ZEB1 is a central mediator of the Epithelial-Mesenchymal Transition

by

Kong Jie Kah

B.A., M.Sci., Natural Sciences
University of Cambridge, 2003

SUBMITTED TO THE DEPARTMENT OF BIOLOGY IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF

DOCTOR OF PHILOSOPHY IN BIOLOGY
AT THE
MASSACHUSSETTS INSTITUTE OF TECHNOLOGY

JUNE 2012

Signature of Author: _____
Department of Biology
April 3, 2012

Certified by: _____
Robert A. Weinberg
Professor of Biology
Thesis Supervisor

Accepted by: _____
Stephen P. Bell
Professor of Biology
Chairperson, Graduate Committee

ZEB1 is a central mediator of the Epithelial-Mesenchymal Transition

by

Kong Jie Kah

Submitted to the Department of Biology on April 5, 2012
in Partial Fulfillment of the Requirements for the
Degree of Doctor of Philosophy in Biology

Abstract

Carcinomas are solid tumors arising from epithelial tissue, and account for the majority of cancer deaths in the United States. In most occurrences of carcinoma, it is the metastases that kill, not the primary tumor. The Epithelial-Mesenchymal Transition (EMT) provides a model by which tightly associated epithelial cancer cells can disseminate to distant sites. Many factors are known to trigger the EMT, but the extent to which the observed phenotypes represent a common process is unknown. There is also little appreciation of the extent to which EMT-inducing factors interact with one another or act on common or redundant pathways. In this study, I sought a common gene expression signature of the EMT by comparing five mesenchymal cell lines independently derived from the same parental epithelial line using different EMT-inducing factors. The resultant EMT core signature strongly suggested a common pathway is involved. Bioinformatics analysis revealed the transcription factor ZEB1 to be a possible mediator of this common pathway. ZEB1 was found to be both sufficient to induce EMT and necessary for maintaining the mesenchymal phenotype in the same cells. ZEB1 and miR-200 were known to reciprocally regulate each other, but their relative importance to the EMT phenotype had never been directly tested. I found that ZEB1 induced EMT regardless of miR-200c levels, thereby excluding the model in which miR-200c downregulation is a necessary step for the EMT. I also show evidence that EMT induced by the transcription factor Snail works at least in part through ZEB1.

Thesis Supervisor: Robert A. Weinberg

Title: Professor of Biology, MIT; Member, Whitehead Institute for Biomedical Research

Biographical Note

Kong Jie Kah
Whitehead Institute for Biomedical Research
9 Cambridge Center, Room 305
Cambridge, MA 02142
(617)-258-5175
kjkah@mit.edu

Education

2005-2012 **Massachusetts Institute of Technology, Cambridge, MA**
Ph.D. in Biology, Degree expected June 2012

1999-2003 **Christ's College, University of Cambridge, Cambridge, UK**
B.A., M.Sci. in Natural Sciences, Biological

Awards and Fellowships Received

2008-2011 Department of Defense Breast Cancer Research Predoctoral Traineeship Award

2005 MIT Presidential Graduate Fellowship

2002 College Book Prize and S. W. Greig Prize, Christ's College (Ranked 2nd in NST II Biochemistry)

2001 College Book Prize and Rosabel Spencer-Thomas Prize, Christ's College (Topped University list in NST IB Biochemistry)

2000 College Book Prize and S. W. Greig Prize, Christ's College

1999 Overseas Merit Scholarship (Open), Public Service Commission, Singapore

1999 Lee Kuan Yew Award for Mathematics and Science, "A" Level

1997 Lee Kuan Yew Award for Mathematics and Science, "O" Level

Research Experience

2006-2012 **Whitehead Institute for Biomedical Research, Cambridge, MA**
Thesis research work in the Robert A. Weinberg laboratory studying the role of ZEB1 as a central mediator of the Epithelial-to-Mesenchymal Transition.

2002-2003 **Department of Biochemistry. University of Cambridge, Cambridge, UK**
Analysis of heterogeneous nuclear ribonucleoprotein (hnRNP) E. Undergraduate research for NSTIII Biochemistry in the Kevin Brindle laboratory supervised by Dr Sandra Fulton.

2002	**Bioprocessing Technology Centre, Singapore** Assisting in the start-up of a haematopoietic stem cell project at the BTC. Short formal attachment arranged by the Public Service Commission through the Agency for Science, Technology and Research.
2002	**Department of Biochemistry. University of Cambridge, Cambridge, UK** Purification and characterization of follostatin domain FS2. Undergraduate research for NSTII Biochemistry in the Timothy Blundell laboratory supervised by Dr Marko Hyvönen.
2001	**Whitehead Institute for Biomedical Research, Cambridge, MA** Investigating anti-apoptotic gene expression in erythroid development using a novel FACS methodology. Work carried out as full-time undergraduate assistant in the Harvey Lodish laboratory, supervised by Dr Merav Socolovsky.
2000	**National Taiwan University, Taipei, Taiwan** Characterization of Poly(ADP-ribose) Polymerase interaction with nucleolin in Sf21 cells. Informal attachment based at the Institute of Molecular Medicine, a part of the National Taiwan University.

Other Academic Experience

2011	Poster presenter, Era of Hope Conference for Breast Cancer Research Program, Orlando, FL
2009	Teaching Assistant for MIT Undergraduate Biology Course 7.17, a molecular biology laboratory course taught by Mary-Lou Pardue and Frank Gertler
2006	Teaching Assistant for MIT Graduate Biology Course 7.52, an advanced genetics course taught by H. Robert Horvitz, Susan L. Lindquist and David E. Housman
1998	Represented Singapore in the International Physics Olympiad in Reykjavik, Iceland
1997	Participated in the Singapore Physics Olympiad and Singapore Chemistry Olympiad

Acknowledgements

I would like to thank first and foremost my thesis advisor, Dr. Robert Weinberg, for his support and guidance throughout my stay in the Weinberg lab. I have learned much from him, both inside and outside science, in particular the spirit of constant learning and the importance of being willing to question old assumptions. I am grateful for the collegiate environment he has fostered in the lab and the freedom he has given me to develop my own ideas.

I am thankful to the members of my thesis committee, Dr. Robert Horvitz and Dr. Michael Hemann, for their many years of guidance and advice which have been invaluable to my thesis work.

I am grateful to Dr. Kornelia Polyak for graciously taking time to serve on my thesis defense committee.

I would like to thank Dr. Richard Goldsby, for many discussions and helpful advice, especially in the past year.

I would like to thank former lab members Tamer Onder and Sendurai Mani, whose work provided the basis for my thesis. Dr. Piyush Gupta, also formerly of the lab and at the time at the Broad Institute, was invaluable in facilitating the original microarray experiment that led to this work. I was also greatly assisted in this work by the Whitehead Institute Genome Technology Core, in particular Jeong-Ah Kwon, as well as George Bell of BaRC. I would like to thank Michael Lodato and Tobias Brambrink of the Jaenisch lab for valuable reagents, as well as Julia Zeitlinger of the Young lab for sharing her unpublished data.

I would like to express my gratitude to both past and present members of the Weinberg lab, for being such a big part of my life for so many years. In particular, thanks to Leonardo Rodrigues, for the many enjoyable and helpful discussions on technology and bioinformatics and many things besides. Thanks to Li Ma and Ann Li for always being so gracious with their time and reagents during the times I had to do protein work. Thanks to Wai Leong Tam, my fellow Singaporean, for being there to have someone to speak Singlish to, and on occasion to talk about science as well.

I am extremely thankful to the Elinor Eaton, Joana Liu Donaher, Mary Brooks and Ann Gifford, who over the years have been surrogate moms to a student living alone far from home. Just as NCOs are the backbone of the army, I believe our technicians are the backbone of the lab. They help maintain the character of the lab over the years, and things would rapidly fall apart without them. I cannot appreciate their tireless work enough.

I am thankful to my Singaporean friends in Boston, who have reminded me of home and given me the excuse to organize social and cultural gatherings and teach people to cook traditional Singaporean cuisine. For someone who has spent more than a third of his life away from home and who could not afford regular visits, this was an anchor for my identity as a Singaporean.

Finally, I would like to thank my parents, Yin Muk and Seow Wah, for the years of love and support. My dad has always been there to make sure my feet were firmly rooted to the ground, even as I followed my ambitions. My mom has been supportive of my decisions, even as she wished I would spend more time at home.

Table of Contents

Abstract **2**

Biographical Note **3**

Acknowledgements **5**

Table of Contents **7**

Chapter 1: Introduction **10**

1.1 Cancer is a genetic disease 11

1.2 Metastasis is the major cause of mortality in cancer 12

1.2.1 Metastasis is a multi-step process 12

1.2.2 Mechanical constraints of carcinoma metastasis 13

1.3 Epithelial-Mesenchymal Transition (EMT): A developmental process 14

1.4 EMT in cancer progression 18

1.5 EMT-inducing factors 20

1.6 EMT-inducing transcription factors 21

1.6.1 Snail/Slug family 21

1.6.2 bHLH family 22

1.6.3 ZEB family 22

1.6.4 Molecular interactions of ZEB1 24

1.7 A common circuit for the EMT 26

1.8 References 27

Chapter 2: Creation and Analysis of the EMT Core Signature **36**

2.1 Introduction 37

2.2 Results and Discussion 38

2.2.1 Generating the EMT core signature 38

2.2.2 Comparing EMT core signature to publicly available gene signatures 42
2.2.3 Selecting candidates by transcription factor binding site enrichment 46
2.2.4 Gene Set Enrichment Analysis (GSEA) 50
2.2.5 Comparing EMT core signature to Drosophila ChIP-chip for Twist and Snail 53
2.3 Conclusions 55
2.4 Materials and Methods 56
2.4.1 Microarray Data Analysis 56
2.4.2 Transcription factor binding site analysis 57
2.4.3 Gene Set Enrichment Analysis (GSEA) 57
2.5 Supplementary Data 58
2.6 References 62
Chapter 3: ZEB1 in EMT and its relation to other EMT mediators 66
3.1 Introduction 67
3.2 Results and Discussion 67
3.2.1 Generating a lentiviral vector for rapid confirmation of protein expression 67
3.2.2 ZEB1 is sufficient to induce EMT in HMLE cells 70
3.2.3 ZEB1-induced EMT is rapid 71
3.2.4 Twist-induced EMT displays characteristics of selective outgrowth 75
3.2.5 ZEB1-induced EMT is apparently not stable 77
3.2.6 ZEB1-inducible single-cell clones can stably maintain a mesenchymal state 79
3.2.7 Endogenous ZEB1 is necessary for maintaining the mesenchymal state 82
3.2.8 ZEB1 overrides miR-200c in determining EMT status 85
3.2.9 ZEB1 is a mediator of Snail-induced EMT 91
3.2.10 Interfering with the mesenchymal state of ZEB1-induced cells 95
3.2.11 Downstream target genes of ZEB1 96

3.3 Conclusions 100
3.4 Materials and Methods 101
3.4.1 Cell culture and reagents 101
3.4.2 mRNA and miRNA extraction and detection 101
3.4.3 Immunofluorescence 101
3.4.4 Transcription factor binding site analysis 102
3.5 Supplementary Data 103
3.6 References 107
Chapter 4: Conclusions and Future Directions 111
4.1 Significance and utility of the EMT core signature 112
4.2 Relationship between EMT core signature and published gene signatures 116
4.3 ZEB1 is qualitatively different from other EMT-inducing factors 119
4.4 ZEB1 can override miR-200c 121
4.5 ZEB1 downstream targets 122
4.6 Final Perspective 124
4.7 References 125

Chapter 1:

Introduction

1.1 Cancer is a genetic disease

Cancer is a leading cause of death in the developed world, and with growing affluence and lifespans, is expected to increasingly dominate in the developing world as well (Center et al., 2011). Fundamentally a genetic disease, cancer arises from cells accumulating genetic and epigenetic lesions, some inherited, some spontaneous, and others caused by external agents such as radiation, chemical damage or viral infection. The development of cancer is hindered by numerous barriers, each of which has to be overcome in turn. The widely recognized hallmark abilities include being able to proliferate uncontrollably, escaping replicative senescence and resisting cell death (Hanahan and Weinberg, 2011). No one genetic lesion is able to provide a cell with all of the hallmarks; several rate-limiting steps are therefore needed to achieve malignancy (Vogelstein and Kinzler, 1993).

With the advent of affordable DNA sequencing, cancer genome sequencing projects have sought to identify the genetic lesions responsible for the underlying cancer (Sjöblom et al., 2006; Berger et al., 2010; Chin et al., 2011). Beyond finding the well-known lesions such as TP53 (Bell et al., 2011), the identification of significant new mutations is hindered by a high background of "passenger" mutations and the apparent low recurrence of "driver" mutations in numerous samples. The term "driver" carries the connotation that the mutation confers a growth advantage to cells that carry them, while "passenger" mutations do not. However, the actual identification of a "driver" mutation from cancer genome sequencing is based on a statistical overrepresentation of mutations in the particular gene. This process carries the underlying assumption that "passenger"

mutations are randomly distributed and will not be overrepresented. If this assumption is broken, as it often is due to the non-uniform structure of the genome, "passenger" mutations that happen to have a higher chance of appearing will be called as "drivers" (Bignell et al., 2010). Empirically, genes that are called as "drivers" rarely become confirmed experimentally, save the ones already known.

1.2 Metastasis is the major cause of mortality in cancer

The majority of cancers seen in the clinic are carcinomas, stemming from various epithelial tissues of the body. The top three common cancers in both males and females, accounting for half the reported cancer deaths in the United States, are all carcinomas (Siegel et al., 2011). Epithelial tissues such as the breast and prostate are not strictly essential, and the disruption of their function as a result of a local neoplastic growth is not a life-threatening event. Similarly, many alimentary canal neoplasias arising from the epithelial lining can be surgically removed along with substantial portions of surrounding healthy tissue. In the vast majority, carcinomas do not kill patients by disrupting the tissues in which they arise. Rather, they metastasize to distant sites and disrupt essential organs and tissues, such as the brain, liver and bone marrow. This final stage of cancer progression is what usually leads to the most suffering and eventual mortality (Gupta et al., 2006). Hence the study of the processes involved in metastatic progression is a key front in the war on cancer.

1.2.1 Metastasis is a multi-step process

Just as tumorigenesis is hindered by numerous barriers, so too, cancer cells must acquire several abilities to overcome the obstacles to generating full-fledged metastases (Steeg, 2006). At the minimum, the cells must first be able to invade locally to access the lymphatic or circulatory network. Next, they must intravasate into these long-range transport networks and survive in what is, to them, a foreign environment. At their destination, cells may or may not need to extravasate from the circulatory system, depending on whether they can survive indefinitely inside the blood vessels (Vaage, 1989). Even if the cells do survive, the majority will remain dormant and never grow to a size of clinical significance (Luzzi et al., 1998). A few may eventually overcome the absence of favorable microenvironmental factors to colonize the new site and grow into macrometastases.

