliceoforms
into the network hierarchy, and it will be very interesting to see how many regulate the
same set of targets versus a unique set of targets.

Of course, the most interesting experiment will be tying the two studies together
by re-examining the in vivo time-lapse assay for all of the LvErg spliceoforms and
determining whether there are functional EMT differences for different alternatively
spliced transcription factors.
Appendix

A supplemental DVD is provided to show time-lapse movies which are summarized in Figure 2.16 on pages 59-60.
References


Biography

I was born in Hartford, CT on Dec 16th, 1983. I was awarded the Presidential Fellowship in 2002 to study and Simpson College in Indianola, Iowa. I received my Bachelor of Arts in Biology with a Minor in Mathematics from Simpson College in May of 2006. I was granted a summer fellowship from the National Institutes of Health in Bethesda, Maryland to study at the National Institute of Child Health and Human Development for the summer of 2005. In the fall of 2006, I enrolled as a PhD student in the Developmental Biology Training Program at Duke University in Durham, North Carolina and joined the Duke University Biology Department in 2008.