1.2.2 Mechanical constraints of carcinoma metastasis

The movement of cancer cells to distant sites (excluding the draining lymph nodes) implies travel through the circulatory system. To reach most sites, the cells must follow the venous circulation into the heart and pass through the capillary networks of the lungs before being distributed through the systemic arterial circulation (Chambers et al., 2002; Steeg, 2006). Even the largest shunt metarterioles that bypass the capillary bed have a diameter of no more than 20 μm, which barely allows single non-blood cells to fit. This size restriction implies that metastasizing cancer cells should exist as single cells at some point in their history, in order to pass through the lung circulation. Indeed, circulating tumor cells, detected as single cells, are an independent prognostic indicator in metastatic breast cancer (Cristofanilli et al., 2004). The dissemination of single tumor cells from a bulk primary tumor has also been directly observed by intra-vital multiphoton

microscopy (Wyckoff et al., 2007). Epithelial cells, from which carcinomas derive, establish close contacts with each other and organize into tightly-bound sheets in their natural context. The single-cell state is not one that epithelial cells normally find themselves in, as tight cell-cell attachments are among their defining characteristics. Even when epithelial cells replicate in an uncontrolled manner, they will still maintain these characteristics and respect topological boundaries defined by their underlying basal lamina (Figure 1).

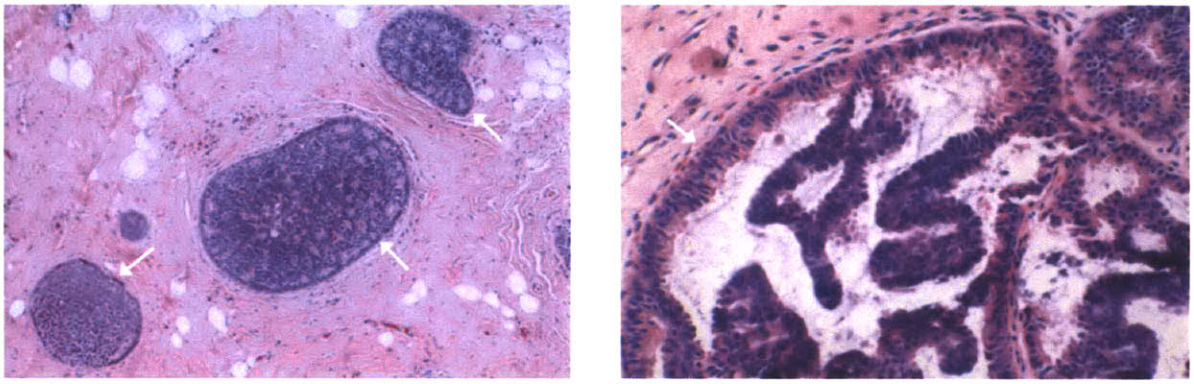


Figure 1. Photomicrographs of human mammary Ductal Carcinoma *In Situ* (DCIS). Arrows indicate where the carcinoma cells respect the basal lamina boundary, despite overgrowing into the luminal space. From http://tgmouse.compmed.ucdavis.edu/jensen-mamm2000/

1.3 Epithelial-Mesenchymal Transition (EMT): A developmental process

The Epithelial-Mesenchymal Transition (EMT), first described as a developmental process, provides a possible explanation for the ability of epithelial cancer cells to undergo the changes necessary to escape their normal topological confines. The observation that cells could convert from an epithelial morphology to a more mobile form was made more than a century ago, both in the context of cancer and development (Ramón y Cajal, 1890; Lillie, 1908), but the actual term

"epithelial-to-mesenchymal transformation" (later renamed as transition) was introduced later, largely in the context of developmental biology (Hay, 1968).

The EMT has been observed at multiple stages of development. In early development, an EMT occurs during the onset of gastrulation, as classically illustrated in the sea urchin embryo (Figure 2A) by the formation of primary mesenchyme cells and their ingression through the basal lamina into the blastocoel (Katow and Solursh, 1980). The analogous process has also been observed in embryos of other model organisms, including mammals (Viebahn, 1995). In all cases, the primitive ectoderm, which is epithelial in nature, gives rise to the mesoderm, which is mesenchymal. The process starts with the invagination of the epithelial cells in a specific area of the embryo, brought about by morphological changes to shrink the apical end of the cell and expand the basal end. The future mesoderm cells breach the basement membrane and lose the tight cell-cell interactions characteristic of epithelial cells, whereupon they migrate outwards, spreading to form the mesoderm layer.

In vertebrates, another well-studied occurrence of an EMT is the formation of neural crest cells (Duband et al., 1995). After the invagination of the neural plate from the ectoderm, the neural tube is formed. Interactions between the neural epithelium and the non-neural ectoderm remaining above it lead to the emergence of neural crest cells from the neural epithelium (Selleck and Bronner-Fraser, 1995; Mancilla and Mayor, 1996). The neural crest cells lose tight cell-cell interactions and delaminate from the epithelia, then proceed to migrate as individual cells throughout the developing embryo (Figure 2B). They eventually contribute to a wide variety of

tissue, such as craniofacial cartilage and bone, melanocytes, adrenal medulla and various components of the nervous system (Knecht and Bronner-Fraser, 2002).

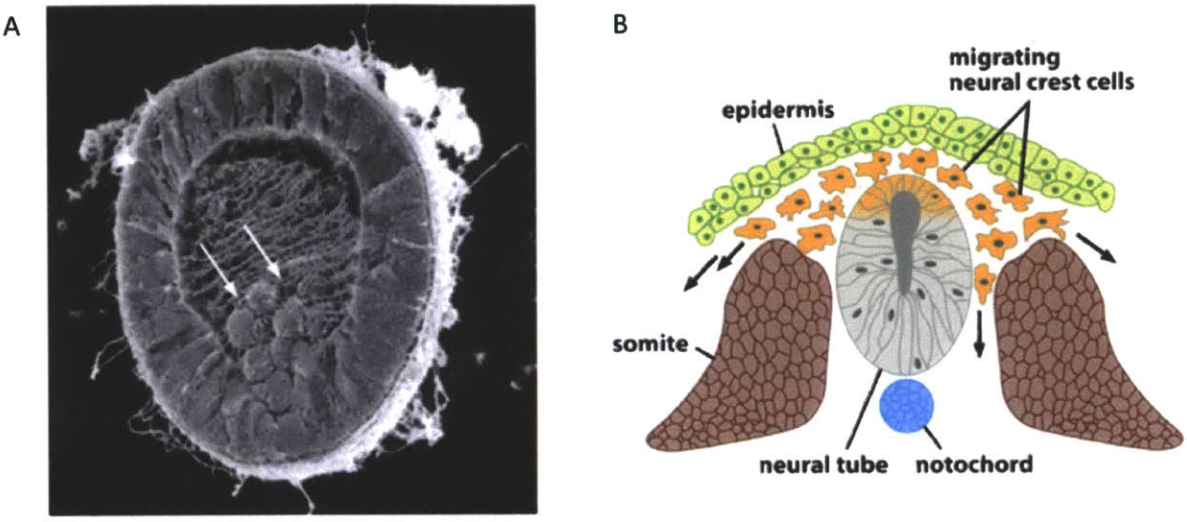


Figure 2. (A) Electron micrograph of sea urchin embryo. Arrows indicate primary mesenchyme cells arising by EMT from the adjacent ectoderm. (B) Neural crest cells arise through EMT of the neural epithelium of the dorsal neural tube and subsequently migrate outwards to form new tissues. From "The Biology of Cancer" by R. Weinberg, 2006, Garland Science, USA.

At the most basic level, an EMT is described by the acquisition of mesenchymal characteristics and loss of epithelial characteristics. Epithelial cells in their natural tissue context form sheets and establish close contacts with their neighbors through adherens junctions, desmosomes and tight junctions, with a well-defined apicobasal axis of polarity. They are also separated from adjacent tissue by a basal lamina, to which they are tightly associated through focal adhesions and hemidesmosomes. The components of these structures can serve as molecular markers of the epithelial state, including desmoplakin (DSP, part of the desmosomes), various claudins (part of the tight junctions) and laminins (part of the basal lamina). E-cadherin, which plays a role in organizing adherens junctions, is the most commonly used marker by virtue of its abundance,

which allows for easy detection of changes in its levels. Growing on 2D culture *in vitro*, the various epithelial structural features are not as well defined, but their components are still detectable. They will also organize into islands of cells on a subconfluent plate, even if initially plated as a single cell suspension, rather than distribute evenly to occupy all available space (Figure 3A).

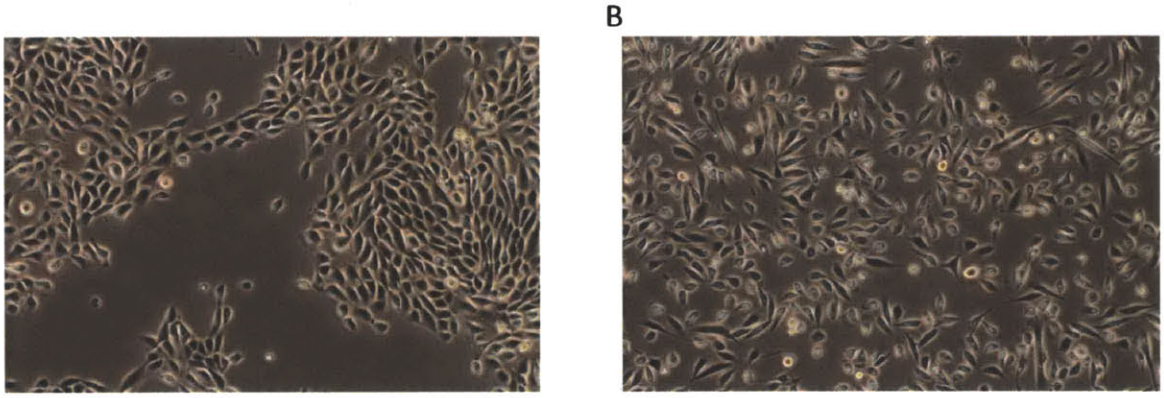


Figure 3. (A) HMLE cells growing on tissue culture plates displaying epithelial morphology. (B) HMLE cells after undergoing ZEB1-induced EMT, displaying mesenchymal morphology.

Mesenchymal cells, on the other hand, are loosely associated in a three-dimensional matrix and do not form organized layers in the tissue context, nor do they form or associate with a basal lamina. Unlike epithelial cells, which derive their structural integrity from their various attachments to their neighboring cells, mesenchymal cells maintain their structural integrity through association with the surrounding matrix, and have a different cytoskeletal composition.

Vimentin, an intermediate filament protein, is a major cytoskeletal component of mesenchymal cells, although it is not exclusive to them. Fibronectin, a component of the extracellular matrix, is also highly expressed by mesenchymal cells. N-cadherin is not strictly a mesenchymal marker, since it is expressed in the epithelium that eventually becomes neural crest cells, and is actually

lost in that EMT process (Pla et al., 2001). However, it often replaces E-cadherin as the main cadherin of the cell prior to or during an EMT, and is found to promote adhesion between invasive breast cancer cells and non-epithelial cells (Hazan et al., 1997, 2000). This "cadherin switch" is also observed in several models of EMT (Hartwell et al., 2006; Yang et al., 2004; Onder et al., 2008; Gravdal et al., 2007).

In 2D culture, mesenchymal cells tend to remain as single cells (Figure 3B), and even if forced to grow to confluency, will still not form adherens junctions, desmosomes and tight junctions with their neighbors. Identification of an EMT occurring in culture can be a simple matter of looking for the presence or absence of islands of cells growing in tight association on a non-confluent plate. However, there are situations where this does not hold. Hepatocyte growth factor (HGF) can in some circumstances cause epithelial islands to break apart, without strongly altering the expression of typical epithelial or mesenchymal markers in those cells (Janda et al., 2002). Conversely, and more commonly, when the transition between epithelial and mesenchymal states is only partial, the change in cell morphology in culture may not be obvious, even though changes in the molecular markers discussed above are detectable.

1.4 EMT in cancer progression

Just as tumor formation involves the deregulation of the existing replicative program of the cell, so the EMT may provide a built-in process for converting tightly bound epithelial cancer cells into loosely associated individual cells with mesenchymal characteristics, able to disseminate more easily to distant sites. However, the role of EMT in cancer was not strongly pursued for a long

time, because evidence was hard to come by in the clinic. Primary carcinomas and their distant metastases tend to look similar and largely retain epithelial morphology. It is this fact that allows pathologists to deduce the origin of occult primary tumors. The presence of epithelial starting and ending points means that any EMT that occurred was likely to be transient and reversible. The difficulty in identifying what may be a transient, partial transition occurring in only a subset of cells in the tumor had long been a point of contention with regard to the relevance of the EMT in cancer metastasis. Some early indications of this tumor heterogeneity came from the observation that in colorectal cancer, β-catenin staining is variable, showing membranous localization in the tumor center and nuclear localization in isolated, scattered cells at the invasive front (Brabletz et al., 1998; Kirchner and Brabletz, 2000).

In recent years, with improved molecular characterization, the case for EMT in cancer has been significantly strengthened (Polyak and Weinberg, 2009; Thiery et al., 2009; Nieto, 2010). Carcinomas with more mesenchymal characteristics have been found to be associated with increased tumor initiation, anoikis resistance, and resistance to chemo- and radiotherapy (Creighton et al., 2009, 2010). In colorectal cancer, partial loss of epithelial characteristics, such as the basement membrane at the edge of the invasive front of tumor, has been correlated with decreased survival and increased chance of distant metastasis (Spaderna et al., 2006). In prostate cancer, loss of epithelial markers and gain of mesenchymal markers is associated with higher Gleason score and clinical recurrence (Gravdal et al., 2007). In model systems, the direct observation of carcinoma cells separating and disseminating as single cells (Giampieri et al., 2009; Wyckoff et al., 2007), as well as the demonstration that triggering EMT enhances metastasis

(Onder et al., 2008), lend credence to the idea that EMT indeed has a role in metastatic progression.

1.5 EMT-inducing factors

In vitro study of the EMT process was advanced with the discovery that Madin-Darby canine kidney (MDCK) cells in 2D-culture underwent the process quite readily (Stoker and Perryman, 1985). These cells, along with NBT-II rat bladder carcinoma cells (Boyer et al., 1989), served as the early *in vitro* models for dissecting the molecular mechanism of the EMT. Many factors have been reported to trigger an EMT. These include transcription factors like ZEB1 (Eger et al., 2005), Snail (Cano et al., 2000) and Twist (Yang et al., 2004), extracellular signaling molecules like TGF-β (Miettinen et al., 1994; Zavadil and Bottinger, 2005) and HGF (Stoker and Perryman, 1985), and even certain conditions of culture such as hypoxia (Cannito et al., 2008; Lundgren et al., 2009). It was in the MDCK system that conditioned medium from cultured fibroblasts was found to induce an EMT (Stoker and Perryman, 1985). Subsequently, the responsible factor in the medium was identified as HGF (Naldini et al., 1991), the first defined EMT-inducing factor.

There are substantial differences in the behaviors of different EMT-inducing signals in different models. TGF-β is able to induce NMuMG murine mammary epithelial cells to undergo an EMT, although this is accompanied by apoptosis in the majority of cells (Gal et al., 2008). In EpH4 murine mammary epithelial cells, TGF-β exposure causes cell cycle arrest and apoptosis, without an EMT. In the Ha-Ras transformed derivative, EpRas, TGF-β-induced EMT can occur without arrest or apoptosis, but HGF cannot induce EMT, unlike in MDCK cells (Janda et al., 2002). Several

other extracellular cues governing the EMT have been uncovered, both in developmental and *in vitro* models (Thiery and Sleeman, 2006). Wnt signaling can induce ectopic neural crest formation in primitive neural epithelium (Wu et al., 2005), and is found to contribute to the maintenance of the mesenchymal phenotype (Scheel et al., 2011), but is not known to be sufficient on its own to induce an EMT in most *in vitro* models. FGF signaling is required for neural crest specification and the specification of mesoderm in gastrulation (Monsoro-Burq et al., 2003; Ciruna and Rossant, 2001). FGF can also induce EMT in NBT-II cells (Vallés et al., 1996). EGF can similarly induce EMT in NBT-II cells (Edme et al., 2002), but its role in developmental EMT is less known.

1.6 EMT-inducing transcription factors

1.6.1 Snail/Slug family

Much work in the past decade has focused on transcription factors that are able to mediate the EMT. FGF signaling was found to work through the C_2H_2-class zinc finger transcription factor Snail to mediate the EMT in the primitive streak (Ciruna and Rossant, 2001). Snail-null mouse embryos die at E8.5 with phenotypes widely indicative of defective mesoderm formation and EMT failure (Carver et al., 2001; Wu and McClay, 2007). Snail is sufficient to trigger the EMT in a variety of epithelial cell lines (Cano et al., 2000). The Slug transcription factor, belonging to the same family as Snail, also features in developmental EMT, marking premigratory neural crest cells in the developing chick embryo and playing a role in their ability to emigrate (Nieto et al., 1994). Slug and Snail expression have been observed to be complementary, occurring in different mesoderm compartments during development, and there is evidence that the role played by Snail in

mammalian EMT is played by Slug in avian EMT (Sefton et al., 1998). Interfering with Slug function in chick embryos results in a failure of mesoderm formation and neural crest migration (Nieto et al., 1994), whereas mice lacking the Slug gene can develop to adulthood. *In vitro*, Slug is sufficient to induce the EMT in MDCK cells (Bolós et al., 2003).

1.6.2 bHLH family

The basic helix-loop-helix (bHLH) transcription factor Twist has been shown to induce EMT *in vitro*, and its knockdown can reduce circulating cancer cells and lung metastases from an orthotopic site in a mouse mammary carcinoma model (Yang et al., 2004). Twist is involved in mesoderm specification in vertebrates (Thisse et al., 1987; Leptin and Grunewald, 1990; Wu et al., 2008; Barnes and Firulli, 2009). Twist-null mutant mouse embryos die around E11 with neural tube closure defects (Chen and Behringer, 1995), while heterozygotes display craniofacial and limb skeletal defects (Bourgeois et al., 1998), phenotypes associated with dysfunctional neural crest cells. E12/E47, a heterodimer composed of alternative splice forms of the E2A gene, is another bHLH factor which has been shown to trigger EMT, but only in MDCK cells (Perez-Moreno et al., 2001). E2A-null mice can survive to term, and the more obvious phenotypes are defects in B- and T-cell development.

1.6.3 ZEB family

The transcription factor ZEB1, the central focus of this study, is a relatively new player in the EMT field. It was first found to play a role in EMT in the EpFosER murine mammary epithelial model system (Eger et al., 2005). When Fos activity was induced by means of a Tamoxifen-inducible Fos-

ER fusion construct, these cells underwent EMT, and δEF1, the murine ortholog of ZEB1, was found among the differentially regulated genes. When ectopically expressed, δEF1 induced an EMT in these cells without the need for Tamoxifen induction of Fos-ER. The same group later found that ZEB1 knockdown could cause mesenchymal MDA-MB-231 cells to express epithelial markers (Aigner et al., 2007a). ZEB1 knockdown also reduced the metastasis-forming ability of the HCT116 colon colorectal cancer cell line (Spaderna et al., 2008).

In the developmental context, ZEB1 is expressed in the notochord, somites, skeletal and limb precursors and several neural crest derivatives (Funahashi et al., 1993; Takagi et al., 1998). Mutant mouse embryos that are null for δEF1 develop to term but die shortly after birth. The mutants display growth retardation and shortened bones, as well as underdeveloped thymi, but do not display defects that would suggest generalized or widespread failure of EMT during early development. This may be one of the reasons why, unlike the other EMT-related transcription factors, ZEB1 was first implicated in EMT *in vitro* rather than in the developmental context. ZEB1 represents only one half of a family of transcription factors, the other being ZEB2, which shares the same general structural features and targets the same DNA binding sequence. The null phenotype for Sip1, the murine ZEB2 ortholog, is more severe than for δEF1 (Van de Putte et al., 2003). Embryos arrest at E8.5 and the neural tube fails to close. Neural crest cells either fail to migrate or fail to form at all, suggesting that EMT failure is more pronounced in these mutants. ZEB1 and ZEB2 orthologs in mice are appear to be expressed in a largely complementary pattern during development, with little overlap between the two (Miyoshi et al., 2006). Double null mutants for both δEF1 and Sip1 closely resemble Sip1 mutants, and arrest at the same stage.

However, while heterozygotes of Sip1 are healthy, compound mutants that are heterozygous for Sip1 and null for δEF1 die earlier than δEF1-null single mutants. The compound mutants additionally display exaggerated versions of mild defects found in the single mutants. This suggests that the two factors may cooperate and have additive effects, notwithstanding their apparently complementary expression.

1.6.4 Molecular interactions of ZEB1

At the molecular level, the atomic structure of ZEB1 is yet to be defined, but sequence analysis shows that it contains zinc-finger clusters in both the N-terminal and C-terminal regions, as well as a homeodomain in the central region. The zinc-fingers play the primary role in DNA-binding, selectively targeting an E-box sequence CACCTG, but capable of binding similar sequences to varying extents (Postigo, 2000). ZEB1 is capable of acting as both a transcriptional repressor and activator (Ikeda and Kawakami, 1995), although most present work on ZEB1 focuses on its repressive role only. The best-known target gene of ZEB1 is E-cadherin. ZEB1 is shown to bind to its promoter, repressing its expression (Eger et al., 2005). ZEB1 also directly binds and represses the promoters of several other epithelial cell components such as Crumbs3 and Pals1-associated tight junction protein (PATJ) (Aigner et al., 2007b, 2007a). These observations collectively reinforce the impression that ZEB1 mediates the EMT largely through its repression of epithelial genes.

TGF-β and NF-κB have been implicated in upregulating ZEB1 (Shirakihara et al., 2007; Chua et al., 2007), but direct regulation of ZEB1 is not well known, with the exception of the relationship

between ZEB1 and the miR-200 family of microRNAs. The miR-200 family comprises five members grouped into two clusters by genomic location(miR-200a, miR-200b and miR-429 in one, miR-200c and miR-141 in the other), and also by seed sequence (miR-200a and miR-141 in one, miR-200b, miR-200c and miR-429 in the other).

Several different groups independently showed that members of this family are downregulated in mesenchymal cells (Park et al., 2008; Gregory et al., 2008; Bracken et al., 2008; Burk et al., 2008), and that they may exert RNA interference (RNAi) on ZEB1 through seed sequences found in the 3' untranslated region (UTR) of the ZEB1 mRNA, primarily through the miR-200b/c/429 seed sequence. Two of these groups further showed that ZEB1 itself can repress expression of the miR-200 family (Burk et al., 2008; Bracken et al., 2008) by binding to the promoter regions of the two clusters. This reciprocal negative regulation sets up a control loop that, under certain conditions, can behave like a bistable switch that self-reinforces an existing state. Increasing the levels of ZEB1, for example, results in repression of miR-200 members, which leads to decreased RNAi on the ZEB1 mRNA, allowing increased levels of ZEB1 protein. Conversely, increasing the levels of miR-200 lowers ZEB1 levels through RNAi, which in turn decreases repression of miR-200 promoters, allowing further increase in miR-200 expression. Note that perturbation of either member does not actively cause an increased transcription of the other. So for example, decreasing levels of ZEB1 would not automatically result in an increase in miR-200 if there was no transcriptional activation of miR-200 expression to begin with (Cochrane et al., 2010). Increasing miR-200 levels reduces ZEB1 protein and transcript levels. but is not known to directly affect ZEB1 at the level of transcription initiation.

1.7 A common circuit for the EMT

The multitude of signals capable of activating EMT programs begs the question of whether the processes being observed are in fact entirely different processes bearing only a superficial resemblance, or whether they share deeper connections, representing diverse manifestations of a common cell-biological program. During the course of my work, Snail and ZEB1 became arguably the most extensively studied inducing factors with respect to the EMT. Both are transcription factors known to directly bind to and repress the E-cadherin promoter, and both bind to E-box sequences (CANNTG). Although this similarity eventually led to some speculation that they may play similar roles in EMT induction (Peinado et al., 2007), at the beginning of my research, there was little appreciation for the possibility that EMT-inducing factors may interact with one another, act redundantly, or impinge on a common regulatory pathway. Each factor was studied in isolation. ZEB1 had only just been discovered to play a role in the EMT, and was simply considered yet another of a number of EMT-inducing factors. In the course of searching for a common circuit for the EMT mediated by several different factors, I independently discovered a role for ZEB1 in the EMT model I was using and developed evidence for how ZEB1 was qualitatively different from other EMT-inducing factors known to work in the same system.

1.8 References

Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., Mikulits, W., Brabletz, T., Strand, D., Obrist, P., Sommergruber, W., Schweifer, N., Wernitznig, A., Beug, H., Foisner, R., and Eger, A. (2007a). The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. Oncogene.

Aigner, K., Descovich, L., Mikula, M., Sultan, A., Dampier, B., Bonne, S., van Roy, F., Mikulits, W., Schreiber, M., Brabletz, T., Sommergruber, W., Schweifer, N., Wernitznig, A., Beug, H., Foisner, R., and Eger, A. (2007b). The transcription factor ZEB1 (deltaEF1) represses Plakophilin 3 during human cancer progression. FEBS Lett *581*, 1617–1624.

Barnes, R. M., and Firulli, A. B. (2009). A twist of insight - the role of Twist-family bHLH factors in development. The International journal of developmental biology *53*, 909–924.

Bell, D., Berchuck, A., Birrer, M., Chien, J., Cramer, D. W., et al. (2011). Integrated genomic analyses of ovarian carcinoma. Nature *474*, 609–615.

Berger, M. F., Levin, J. Z., Vijayendran, K., Sivachenko, A., Adiconis, X., Maguire, J., Johnson, L. A., Robinson, J., Verhaak, R. G., Sougnez, C., Onofrio, R. C., Ziaugra, L., Cibulskis, K., Laine, E., Barretina, J., Winckler, W., Fisher, D. E., Getz, G., Meyerson, M., Jaffe, D. B., Gabriel, S. B., Lander, E. S., Dummer, R., Gnirke, A., Nusbaum, C., and Garraway, L. A. (2010). Integrative analysis of the melanoma transcriptome. Genome research *20*, 413–427.

Bignell, G. R., Greenman, C. D., Davies, H., Butler, A. P., Edkins, S., Andrews, J. M., Buck, G., Chen, L., Beare, D., Latimer, C., Widaa, S., Hinton, J., Fahey, C., Fu, B., Swamy, S., Dalgliesh, G. L., Teh, B. T., Deloukas, P., Yang, F., Campbell, P. J., Futreal, P. A., and Stratton, M. R. (2010). Signatures of mutation and selection in the cancer genome. Nature *463*, 893–898.

Bolós, V., Peinado, H., Pérez-Moreno, M. A., Fraga, M. F., Esteller, M., and Cano, A. (2003). The transcription factor Slug represses E-cadherin expression and induces epithelial to mesenchymal transitions: a comparison with Snail and E47 repressors. Journal of cell science *116*, 499–511.

Bourgeois, P., Bolcato-Bellemin, A. L., Danse, J. M., Bloch-Zupan, A., Yoshiba, K., Stoetzel, C., and Perrin-Schmitt, F. (1998). The variable expressivity and incomplete penetrance of the twist-null heterozygous mouse phenotype resemble those of human Saethre-Chotzen syndrome. Human molecular genetics *7*, 945–957.

Boyer, B., Tucker, G. C., Vallés, A. M., Franke, W. W., and Thiery, J. P. (1989). Rearrangements of desmosomal and cytoskeletal proteins during the transition from epithelial to fibroblastoid organization in cultured rat bladder carcinoma cells. The Journal of cell biology *109*, 1495–1509.

Brabletz, T., Jung, A., Hermann, K., Günther, K., Hohenberger, W., and Kirchner, T. (1998). Nuclear overexpression of the oncoprotein beta-catenin in colorectal cancer is localized predominantly at the invasion front. Pathology, research and practice *194*, 701–704.

Bracken, C. P., Gregory, P. A., Kolesnikoff, N., Bert, A. G., Wang, J., Shannon, M. F., and Goodall, G. J. (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. Cancer research *68*, 7846–7854.

Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S., and Brabletz, T. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep *9*, 582–589.

Cannito, S., Novo, E., Compagnone, A., Valfrè di Bonzo, L., Busletta, C., Zamara, E., Paternostro, C., Povero, D., Bandino, A., Bozzo, F., Cravanzola, C., Bravoco, V., Colombatto, S., and Parola, M. (2008). Redox mechanisms switch on hypoxia-dependent epithelial-mesenchymal transition in cancer cells. Carcinogenesis *29*, 2267–2278.

Cano, A., Perez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F., and Nieto, M. A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol *2*, 76–83.

Carver, E. A., Jiang, R., Lan, Y., Oram, K. F., and Gridley, T. (2001). The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. Molecular and cellular biology *21*, 8184–8188.

Center, M., Siegel, R., and Jemal, A. (2011). Global Cancer Facs & Figures 2nd Edition (American Cancer Society, Atlanta).

Chambers, A. F., Groom, A. C., and MacDonald, I. C. (2002). Metastasis: Dissemination and growth of cancer cells in metastatic sites. Nat Rev Cancer *2*, 563–572.

Chen, Z. F., and Behringer, R. R. (1995). twist is required in head mesenchyme for cranial neural tube morphogenesis. Genes & development *9*, 686–699.

Chin, L., Hahn, W. C., Getz, G., and Meyerson, M. (2011). Making sense of cancer genomic data. Genes & development *25*, 534–555.

Chua, H. L. L., Bhat-Nakshatri, P., Clare, S. E. E., Morimiya, A., Badve, S., and Nakshatri, H. (2007). NF-kappaB represses E-cadherin expression and enhances epithelial to mesenchymal transition of mammary epithelial cells: potential involvement of ZEB-1 and ZEB-2. Oncogene *26*, 711–724.

Ciruna, B., and Rossant, J. (2001). FGF signaling regulates mesoderm cell fate specification and morphogenetic movement at the primitive streak. Developmental cell *1*, 37–49.

Cochrane, D. R., Howe, E. N., Spoelstra, N. S., and Richer, J. K. (2010). Loss of miR-200c: A Marker of Aggressiveness and Chemoresistance in Female Reproductive Cancers. Journal of oncology *2010*, 821717.

Creighton, C. J., Chang, J. C., and Rosen, J. M. (2010). Epithelial-mesenchymal transition (EMT) in tumor-initiating cells and its clinical implications in breast cancer. Journal of mammary gland biology and neoplasia *15*, 253–260.

Creighton, C. J., Li, X., Landis, M., Dixon, J. M., Neumeister, V. M., Sjolund, A., Rimm, D. L., Wong, H., Rodriguez, A., Herschkowitz, J. I., Fan, C., Zhang, X., He, X., Pavlick, A., Gutierrez, M. C., Renshaw, L., Larionov, A. A., Faratian, D., Hilsenbeck, S. G., Perou, C. M., Lewis, M. T., Rosen, J. M., and Chang, J. C. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proceedings of the National Academy of Sciences of the United States of America *106*, 13820–13825.

Cristofanilli, M., Budd, G. T., Ellis, M. J., Stopeck, A., Matera, J., Miller, M. C., Reuben, J. M., Doyle, G. V., Allard, W. J., Terstappen, L. W. M. M., and Hayes, D. F. (2004). Circulating Tumor Cells, Disease Progression, and Survival in Metastatic Breast Cancer. New England Journal of Medicine *351*, 781–791.

Duband, J. L., Monier, F., Delannet, M., and Newgreen, D. (1995). Epithelium-mesenchyme transition during neural crest development. Acta anatomica *154*, 63–78.

Edme, N., Downward, J., Thiery, J.-P., and Boyer, B. (2002). Ras induces NBT-II epithelial cell scattering through the coordinate activities of Rac and MAPK pathways. Journal of cell science *115*, 2591–2601.

Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., Berx, G., Cano, A., Beug, H., and Foisner, R. (2005). DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. Oncogene *24*, 2375–2385.

Funahashi, J., Sekido, R., Murai, K., Kamachi, Y., and Kondoh, H. (1993). Delta-crystallin enhancer binding protein delta EF1 is a zinc finger-homeodomain protein implicated in postgastrulation embryogenesis. Development (Cambridge, England) *119*, 433–446.

Gal, A., Sjöblom, T., Fedorova, L., Imreh, S., Beug, H., and Moustakas, A. (2008). Sustained TGF beta exposure suppresses Smad and non-Smad signalling in mammary epithelial cells, leading to EMT and inhibition of growth arrest and apoptosis. Oncogene *27*, 1218–1230.

Giampieri, S., Manning, C., Hooper, S., Jones, L., Hill, C. S., and Sahai, E. (2009). Localized and reversible TGFbeta signalling switches breast cancer cells from cohesive to single cell motility. Nature cell biology *11*, 1287–1296.

Gravdal, K., Halvorsen, O. J. J., Haukaas, S. A. A., and Akslen, L. A. A. (2007). A switch from E-cadherin to N-cadherin expression indicates epithelial to mesenchymal transition and is of strong and independent importance for the progress of prostate cancer. Clinical cancer research : an official journal of the American Association for Cancer Research *13*, 7003–7011.

Gregory, P. A., Bert, A. G., Paterson, E. L., Barry, S. C., Tsykin, A., Farshid, G., Vadas, M. A., Khew-Goodall, Y., and Goodall, G. J. (2008). The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. Nature cell biology *10*, 593–601.

Gupta, G. P., Massague, J., and Massagué, J. (2006). Cancer metastasis: building a framework. Cell *127*, 679–695.

Hanahan, D., and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Hartwell, K. A., Muir, B., Reinhardt, F., Carpenter, A. E., Sgroi, D. C., and Weinberg, R. A. (2006). The Spemann organizer gene, Goosecoid, promotes tumor metastasis. Proc Natl Acad Sci U S A *103*, 18969–18974.

Hay, E. D. (1968). Organization and fine structure of epithelium andmesenchyme in the developing chick embryos. In Epithelial-Mesenchymal Interactions: 18th Hahnemann Symposium, R. Fleischmajer and R. Billingham, eds. (Baltimore: Williams & Wilkins), pp. 31–35.

Hazan, R. B. B., Kang, L., Whooley, B. P. P., and Borgen, P. I. I. (1997). N-cadherin promotes adhesion between invasive breast cancer cells and the stroma. Cell adhesion and communication *4*, 399–411.

Hazan, R. B. B., Phillips, G. R. R., Qiao, R. F. F., Norton, L., and Aaronson, S. A. A. (2000). Exogenous expression of N-cadherin in breast cancer cells induces cell migration, invasion, and metastasis. The Journal of cell biology *148*, 779–790.

Ikeda, K., and Kawakami, K. (1995). DNA binding through distinct domains of zinc-finger-homeodomain protein AREB6 has different effects on gene transcription. European journal of biochemistry / FEBS *233*, 73–82.

Janda, E., Lehmann, K., Killisch, I., Jechlinger, M., Herzig, M., Downward, J., Beug, H., and Grünert, S. (2002). Ras and TGF[beta] cooperatively regulate epithelial cell plasticity and metastasis: dissection of Ras signaling pathways. The Journal of cell biology *156*, 299–313.

Katow, H., and Solursh, M. (1980). Ultrastructure of primary mesenchyme cell ingression in the sea urchinLytechinus pictus. Journal of Experimental Zoology *213*, 231–246.

Kirchner, T., and Brabletz, T. (2000). Patterning and nuclear beta-catenin expression in the colonic adenoma-carcinoma sequence. Analogies with embryonic gastrulation. The American journal of pathology *157*, 1113–1121.

Knecht, A. K., and Bronner-Fraser, M. (2002). Induction of the neural crest: a multigene process. Nature reviews. Genetics *3*, 453–461.

Leptin, M., and Grunewald, B. (1990). Cell shape changes during gastrulation in Drosophila. Development (Cambridge, England) *110*, 73–84.

Lillie, F. R. (1908). The Development of the Chick (Henry Holt and Co, New York).

Lundgren, K., Nordenskjold, B., and Landberg, G. (2009). Hypoxia, Snail and incomplete epithelial-mesenchymal transition in breast cancer. Br J Cancer *101*, 1769–1781.

Luzzi, K. J., MacDonald, I. C., Schmidt, E. E., Kerkvliet, N., Morris, V. L., Chambers, A. F., and Groom, A. C. (1998). Multistep nature of metastatic inefficiency: dormancy of solitary cells after successful extravasation and limited survival of early micrometastases. The American journal of pathology *153*, 865–873.

Mancilla, A., and Mayor, R. (1996). Neural crest formation in Xenopus laevis: mechanisms of Xslug induction. Developmental biology *177*, 580–589.

Miettinen, P. J., Ebner, R., Lopez, A. R., and Derynck, R. (1994). TGF-beta induced transdifferentiation of mammary epithelial cells to mesenchymal cells: involvement of type I receptors. . The Journal of Cell Biology *127* , 2021–2036.

Miyoshi, T., Maruhashi, M., Van De Putte, T., Kondoh, H., Huylebroeck, D., and Higashi, Y. (2006). Complementary expression pattern of Zfhx1 genes Sip1 and deltaEF1 in the mouse embryo and their genetic interaction revealed by compound mutants. Developmental dynamics : an official publication of the American Association of Anatomists *235*, 1941–1952.

Monsoro-Burq, A.-H., Fletcher, R. B., and Harland, R. M. (2003). Neural crest induction by paraxial mesoderm in Xenopus embryos requires FGF signals. Development (Cambridge, England) *130*, 3111–3124.

Naldini, L., Weidner, K. M., Vigna, E., Gaudino, G., Bardelli, A., Ponzetto, C., Narsimhan, R. P., Hartmann, G., Zarnegar, R., and Michalopoulos, G. K. (1991). Scatter factor and hepatocyte growth factor are indistinguishable ligands for the MET receptor. The EMBO journal *10*, 2867–2878.

Nieto, M. A. (2010). The Ins and Outs of the Epithelial to Mesenchymal Transition in Health and Disease. Annual review of cell and developmental biology *27*, 347–376.

Nieto, M. A., Sargent, M. G., Wilkinson, D. G., and Cooke, J. (1994). Control of cell behavior during vertebrate development by Slug, a zinc finger gene. Science (New York, N.Y.) *264*, 835–839.

Onder, T. T., Gupta, P. B., Mani, S. A., Yang, J., Lander, E. S., and Weinberg, R. A. (2008). Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. Cancer Res *68*, 3645–3654.

Park, S.-M., Gaur, A. B., Lengyel, E., and Peter, M. E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. Genes & development *22*, 894–907.

Peinado, H., Olmeda, D., and Cano, A. (2007). Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? Nature reviews. Cancer *7*, 415–428.

Perez-Moreno, M. A., Locascio, A., Rodrigo, I., Dhondt, G., Portillo, F., Nieto, M. A., and Cano, A. (2001). A new role for E12/E47 in the repression of E-cadherin expression and epithelial-mesenchymal transitions. J Biol Chem *276*, 27424–27431.

Pla, P., Moore, R., Morali, O. G. G., Grille, S., Martinozzi, S., Delmas, V., and Larue, L. (2001). Cadherins in neural crest cell development and transformation. Journal of cellular physiology *189*, 121–132.

Polyak, K., and Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nat Rev Cancer *9*, 265–273.

Postigo, A. (2000). Differential expression and function of members of the zfh-1 family of zinc finger/homeodomain repressors. Proceedings of the National *97*, 6391–6396.

Van de Putte, T., Maruhashi, M., Francis, A., Nelles, L., Kondoh, H., Huylebroeck, D., and Higashi, Y. (2003). Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. American journal of human genetics *72*, 465–470.

Ramón y Cajal, S. (1890). Manual de Anatomía Patológica General (Casa Provincial Caridad, Barcelona).

Scheel, C., Eaton, E. N., Li, S. H.-J., Chaffer, C. L., Reinhardt, F., Kah, K.-J., Bell, G., Guo, W., Rubin, J., Richardson, A. L., and Weinberg, R. A. (2011). Paracrine and autocrine signals induce and maintain mesenchymal and stem cell States in the breast. Cell *145*, 926–940.

Sefton, M., Sanchez, S., and Nieto, M. (1998). Conserved and divergent roles for members of the Snail family of transcription factors in the chick and mouse embryo. Development *125*, 3111–3121.

Selleck, M. A., and Bronner-Fraser, M. (1995). Origins of the avian neural crest: the role of neural plate-epidermal interactions. Development (Cambridge, England) *121*, 525–538.

Shirakihara, T., Saitoh, M., and Miyazono, K. (2007). Differential regulation of epithelial and mesenchymal markers by deltaEF1 proteins in epithelial mesenchymal transition induced by TGF-beta. Molecular biology of the cell *18*, 3533–3544.

Siegel, R., Naishadham, D., and Jemal, A. (2011). Cancer statistics, 2012. CA: A Cancer Journal for Clinicians *00*, n/a-n/a.

Sjöblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K. V., Gazdar, A. F., Hartigan, J., Wu, L., Liu, C., Parmigiani, G., Park, B. H., Bachman, K. E., Papadopoulos, N., Vogelstein, B., Kinzler, K. W., and Velculescu, V. E. (2006). The consensus coding sequences of human breast and colorectal cancers. Science (New York, N.Y.) *314*, 268–274.

Spaderna, S., Schmalhofer, O., Hlubek, F., Berx, G., Eger, A., Merkel, S., Jung, A., Kirchner, T., and Brabletz, T. (2006). A Transient, EMT-Linked Loss of Basement Membranes Indicates Metastasis and Poor Survival in Colorectal Cancer. Gastroenterology *131*, 830–840.

Spaderna, S., Schmalhofer, O., Wahlbuhl, M., Dimmler, A., Bauer, K., Sultan, A., Hlubek, F., Jung, A., Strand, D., Eger, A., Kirchner, T., Behrens, J., and Brabletz, T. (2008). The transcriptional repressor ZEB1 promotes metastasis and loss of cell polarity in cancer. Cancer Res *68*, 537–544.

Steeg, P. S. (2006). Tumor metastasis: mechanistic insights and clinical challenges. Nature medicine *12*, 895–904.

Stoker, M., and Perryman, M. (1985). An epithelial scatter factor released by embryo fibroblasts. Journal of cell science *77*, 209–223.

Takagi, T., Moribe, H., Kondoh, H., and Higashi, Y. (1998). DeltaEF1, a zinc finger and homeodomain transcription factor, is required for skeleton patterning in multiple lineages. Development (Cambridge, England) *125*, 21–31.

Thiery, J. P., Acloque, H., Huang, R. Y. J., and Nieto, M. A. (2009). Epithelial-mesenchymal transitions in development and disease. Cell *139*, 871–890.

Thiery, J. P., and Sleeman, J. P. (2006). Complex networks orchestrate epithelial-mesenchymal transitions. Nature reviews. Molecular cell biology *7*, 131–142.

Thisse, B., el Messal, M., and Perrin-Schmitt, F. (1987). The twist gene: isolation of a Drosophila zygotic gene necessary for the establishment of dorsoventral pattern. Nucleic acids research *15*, 3439–3453.

Vaage, J. (1989). A survey of in vivo growth characteristics and spontaneous metastasizing potentials of C3H mouse mammary tumors. International journal of cancer. Journal international du cancer *43*, 910–914.

Vallés, A. M., Boyer, B., Tarone, G., and Thiery, J. P. (1996). Alpha 2 beta 1 integrin is required for the collagen and FGF-1 induced cell dispersion in a rat bladder carcinoma cell line. Cell adhesion and communication *4*, 187–199.

Viebahn, C. (1995). Epithelio-mesenchymal transformation during formation of the mesoderm in the mammalian embryo. Acta anatomica *154*, 79–97.

Vogelstein, B., and Kinzler, K. W. (1993). The multistep nature of cancer. Trends in genetics : TIG *9*, 138–141.

Weinberg, R. A. (2006). The Biology of Cancer 1st ed. (Garland Science).

Wu, J., Yang, J., and Klein, P. S. (2005). Neural crest induction by the canonical Wnt pathway can be dissociated from anterior-posterior neural patterning in Xenopus. Developmental biology *279*, 220–232.

Wu, S.-Y., Yang, Y.-P., and McClay, D. R. (2008). Twist is an essential regulator of the skeletogenic gene regulatory network in the sea urchin embryo. Developmental biology *319*, 406–415.

Wu, S.-Y., and McClay, D. R. (2007). The Snail repressor is required for PMC ingression in the sea urchin embryo. Development (Cambridge, England) *134*, 1061–1070.

Wyckoff, J. B., Wang, Y., Lin, E. Y., Li, J.-feng, Goswami, S., Stanley, E. R., Segall, J. E., Pollard, J. W., and Condeelis, J. (2007). Direct visualization of macrophage-assisted tumor cell intravasation in mammary tumors. Cancer research *67*, 2649–2656.

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927–939.

Zavadil, J., and Bottinger, E. P. (2005). TGF-[beta] and epithelial-to-mesenchymal transitions. Oncogene *24*, 5764–5774.

Chapter 2:

Creation and Analysis of the EMT Core Signature

2.1 Introduction

The EMT-inducing signals described in Chapter 1 function in different experimental models, each with its own starting and ending points, which has made cross-comparison problematic and complicated the search for a common regulatory circuit. The HMLE cell line (Elenbaas et al., 2001) serves as a useful experimental model, because it is able to undergo EMT in response to a number of different signals, including the overexpression of TGF-β1, Snail, Twist (Yang et al., 2004), Goosecoid (Hartwell et al., 2006) and the knockdown of E-cadherin (Onder et al., 2008). This cell line was originally created by immortalizing normal human mammary epithelial cells through the introduction of the *hTERT* gene, which specifies the catalytic component of telomerase, as well as the SV40 early region, whose products inactivate the p53 and RB tumor suppressor proteins. Only MDCK cells can accept a comparably broad range of inputs to trigger the EMT, and the canine origin of those cells poses technical challenges, particularly in microarray gene expression profiling. Exploiting the fact that the various mesenchymal derivatives of the HMLE cells have the same starting point, one could eliminate cell-type-specific differences that would plague analysis of, for example, the Twist-EMT signature in one cell type versus the Snail-EMT signature in a different cell type.

Prior to this study, it was known that the HMLE mesenchymal derivatives all downregulated E-cadherin, upregulated vimentin, and changed their cell morphology from epithelial to mesenchymal, but it was unclear if they were achieving this through separate pathways, or if these responses reflected the actions of a common underlying regulatory circuit. Microarray gene expression profiling, data for which were subsequently published (Onder et al., 2008; Taube et al.,

2010), was used to extract a common gene signature from all five mesenchymal derivatives compared to the epithelial controls. Analysis of the signature revealed that ZEB1, hitherto untested in HMLE cells, might play a role in mediating the EMT of all of the other five inducing factors.

2.2 Results and Discussion

2.2.1 Generating the EMT core signature

HMLE immortalized human mammary epithelial cells can be made to undergo EMT using several different inducing factors. Stable mesenchymal derivatives previously created include cells overexpressing Twist (Yang et al., 2004), Snail, Goosecoid (Hartwell et al., 2006), TGF-β1 and an shRNA against E-cadherin (Onder et al., 2008). I wanted to test the hypothesis that the various derivative lines, upon passing through an EMT, shared more than a superficial morphological similarity. At the time, this was assumed to be true, but little data existed to demonstrate this beyond the shared behavior of E-cadherin, vimentin and a few other molecular markers.

Microarray gene expression profiling was carried out to provide an unbiased description of the transcriptional changes that these various cells had undergone in response to the EMT-inducing signals (Onder et al., 2008; Taube et al., 2010). The HMLE-TGF-β1, HMLE-Goosecoid, HMLE-Twist and HMLE-Snail cells were compared to a HMLE-GFP vector control (Taube et al., 2010), while the HMLE-shRNA-E-cadherin cells were compared to a HMLE-shRNA-GFP vector control (Onder et al., 2008). For each pairwise comparison between a mesenchymal line and its epithelial control, the

number of significantly differentially regulated probes ranged from 2173 in the HMLE-TGF-β1 line to 6980 in the HMLE-Goosecoid line (Table 1).

EMT-inducing factor	Goosecoid	Snail	Twist	TGF-β1	shRNA E-cadherin	Expected overlap (E)	Observed overlap (O)	% signal $\frac{E-O}{O} \times 100\%$
Upregulated (probes)	3577	2333	1896	1158	3551	4.78	320	98.5
Fraction upregulated	0.331	0.216	0.176	0.107	0.329	0.000443	0.0296	-
Downregulated (probes)	3403	2027	1723	1015	2799	2.48	304	99.2
Fraction downregulated	0.315	0.188	0.160	0.094	0.259	0.000230	0.0281	-

Table 1. Observed and expected number of differentially regulated probes obtained from five different mesenchymal HMLE cell lines versus epithelial control cells.

Approximately half the probes present on the microarray (10,800 out of 22,277 probes) were significantly differentially regulated in at least one of these mesenchymal lines relative to the epithelial control line. This large number was in fact anticipated, given that each overexpressed protein or shRNA was likely to have exerted certain effects on cells that were unique to itself and unrelated to the EMT program. However, when assessing commonly shared changes among the five mesenchymal cell populations, I found 320 probes that were consistently upregulated and 304 that were consistently downregulated (Table 1).

I took into account the fact that any comparison of two or more gene sets is expected to produce a certain overlap, if only by chance. Accordingly, I compared the size of the experimentally obtained overlap with that predicted to occur by chance. The theoretical expected overlap is the product of the fraction of differentially regulated probes in all five lines, multiplied by the total number of probes (22,277). This calculation resulted in a value of 0.264 expected upregulated

probes and 0.137 expected downregulated probes, meaning not a single probe was expected to be shared in common by all five mesenchymal lines by chance.

Independent of this consideration, I note that, in general for a given cell type, there are genes that are permanently silenced and will never appear in any experimentally obtained gene signature. This reduces the effective total population of probes that can be perturbed and increases the apparent fraction of differentially regulated probes. Failure to account for this effect resulted in smaller expected overlaps. Taking the most conservative position that the total population of alterable probes (and corresponding genes) is 10,800 (the number of probes significantly altered in at least one mesenchymal line), I predicted 4.78 upregulated and 2.48 downregulated probes to overlap in all five mesenchymal lines by chance, which is still far fewer than the 320 upregulated and 304 downregulated probes that I actually observed (Table 1). Hence, the shared set of probes was greatly enriched over the number expected by chance alone, and over 98% of the shared probes were expected to be informative (i.e. not due to chance). I interpreted this to mean that a common biological program underlies the mesenchymal state induced in all five lines. Henceforth, the 624 overlapping up- and down-regulated probes are collectively referred to as the "EMT core signature".

Hierarchical clustering of the gene expression data filtered for the EMT core signature provided a rough guide to the relationships between the five mesenchymal lines and their epithelial counterpart. Clustering by probes separated the commonly upregulated from commonly downregulated probes, as expected (Figure 1). When clustering by sample, the Goosecoid-induced mesenchymal line was the most distantly related among the five, while the Snail- and Twist-induced mesenchymal lines were far more closely related to one another. A priori, I cannot

say which factor is a "better" inducer of EMT, because there is no perfect "reference EMT signature" to which these data could be compared, to say that any one of these samples is closer to the reference than any other.

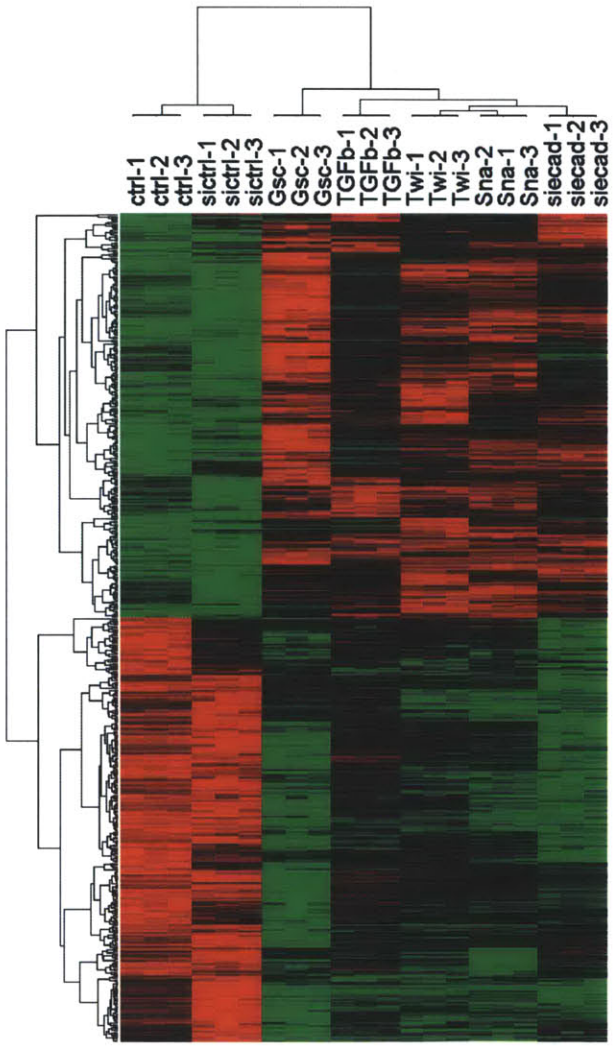


Figure 1. Hierarchical clustering of gene expression microarray data from five mesenchymal HMLE lines and two epithelial HMLE control lines. Clustering by sample (dendrogram on top) reveals that the Gsc-induced mesenchymal line is more distantly related to the other four lines. Clustering by probeset (dendrogram on left) separates genes upregulated across all five lines from those downregulated across all five lines.

2.2.2 Comparing EMT core signature to publicly available gene signatures

In order to further understand the EMT program, I undertook to narrow down the list of candidate genes of special interest, doing so by searching for connections between the EMT core signature described above and preclinical models or clinical data of tumor progression. While many studies purport to identify gene signatures of metastasis or tumor progression using gene expression profiling, a great majority of the studies have analyzed bulk tumors, and hence included undefined levels of stromal cells in addition to cancer cells (Ramaswamy et al., 2003; van t Veer et al., 2002; Wang et al., 2005). While this is acceptable for the purposes of prognosis, it is not ideal for the identification of biological processes happening within the cancer cells themselves.

Four different signatures, representing different aspects of tumor progression and generated from different tissue types, were compared to the EMT core signature. All four studies used the same microarray platform as that used to generate the EMT core signature, allowing for easier comparison, since it is known that cross-platform comparison is less reliable (Irizarry et al., 2005). The first signature represented here is the differentially regulated genes between primary oral squamous cell carcinomas tumors from patients that were either lymph node positive or lymph node negative (O'Donnell et al., 2005). The second represented expression differences between two isogenic human colorectal cell lines, one derived from an invasive primary tumor and the other from a lymph node metastasis from the same individual (Provenzani et al., 2006). The third is the difference between matched pairs of invasive ductal carcinoma (IDC) and ductal carcinoma *in situ* (DCIS), using laser capture microdissection to minimize stromal cell contamination of the sample (Schuetz et al., 2006). The last of these derives from a comparison between unmatched

melanoma metastases and primary tumors (Jaeger et al., 2007), also harvested using laser capture microdissection.

The EMT core signature represents approximately 2.8% of the probes on the microarray. A simple comparison with a random signature is therefore expected to result in an overlap of about 2.8%, purely by mathematical chance. The three comparisons to the signatures of Provenzani et al., Schuetz et al. and Jaeger et al. produced levels of overlap 3.7–7.2 times the mathematically expected result (Table 2). The fact that the EMT core signature showed a high overlap with the latter three signatures suggested that parts of the EMT process may be involved in turning localized tumor regions invasive, and also appeared as differences between primary tumors and their metastases. The weak overlap with the first signature may indicate the difficulty in searching for EMT by looking at whole, unfractionated primary tumors, lending credence to the idea that EMT in clinical cases is a localized phenomenon occurring only in some parts of the tumor rather than across the whole tumor.

	Total size	Expected overlap (E)	Observed overlap (O)	Consistent (C)	Inconsistent (I)	% consistent $\frac{C}{O} \times 100\%$	% signal $\frac{C-I}{C} \times 100\%$
Squamous oral carcinoma, metastatic primary vs. nonmetastatic primary (O'Donnell et al., 2005)	119	3.33	5	5	0	100	100
Colorectal carcinoma cell line from metastasis vs. invasive primary (Provenzani et al., 2006)	378	10.6	39	26	13	66.7	50
Breast invasive ductal carcinoma vs. ductal carcinoma *in-situ* (Schuetz et al., 2006)	546	15.3	88	78	10	88.6	87.2
Melanoma, metastases vs. primary tumors (Jaeger et al., 2007)	389	10.9	78	73	5	93.6	93.2

Table 2. Comparison of microarray probes between EMT core signature and other published gene signatures. The breast carcinoma and melanoma signature overlaps were several fold higher than expected by chance, and highly consistent.

In order to further stratify the results, another metric could be used to assess the quality of the overlap. Every gene in a gene signature is either up- or down-regulated. A signature of randomly up- or down-regulated genes may overlap by chance with the reference set, in this case the EMT core signature, but the further chance that the up or down movement of gene activity agrees with the existing signature is only 50%. A biologically meaningful overlap between two gene signatures should result in a high percentage of the overlapping genes moving in a consistent direction. In this case, since the EMT is hypothesized to play a role in advancing tumor progression, a gene that is upregulated in the more invasive or metastatic class in the published signatures should also be upregulated in the mesenchymal lines. The closer the percentage consistency is to 50%, the less likely the overlap is meaningful, since for every inconsistent overlap by chance, there is likely to be one consistent overlap by chance as well. A consistency of 50% indicates that the entire overlap is noise, and thus of no utility.

Of the three signatures with much higher than expected overlap with the EMT core signature, Schuetz et al. and Jaeger et al. in particular overlapped with a high consistency of 78 out of 88 (88.6%) and 73 out of 78 probes (93.6%) respectively, suggesting a biologically meaningful overlap (Table 2 and Figure 2). In order to build a mechanistic model based on a set of genes, one must be reasonably confident that a large proportion of the genes are informative, or contribute to the phenotype under study. As described above, for every inconsistent probe, there is likely to be one consistent probe that appears by chance. I subtracted the number of inconsistent probes from the consistent signal and calculated the proportion of the consistent probes that was likely not due to chance, deriving a percentage-signal measure (Table 2).

Using this method, I calculated a percentage-signal measure of 87.2% for Schuetz et al., 93.2% for Jaeger et al. and only 50% for Provenzani et al. The high proportion of the consistent overlap that was not due to chance (high percentage-signal) meant that the Schuetz et al. and Jaeger et al. overlapping sets would be more informative than the set of Provenzani et al., where half the consistent overlapped probes are expected to be random noise, with no way of knowing which half. The overlapping genes arising from Schuetz et al. and Jaeger et al. were thus used for further investigation.

These findings based on the signature overlaps are particularly significant because both published signatures of Schuetz et al. and Jaeger et al. derived from laser-capture microdissected samples. As mentioned above, bulk tumors contain not just the neoplastic cells, but also non-neoplastic stromal cells such as fibroblasts and myofibroblasts (Sappino et al., 1988; Orimo et al., 2001), which are mesenchymal in nature. Poor prognosis and late-stage samples may contain higher proportions of these cells, and naturally give a stronger mesenchymal gene signature even if the

actual carcinoma cells remain epithelial. The process by which the two signatures analyzed here were derived gave me confidence that the signature overlap represented a change to the cancer cells themselves, and not their surroundings. Thus, the cancer progressions characterized in these two studies may involve the cancer cells themselves undergoing an EMT or EMT-like change. The presence of a meaningful overlap not just with a breast cancer signature but a melanoma signature suggests that the processes involved in an EMT may be at work in a wider variety of cancer progression processes than previously thought.

It was not immediately obvious why a melanoma, which is not a carcinoma, should undergo a process reminiscent of the EMT associated with carcinomas. However, melanocytes are themselves derived from the neural crest (Dupin and Le Douarin, 2003), meaning they have undergone EMT in their developmental history. They may thus be epigenetically poised to re-engage part of the EMT mechanism during metastatic progression. Indeed, an independent study comparing melanoma primary tumors that did or did not go on to develop metastases generated a signature that identified EMT-related genes as a major determinant (Alonso et al., 2007).

2.2.3 Selecting candidates by transcription factor binding site enrichment

Hypothesizing that the genes common to the EMT core signature and the clinical signatures may be coordinately regulated by one or a few pleiotropically acting transcription factors, I further analyzed the overlapping gene sets from Schuetz et al. and Jaeger et al. for enrichment of predicted transcription factor binding sites (TFBS) within their promoter regions. Most transcription factors bind to sites that are defined by a short degenerate nucleotide sequence motif (6-12 bp in length) (Heinemeyer et al., 1998; Loots et al., 2002). These motifs are expected

to occur very often throughout the genome just by chance. Simply searching for matching sequence patterns would result in extraordinarily high hit rates, and many of the predicted TFBS would likely be functionally irrelevant (Fickett and Wasserman, 2000). By exploiting the fact that functional regions of the genome are more evolutionarily conserved, and the fact that transcription factor orthologs across species tend to retain the same DNA-binding specificity, it is possible to minimize the prediction of false positives in such an analysis. A web-based tool called rVISTA utilizes the alignment between the human and mouse genome and provides a ranked list of evolutionarily conserved TFBS for a user-provided set of genes (Loots et al., 2002).

Both the Schuetz et al. and Jaeger et al. overlap sets were submitted to the rVISTA tool. Microarray probe identities were converted to gene identities for this analysis, resulting in 64 unique genes being subjected to analysis for each of the overlap sets, out of the original 78 consistent probes from Schuetz et al. and 73 from Jaeger et al. The top hit for the breast ductal carcinoma overlap from Schuetz et al. was LEF1TCF1 (Figure 3A), which refers to binding site for lymphoid enhancer-binding factor 1 (LEF1). This site was found on 48 of the 64 overlapping genes. LEF1 is a transcription factor involved in the Wnt signaling pathway, associating with its co-activator β-catenin. Working together, the two factors have been reported capable of inducing EMT in certain cell lines (Kim et al., 2002). Wnt signaling is known to play a role in neural crest formation (Wu et al., 2005) and the maintenance of the mesenchymal phenotype (Scheel et al., 2011), while differential β-catenin staining was one of the earliest indications of EMT occurring in clinically observed colorectal cancers (Brabletz et al., 1998). Neither LEF1 nor β-catenin appeared to be differentially regulated on the microarray data, but this was unsurprising given the extensive

post-transcriptional regulation of these factors. The rVISTA result suggested that EMT could play a role in the conversion of DCIS to IDC, with LEF1 and β-catenin signaling mediating the changes.

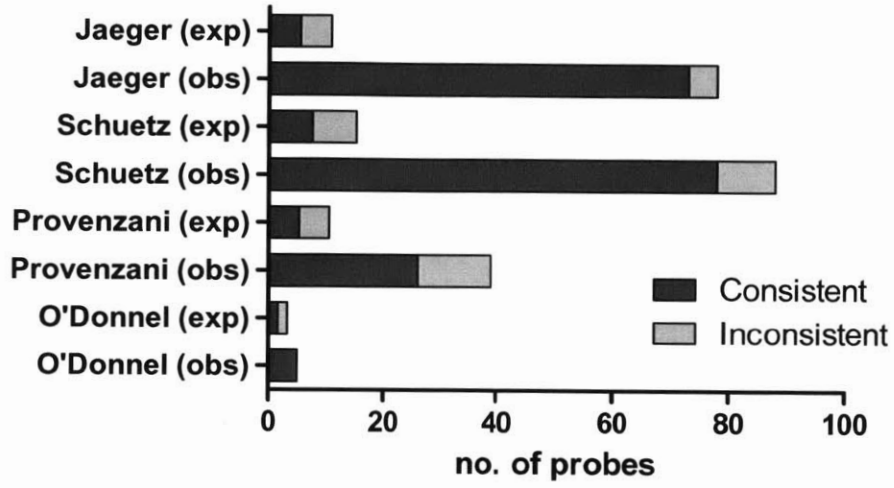


Figure 2. Graphic comparison of the expected (exp) and observed (obs) number of probes overlapping with the EMT core signature, from Table 2. The mathematically expected overlapping probes will always comprise an equal proportion of consistent and inconsistent probes, while a more biologically related overlap will be much larger than the expected overlap, and be made up mostly of consistent probes, as illustrated with Jaeger (obs) and Schuetz (obs).

The top hit for the rVISTA analysis of the melanoma overlap set of Jaeger et al. was a motif called AREB6 (Figure 3B), which appeared on 33 of the 64 overlapping genes. The AREB6 motif is defined by the transcription factor ZEB1. At the time, ZEB1 was not known to have a role in EMT of HMLE cells, and this was the first suggestion that it might. More compellingly, not only were many genes in this overlapping set found to contain ZEB1 binding sites on their promoters, but ZEB1 was itself an EMT core signature gene, upregulated in every mesenchymal HMLE line. These results suggested that not only could ZEB1 be a common mediator of the EMT in five different methods of triggering EMT in HMLE cells, but it could also play a hitherto unappreciated role in melanoma metastasis.

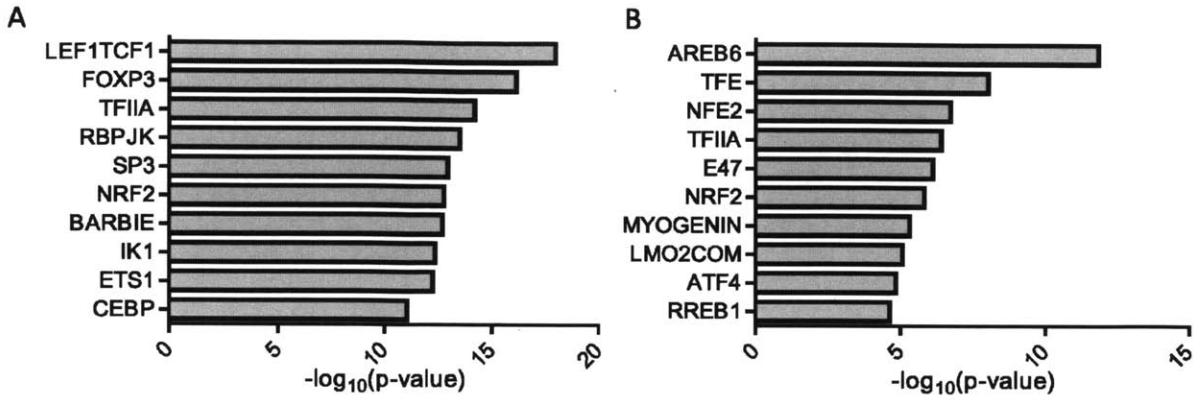


Figure 3. (A) rVISTA output of the evolutionarily conserved TFBS found in the set of overlapping genes between the EMT core signature and the breast invasive ductal carcinoma vs. DCIS signature (Schuetz et al., 2006). (B) rVISTA output of the evolutionarily conserved TFBS found in the set of overlapping genes between the EMT core signature and melanoma metastasis vs. primary signature (Jaeger et al., 2007).

The nature of enrichment analysis is such that results usually have extremely low p-values that are difficult to interpret in isolation. For instance, the LEF1TCF1 motif was overrepresented in the overlapping gene set with a p-value calculated to be less than $1x10^{-17}$, while the AREB6 motif had a p-value less than $1x10^{-11}$. This difference does not automatically mean that the LEF1TCF1 result was more meaningful, as the spectrum of genes and evolutionarily conserved regions being fed into the analysis was different in each case, even if the total number of genes was the same. Empirically, one is able to compare the results within each set of genes (Figure 3). Displayed graphically, the AREB6 result was quite far above the next best result, whereas the LEF1TCF1 result was less outstanding in comparison to its neighbors. The presence of ZEB1/AREB6 in the EMT core signature itself suggested that it would be extremely worthwhile to directly manipulate ZEB1 transcript levels to study the effects on EMT in HMLE cells, while the absence of LEF1 or β-catenin in the signature suggested the signaling pathway involved could be modulated by unknown factors at a post-transcriptional level. I therefore decided to concentrate on ZEB1 for further *in vitro* studies, detailed in Chapter 3.

The use of published gene signatures to derive meaningful subsets of the EMT core signature thus turned out to be a very productive approach. Had the entire EMT core signature been subjected to rVISTA analysis unfiltered, AREB6 would not have appeared among the top hits at all (Supplementary Figure 1). The confluence of signatures derived under entirely different circumstances (*in vitro* perturbation and spontaneous clinical progression) gave added reassurance that the rVISTA results obtained did not merely reflect idiosyncrasies of the particular system being studied (HMLE cells in this case), but represented instead more widely applicable biological processes.

2.2.4 Gene Set Enrichment Analysis (GSEA)

Besides looking only at differentially regulated genes that met statistical significance, there were other methods available to analyze microarray data. One method is Gene Set Enrichment Analysis (Subramanian et al., 2005), which works on the premise that individual gene changes may be insignificant, but if the genes are considered as sets, the collective behavior of the genes constituting the set may show enrichment in a ranked list, either at the top or bottom end, signifying some correlation between the data and the gene set. Since this method is based on detecting signals from otherwise statistically insignificant gene expression patterns, the combined microarray dataset, not filtered for the EMT core signature, was subjected to GSEA. The output of GSEA takes the form of the identities of gene sets, along with an associated Normalized Enrichment Score (NES) and a false-discovery rate (FDR) q-value. An NES further from zero indicates a higher enrichment of the gene set in question, while the FDR value controls for the testing of multiple hypotheses using multiple gene sets.

The GSEA is performed with pre-determined gene sets. One collection of gene sets is based on the genes that share particular Gene Ontology (GO) terms (Ashburner et al., 2000). GO terms are managed by the GO Consortium, and consists of a mixture of manual and automated text annotations that describe a gene and its function. Each gene can have many GO terms attached to it, and each GO term can encompass many genes. GSEA performed based on GO term gene sets is equivalent to testing for GO term enrichment. Performing this analysis using the combined microarray dataset, yielded only six gene sets with q-values of less than 0.05, the recommended cutoff for the analysis parameters used (Supplementary Table 1). The top ten enriched sets included GO terms for "ectoderm development" and "structural constituents of cytoskeleton" and "mesoderm development". While it was reassuring to have confirmation that the EMT did affect genes involved in these functions, these terms gave no clue as to the underlying regulatory mechanism.

The GSEA performed based on curated gene pathway sets (KEGG, Biocarta, Reactome) likewise did not yield any useful clues as to the genes or pathways that are mediating the EMT, as almost none of the sets met the q-value <0.05 cutoff (Supplementary Table 2). The three gene sets that did satisfy this criterion, related to smooth muscle contraction, RNA polymerase I promoter escape and tight junction interactions, did not contain any compelling candidate genes.

GSEA performed using sets of genes that contained defined transcription factor binding site motifs (as defined in the TRANSFAC database) around their transcription start sites resulted in 29 gene sets that met the cutoff (Supplementary Table 3). Since each set is defined by a transcription factor, the expression of that transcription factor could be checked within the microarray data.

None of the defining transcription factors were consistently up- or downregulated in the mesenchymal cell lines.

GSEA using sets created from published experimentally or clinically generated gene signatures led to over 200 sets appearing to meet the q-value cutoff, and 49 sets alone having a zero q-value, making it difficult to pick one set over another for further study. However, among the list of enriched signatures that satisfied the q-value cutoff criterion were the breast carcinoma signature of Schuetz et al. and the melanoma signature of Jaeger et al., while the colorectal cell line signature of Provenzani et al. did not make the cutoff (Supplementary Table 4). This provided an independent confirmation in retrospect of the validity of the method of assessing gene signature overlaps described in Section 2.2.2. Had this GSEA been done in isolation, the Schuetz et al. dataset would never have been noticed, being ranked 48th among the 200 sets meeting the q-value cutoff. On the other hand, Jaeger et al. would have been noticed, being ranked 2nd, just below a set derived from Onder et al., which was used to create the EMT core signature in the first place.

Overall, GSEA analysis did not yield great dividends. While the principle behind the approach is sound, part of the reason it was not useful may be that the analysis is only as good as the gene sets being used. Many of the gene sets were computationally generated and may have only weak supporting evidence (du Plessis et al., 2011). The large number of such gene sets represents additional noise in the analysis, increasing the number of hypotheses tested without increasing the likelihood of finding a meaningful enrichment. The relatively small number of samples being tested (21) is also ill suited to GSEA, which was optimally designed to handle large numbers (>30) of samples of each phenotype.

2.2.5 Comparing EMT core signature to Drosophila ChIP-chip for Twist and Snail

Prior to my experiments, Zeitlinger et al. had previously performed whole genome chromatin immunoprecipitation coupled with microarray analysis (ChIP-chip) on *Drosophila* embryos using antibodies for Dorsal (fly ortholog of NF-κB), Twist and Snail (Zeitlinger et al., 2007). In the process, they produced a list of high-confidence target genes that are bound by both Twist and Snail. Since both Twist and Snail can induce EMT in HMLE cells, it was possible that they were acting directly on the same genes. Working on the possibility that transcription factor-target gene relationships may be conserved between flies and humans, I compared the fly gene list to the EMT core signature. From the list, 373 fly genes with human orthologs were found, and these were matched to genes in the EMT core signature. A list of 20 matching human genes was generated (

D. melanogaster gene bound by Twist, Snail	*H. sapiens* ortholog in EMT core signature
CREG	CREG1
shot	DST
Mkp3	DUSP6
Egfr	EGFR
dp	FBN1
dp	FBN2
Fs	FST
fz	FZD7
homer	HOMER1
Antp	HOXA5
Awh	LHX6
shn	MBP
CG2022	MLPH
Pgk	PGK1
Prat2	PPAT
Pkc98E	PRKCH
how	QKI
Rgl	RGL1
gem	TFCP2L1
zfh1	ZEB1

Table 3). Among this list was ZEB1, whose *Drosophila* ortholog, zfh1, appeared to have its promoter bound by *Drosophila* Twist and Snail. Thus, not only did ZEB1 appear to be able to mediate downstream targets relevant to the EMT and melanoma, there was also a suggestion of an evolutionarily conserved relationship between ZEB1, Twist and Snail.

D. melanogaster gene bound by Twist, Snail	*H. sapiens* ortholog in EMT core signature
CREG	CREG1
shot	DST
Mkp3	DUSP6
Egfr	EGFR
dp	FBN1
dp	FBN2
Fs	FST
fz	FZD7
homer	HOMER1
Antp	HOXA5
Awh	LHX6
shn	MBP
CG2022	MLPH
Pgk	PGK1
Prat2	PPAT
Pkc98E	PRKCH
how	QKI
Rgl	RGL1
gem	TFCP2L1
zfh1	ZEB1

Table 3. Genes that are found to be bound by *Drosophila melanogaster* Twist and Snail using chromatin-immunoprecipitation with microarray analysis (ChIP-chip) performed by Zeitlinger et al., and their corresponding human orthologs in the EMT core signature as predicted by Ensembl (Vilella et al., 2009).

2.3 Conclusions

Prior to this study, EMT-inducing signals tended to be studied in isolation. Working on the assumption that the EMT-inducing signals could converge on a common mechanism, I generated an EMT gene signature that led to novel connections to clinical data. I showed that EMT might play a role in the conversion of ductal carcinoma *in situ* into invasive ductal carcinoma, and more surprisingly, that it might be involved in the conversion of melanoma primary tumors to metastases. I found ZEB1 to be a compelling candidate mediator of the EMT using several different lines of evidence, and combining a wide variety of data sources. The rVISTA analysis produced ZEB1 as a candidate through an unbiased approach. The *Drosophila* ChIP-chip data

suggested that the connection between Snail, Twist and ZEB1 could be highly evolutionarily conserved. All the results suggest that ZEB1 is a highly compelling candidate as a central mediator of the EMT.

2.4 Materials and Methods

2.4.1 Microarray Data Analysis

Microarray data for HMLE expressing shCDH1 and the vector control were extracted from the Gene Expression Omnibus (GEO) database under the accession GSE9691. Data for HMLE-Gsc, Snail, Twist, TGF-β1 and the HMLE vector control were extracted from GEO under accession GSE24202. Raw array output CEL files were downloaded and processed using GenePattern (Reich et al., 2006). Each sample class contained triplicates. Raw data were normalized using RMA method with quantile normalization and background correction.

To select for differentially expressed genes, each mesenchymal line was paired with its epithelial control equivalent, and a 2-sided T-test was performed on their microarray data. To correct for multiple hypotheses, a false discovery rate (FDR) metric, the *q* value, for each probeset was calculated using the method of Storey and Tibshirani (Storey and Tibshirani, 2003), and a cutoff value of <0.05 for the *q* value was imposed. Five sets each of up- and down-regulated genes were thus generated. The EMT core signature was composed of the probes that met the *q* value cutoff and moved in a consistent direction in all five mesenchymal lines compared to their epithelial controls.

The heatmap and dendrogram for the EMT core signature was generated by filtering the data for all 21 arrays for only the 624 probes in the signature, collapsing the dataset for unique genes, and performing hierarchical clustering using Pearson correlation for both sample distance and row distance.

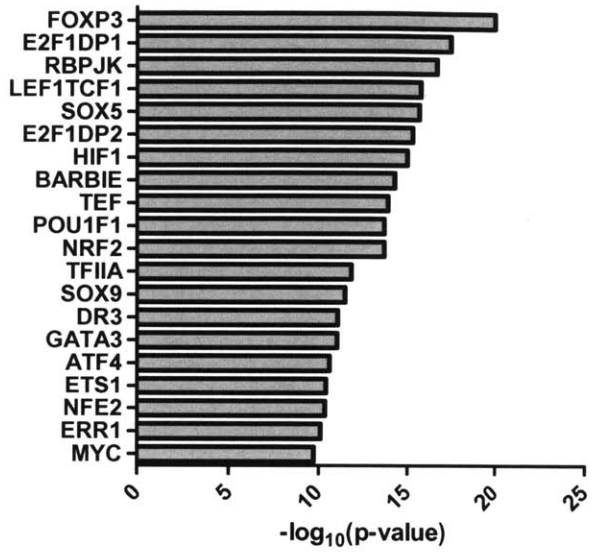
2.4.2 Transcription factor binding site analysis

The VISTA suite of online tools (http://genome.lbl.gov/vista/index.shtml) was used for TFBS analysis. Whole genome rVISTA with the precomputed alignments between the Human March 2006 (hg18) assembly and the Mouse February 2006 (mm8) assembly was selected. The promoter region of genes was designated as the 3000 bp upstream of the coding start site. Consistent overlapping probes were converted into unique gene locus link ids before being submitted to analysis.

2.4.3 Gene Set Enrichment Analysis (GSEA)

GSEA (Subramanian et al., 2005) was performed using the combined, normalized but unfiltered microarray data of the HMLE controls and derivative mesenchymal lines. The software implementation is found in Genepattern (Reich et al., 2006). Due to the low number of samples (6 epithelial and 15 mesenchymal), gene set permutation was used instead of the default phenotype permutation, necessitating a lower recommended q-value cutoff of 0.05. Other parameters were left at their default settings.

2.5 Supplementary Data



Supplementary Figure 1. rVISTA output of the evolutionarily conserved TFBS found in the entire EMT core signature.

GO term gene sets	Normalized enrichment score (NES)	FDR q-val
ECTODERM DEVELOPMENT	2.09	0
EXTRACELLULAR MATRIX	-2.11	0
PROTEINACEOUS EXTRACELLULAR MATRIX	-2.00	0.00941
COLLAGEN	-1.92	0.0221
EPIDERMIS DEVELOPMENT	1.97	0.0311
GUANYL NUCLEOTIDE BINDING	-1.85	0.0447
STRUCTURAL CONSTITUENT OF CYTOSKELETON	1.88	0.0700
EXTRACELLULAR MATRIX PART	-1.80	0.0721
MESODERM DEVELOPMENT	-1.77	0.111
SKELETAL DEVELOPMENT	-1.76	0.113

Supplementary Table 1. Gene Set Enrichment Analysis (GSEA) analysis of combined HMLE microarray data, showing the top ten Gene Ontology (GO) term gene sets enriched during EMT. A negative NES indicates a positive correlation with the mesenchymal phenotype, while a positive NES indicates a positive correlation with the epithelial phenotype.

Curated pathway gene sets	Normalized enrichment score (NES)	FDR q-val
REACTOME SMOOTH MUSCLE CONTRACTION	-2.21	0
REACTOME RNA POLYMERASE I PROMOTER ESCAPE	1.92	0.0307
REACTOME TIGHT JUNCTION INTERACTIONS	1.96	0.0319
BIOCARTA CALCINEURIN PATHWAY	-1.77	0.0853
REACTOME REGULATION OF INSULIN LIKE GROWTH FACTOR ACTIVITY BY INSULIN LIKE GROWTH FACTOR BINDING PROTEINS	-1.80	0.0890
REACTOME NCAM1 INTERACTIONS	-1.72	0.0967
BIOCARTA P53HYPOXIA PATHWAY	-1.84	0.0988
REACTOME SIGNALING BY PDGF	-1.78	0.0991
KEGG NICOTINATE AND NICOTINAMIDE METABOLISM	-1.70	0.107
KEGG PENTOSE AND GLUCURONATE INTERCONVERSIONS	-1.69	0.109

Supplementary Table 2. GSEA analysis of combined HMLE microarray data, showing top ten curated pathway gene sets based on the KEGG, BioCarta and Reactome database of curated biological pathways.

TRANSFAC motif-containing gene sets	Normalized enrichment score (NES)	FDR q-val
V$SRF_Q5_01	-1.83	0.0140
V$HTF_01	-1.94	0.0141
V$CART1_01	-1.75	0.0144
V$ATF6_01	-1.73	0.0182
V$POU6F1_01	-1.68	0.0186
V$RFX1_02	-1.66	0.0215
V$CDX2_Q5	-1.68	0.0217
V$SRF_01	-1.68	0.0233
TGTYNNNNNRGCARM_UNKNOWN	-1.63	0.0283
V$SP1_Q2_01	-1.59	0.0283
ACAWNRNSRCGG_UNKNOWN	-1.60	0.0290
V$SRF_Q4	-1.63	0.0297
V$SREBP1_01	-1.59	0.0297
V$OCT1_03	-1.64	0.0299
AAGWWRNYGGC_UNKNOWN	-1.60	0.0306
V$PITX2_Q2	-1.57	0.0306
V$EGR1_01	-1.63	0.0307
GGGNRMNNYCAT_UNKNOWN	-1.60	0.0307
V$SRF_C	-1.61	0.0310
V$FREAC4_01	-1.61	0.0311
GCCATNTTG_V$YY1_Q6	-1.62	0.0313
V$FOXO3_01	-1.65	0.0316
CCAWWNAAGG_V$SRF_Q4	-1.57	0.0318
V$YY1_Q6	-1.58	0.0325
WYAAANNRNNNGCG_UNKNOWN	-1.61	0.0325
CCAWNWWNNNGGC_UNKNOWN	-1.58	0.0333
V$EVI1_06	-1.58	0.0347
V$EVI1_04	-1.55	0.0386
AGCYRWTTC_UNKNOWN	-1.53	0.0496

Supplementary Table 3. GSEA analysis of combined HMLE microarray data, showing gene sets containing transcription factor binding site (TFBS) motifs based on the TRANSFAC database, with a q-value <0.05.

Published signature gene sets	Normalized enrichment score (NES)	FDR q-val
JAEGER METASTASIS DN	3.15	0
JAEGER METASTASIS UP	-1.68	0.0414
SCHUETZ BREAST CANCER DUCTAL INVASIVE UP	-2.39	0
SCHUETZ BREAST CANCER DUCTAL INVASIVE DN	2.04	0.000596
PROVENZANI METASTASIS UP	-1.44	0.154
PROVENZANI METASTASIS DN	1.26	0.277
ODONNELL METASTASIS DN	1.81	0.0137
ODONNELL METASTASIS UP	-0.766	0.933

Supplementary Table 4. GSEA analysis of combined HMLE microarray data, showing selected gene sets based on published microarray data. The gene signatures upon which these sets were based were earlier compared with the EMT core signature (Table 2) and show agreement with the GSEA results.

2.6 References

Alonso, S. R., Tracey, L., Ortiz, P., Pérez-Gómez, B., Palacios, J., Pollán, M., Linares, J., Serrano, S., Sáez-Castillo, A. I., Sánchez, L., Pajares, R., Sánchez-Aguilera, A., Artiga, M. J., Piris, M. A., and Rodríguez-Peralto, J. L. (2007). A high-throughput study in melanoma identifies epithelial-mesenchymal transition as a major determinant of metastasis. Cancer research *67*, 3450–3460.

Ashburner, M., Ball, C. A., Blake, J. A., Botstein, D., Butler, H., Cherry, J. M., Davis, A. P., Dolinski, K., Dwight, S. S., Eppig, J. T., Harris, M. A., Hill, D. P., Issel-Tarver, L., Kasarskis, A., Lewis, S., Matese, J. C., Richardson, J. E., Ringwald, M., Rubin, G. M., and Sherlock, G. (2000). Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. Nature genetics *25*, 25–29.

Brabletz, T., Jung, A., Hermann, K., Günther, K., Hohenberger, W., and Kirchner, T. (1998). Nuclear overexpression of the oncoprotein beta-catenin in colorectal cancer is localized predominantly at the invasion front. Pathology, research and practice *194*, 701–704.

Dupin, E., and Le Douarin, N. M. (2003). Development of melanocyte precursors from the vertebrate neural crest. Oncogene *22*, 3016–3023.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M. D., Zimonjic, D. B., Donaher, J. L., Popescu, N. C., Hahn, W. C., and Weinberg, R. A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes & development *15*, 50–65.

Fickett, J. W., and Wasserman, W. W. (2000). Discovery and modeling of transcriptional regulatory regions. Current opinion in biotechnology *11*, 19–24.

Hartwell, K. A., Muir, B., Reinhardt, F., Carpenter, A. E., Sgroi, D. C., and Weinberg, R. A. (2006). The Spemann organizer gene, Goosecoid, promotes tumor metastasis. Proc Natl Acad Sci U S A *103*, 18969–18974.

Heinemeyer, T., Wingender, E., Reuter, I., Hermjakob, H., Kel, A. E., Kel, O. V., Ignatieva, E. V., Ananko, E. A., Podkolodnaya, O. A., Kolpakov, F. A., Podkolodny, N. L., and Kolchanov, N. A. (1998). Databases on transcriptional regulation: TRANSFAC, TRRD and COMPEL. Nucleic acids research *26*, 362–367.

Irizarry, R. A., Warren, D., Spencer, F., Kim, I. F., Biswal, S., Frank, B. C., Gabrielson, E., Garcia, J. G. N., Geoghegan, J., Germino, G., Griffin, C., Hilmer, S. C., Hoffman, E., Jedlicka, A. E., Kawasaki, E., Martínez-Murillo, F., Morsberger, L., Lee, H., Petersen, D., Quackenbush, J., Scott, A., Wilson, M., Yang, Y., Ye, S. Q., and Yu, W. (2005). Multiple-laboratory comparison of microarray platforms. Nature methods *2*, 345–350.

Jaeger, J., Koczan, D., Thiesen, H. J., Ibrahim, S. M., Gross, G., Spang, R., and Kunz, M. (2007). Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. Clin Cancer Res *13*, 806–815.

Kim, K., Lu, Z., and Hay, E. D. (2002). Direct evidence for a role of beta-catenin/LEF-1 signaling pathway in induction of EMT. Cell biology international *26*, 463–476.

Loots, G. G., Ovcharenko, I., Pachter, L., Dubchak, I., and Rubin, E. M. (2002). rVista for comparative sequence-based discovery of functional transcription factor binding sites. Genome research *12*, 832–839.

Onder, T. T., Gupta, P. B., Mani, S. A., Yang, J., Lander, E. S., and Weinberg, R. A. (2008). Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. Cancer Res *68*, 3645–3654.

Orimo, A., Tomioka, Y., Shimizu, Y., Sato, M., Oigawa, S., Kamata, K., Nogi, Y., Inoue, S., Takahashi, M., Hata, T., and Muramatsu, M. (2001). Cancer-associated myofibroblasts possess various factors to promote endometrial tumor progression. Clinical cancer research : an official journal of the American Association for Cancer Research *7*, 3097–3105.

O'Donnell, R. K., Kupferman, M., Wei, S. J., Singhal, S., Weber, R., O'Malley, B., Cheng, Y., Putt, M., Feldman, M., Ziober, B., Muschel, R. J., O'Donnell, R. K., and O'Malley, B. (2005). Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. Oncogene *24*, 1244–1251.

du Plessis, L., Skunca, N., and Dessimoz, C. (2011). The what, where, how and why of gene ontology--a primer for bioinformaticians. Briefings in bioinformatics *12*, 723–735.

Provenzani, A., Fronza, R., Loreni, F., Pascale, A., Amadio, M., and Quattrone, A. (2006). Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. Carcinogenesis *27*, 1323–1333.

Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003). A molecular signature of metastasis in primary solid tumors. Nat Genet *33*, 49–54.

Reich, M., Liefeld, T., Gould, J., Lerner, J., Tamayo, P., and Mesirov, J. P. (2006). GenePattern 2.0. Nat Genet *38*, 500–501.

Sappino, A. P., Skalli, O., Jackson, B., Schürch, W., and Gabbiani, G. (1988). Smooth-muscle differentiation in stromal cells of malignant and non-malignant breast tissues. International journal of cancer. Journal international du cancer *41*, 707–712.

Scheel, C., Eaton, E. N., Li, S. H.-J., Chaffer, C. L., Reinhardt, F., Kah, K.-J., Bell, G., Guo, W., Rubin, J., Richardson, A. L., and Weinberg, R. A. (2011). Paracrine and autocrine signals induce and maintain mesenchymal and stem cell States in the breast. Cell *145*, 926–940.

Schuetz, C. S., Bonin, M., Clare, S. E., Nieselt, K., Sotlar, K., Walter, M., Fehm, T., Solomayer, E., Riess, O., Wallwiener, D., Kurek, R., and Neubauer, H. J. (2006). Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer Res *66*, 5278–5286.

Storey, J. D., and Tibshirani, R. (2003). Statistical significance for genomewide studies. Proceedings of the National Academy of Sciences of the United States of America *100*, 9440–9445.

Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., Paulovich, A., Pomeroy, S. L., Golub, T. R., Lander, E. S., and Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proc Natl Acad Sci U S A *102*, 15545–15550.

Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., Hartwell, K., Onder, T. T., Gupta, P. B., Evans, K. W., Hollier, B. G., Ram, P. T., Lander, E. S., Rosen, J. M., Weinberg, R. A., and Mani, S. A. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci U S A *107*, 15449–15454.

van t Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Vilella, A. J., Severin, J., Ureta-Vidal, A., Heng, L., Durbin, R., and Birney, E. (2009). EnsemblCompara GeneTrees: Complete, duplication-aware phylogenetic trees in vertebrates. Genome research *19*, 327–335.

Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., Jatkoe, T., Berns, E. M., Atkins, D., and Foekens, J. A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671–679.

Wu, J., Yang, J., and Klein, P. S. (2005). Neural crest induction by the canonical Wnt pathway can be dissociated from anterior-posterior neural patterning in Xenopus. Developmental biology *279*, 220–232.

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927–939.

Zeitlinger, J., Zinzen, R. P., Stark, A., Kellis, M., Zhang, H., Young, R. A., and Levine, M. (2007). Whole-genome ChIP-chip analysis of Dorsal, Twist, and Snail suggests integration of diverse patterning processes in the Drosophila embryo. Genes Dev *21*, 385–390.

Chapter 3:

ZEB1 in EMT and its relation to other EMT mediators

3.1 Introduction

Several different lines of analysis of the HMLE microarray data detailed in Chapter 2 indicated that ZEB1 may have a role in mediating the EMT program induced by several different signals. ZEB1 was upregulated in the EMT induced by Twist, Snail, Goosecoid, TGF-β1 and shRNA against E-cadherin. Overlapping genes between the EMT core signature and a melanoma metastasis signature were enriched for ZEB1 binding sites at their promoter regions. The promoter region of the fly ortholog of ZEB1, zfh1, was found to be bound by the fly orthologs of Twist and Snail. These combined observations led me to hypothesize that ZEB1 lies downstream of the other known EMT-inducing signals, funneling their inputs through ZEB1 into determining whether or not to trigger EMT. The gene relationships in a distantly related species (*Drosophila*), as well as the appearance of possible ZEB1 target genes in the metastasis signature of a different cell type (melanocytes), suggested that the role of ZEB1 may be far-reaching.

3.2 Results and Discussion

3.2.1 Generating a lentiviral vector for rapid confirmation of protein expression

Previously, constitutive overexpression constructs in the Weinberg lab were made using Moloney Murine Leukemia Virus (MMLV)-derived retroviruses driven by an SV40 immediate early promoter, with infection efficiencies of less than 50%, necessitating the use of antibiotic selection to obtain the infected cell population. I decided to use instead an HIV-based lentiviral vector, which

promised higher infection efficiencies. Starting with a modified FUW lentiviral construct (Lois et al., 2002), which contained a human ubiquitin promoter, I added a puromycin selection marker driven by a separate SV40 immediate early promoter. At the end of the multiple cloning site, I added the gene for the modified red fluorescent protein mStrawberry (Shu et al., 2006) without a starting ATG, preceded by a foot-and-mouth disease virus 2A sequence (Ryan and Drew, 1994). The resulting FUW-2A-mStrawberry plasmid allowed me to clone a gene of interest in-frame to the 2A sequence (Figure 1). The 2A oligopeptide allows for ribosomal skipping (Donnelly et al., 2001), physically separating the first translated polypeptide from the one encoded downstream of the 2A peptide by skipping the synthesis of the glycyl-prolyl peptide bond at the C-terminus of 2A. The first polypeptide is released from the ribosome, which resumes translating the downstream sequence. Crucially, the skipping does not affect the frame of translation, so a successful translation of the downstream polypeptide means that the upstream sequence was successfully translated as well. The most useful feature of virus generated using this plasmid is rapid confirmation of successful protein expression by checking for fluorescence. Cells that exhibit red fluorescence due to the presence of mStrawberry are assumed to have successfully translated the upstream product.

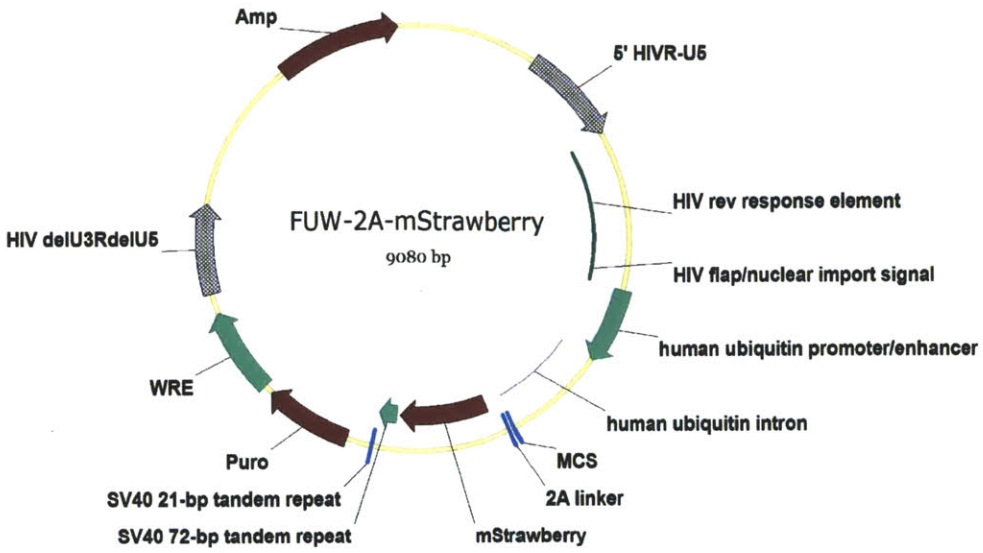


Figure 1. Vector map of modified FUW lentivirus for constitutive expression of gene of interest along with co-transcribed and co-translated mStrawberry fluorescent protein.

This property was tested by cloning eGFP upstream of the 2A sequence. Every cell that is red from mStrawberry should theoretically be green from eGFP as well. The upstream and downstream polypeptides need not be produced in equimolar amounts, and in fact the upstream polypeptide is normally expressed in relative excess (Donnelly et al., 2001). However, the proportion between the two is relatively constant for a given construct. HMLE cells infected with this test virus, henceforth termed HMLE-GFP, were subjected to flow cytometry to measure the relative levels of eGFP and mStrawberry (RFP) fluorescence (Figure 2). The fluorescence-activated cell sorting (FACS) profile of these cells appeared as a sloped line, indicating that in every individual cell, the ratio of green-to-red fluorescence was similar. This contrasted with the profile of cells that were infected with two independent fluorescent marker viruses, where the brightness in one channel was unrelated to the brightness in the second channel. The HMLE-GFP cells displayed a minor

population of eGFP-positive, mStrawberry-negative cells, indicating that there were cells that produced a functional pre-2A polypeptide without the post-2A polypeptide. An even smaller proportion displayed the reverse phenotype, being mStrawberry-positive and eGFP-negative.

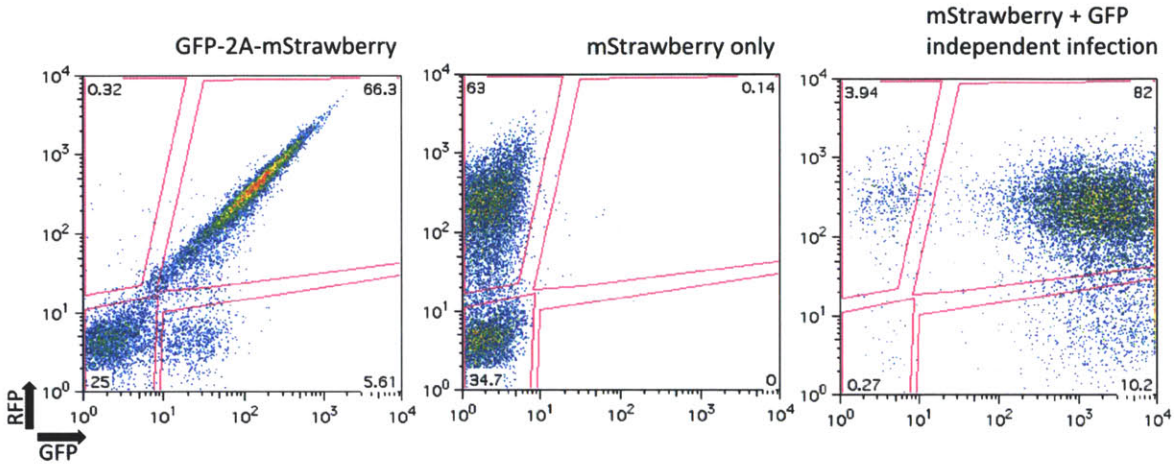


Figure 2. Flow cytometry analysis of HMLE cells infected with the FUW-GFP-2A-mStrawberry virus, compared with cells infected with virus encoding only one fluorophore and cells infected with two viruses encoding two fluorophores. The distribution of the GFP-2A-mStrawberry infected cells showed high correlation between GFP and RFP signal on an individual cell basis, unlike the doubly infected cells.

3.2.2 ZEB1 is sufficient to induce EMT in HMLE cells

In order to test whether ZEB1 was sufficient to induce an EMT in HMLE cells, the coding region of ZEB1 was cloned into the FUW-2A-mStrawberry construct. HMLE-ZEB1 cells took on an elongated appearance and failed to form tightly packed epithelial islands, compared with control HMLE-GFP cells (Figure 3A). Realtime PCR indicated a downregulation of the epithelial markers E-cadherin and desmoplakin, as well as upregulation of mesenchymal markers fibronectin and vimentin (Figure 3B). These changes are characteristic of an EMT in HMLE cells as previously characterized

(Hartwell et al., 2006; Yang et al., 2004; Onder et al., 2008; Whipple et al., 2010; Casas et al., 2011; Shapiro et al., 2011).

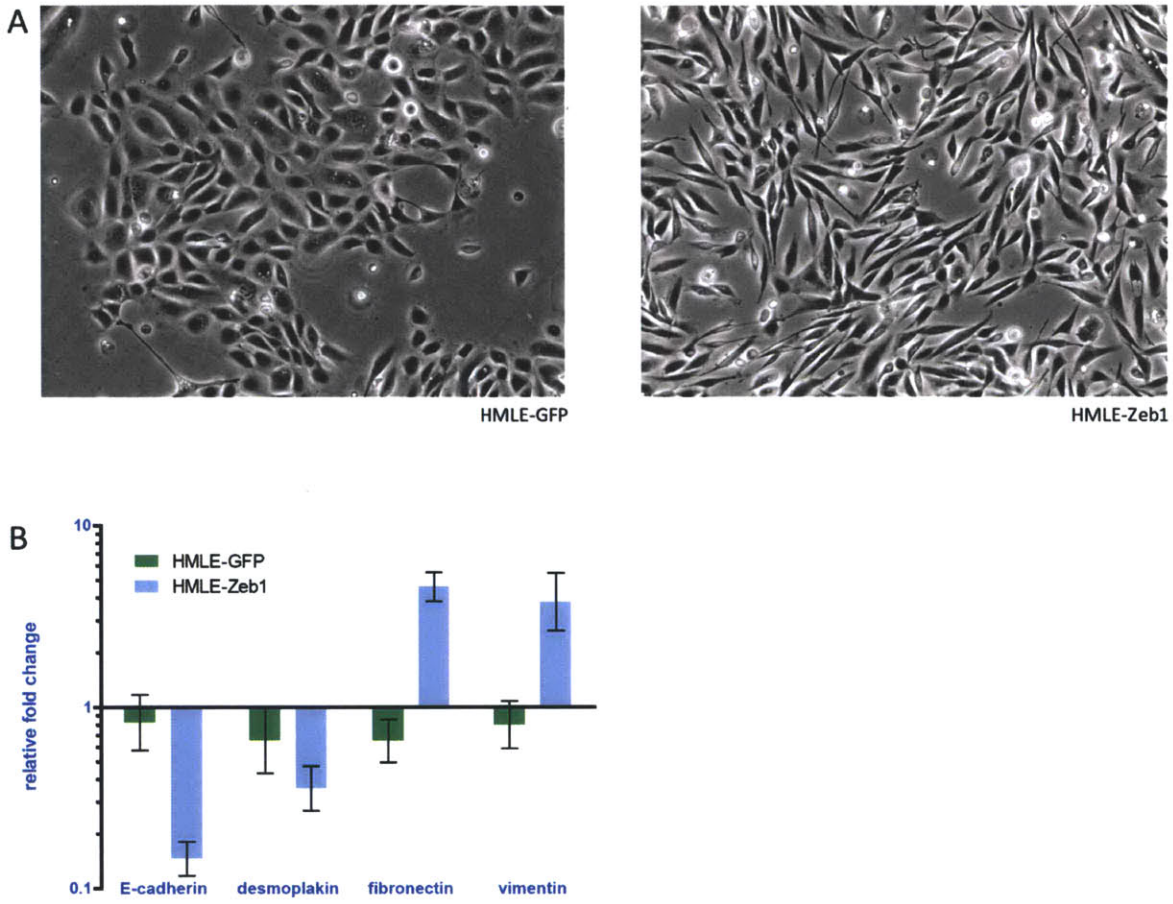


Figure 3. (A) Morphology of HMLE cells after infection with control virus or ZEB1-expressing virus. (B) Realtime PCR of EMT markers, relative to uninfected controls, showing agreement with morphological changes.

3.2.3 ZEB1-induced EMT is rapid

Upon repeated independent infections, I observed that the ZEB1-expressing HMLE cells acquired their mesenchymal characteristics in a relatively short timeframe, usually within two cell passages,

or three days. This observation was at odds with previous experience at inducing EMT in HMLE cells. While it was not reported in published form, prior attempts to induce EMT in HMLE cells took at least a week, sometimes more. This raised the question of whether the observed phenotype change resulted from the induction of a new phenotype or selection for the outgrowth of a pre-existing mesenchymal subpopulation, an issue that I explored in greater detail later.

The ZEB1 expression construct was very different from the previously used MMLV-based retroviruses, raising the possibility that the difference in speed of EMT induction was due to the different vector constructions. In order to address this possibility, Snail and Twist were subcloned into the new lentiviral construct and used to infect HMLE cells. HMLE-ZEB1 cells continued to acquire a mesenchymal morphology faster than HMLE-Snail and HMLE-Twist cells, even when the genes were subcloned into the same lentiviral context. The HMLE-ZEB1 cells appeared as individual, elongated cells on the culture plate by 6 days post-infection, while HMLE-Twist and HMLE-Snail cells at the same time point still grew in epithelial islands and looked almost indistinguishable from HMLE-GFP control cells (Figure 4A). At 5 days post-infection, HMLE-ZEB1 cells had downregulated E-cadherin and upregulated fibronectin and vimentin to a greater extent than HMLE-Snail cells, while HMLE-Twist cells had barely registered any change in these markers at all (Figure 4B). By 22 days post-infection, HMLE-Snail and HMLE-Twist cells did undergo EMT, downregulating E-cadherin and upregulating fibronectin and vimentin to extents similar to or greater than those observed in HMLE-ZEB1 cells (Figure 4C), showing that the constructs were functional.

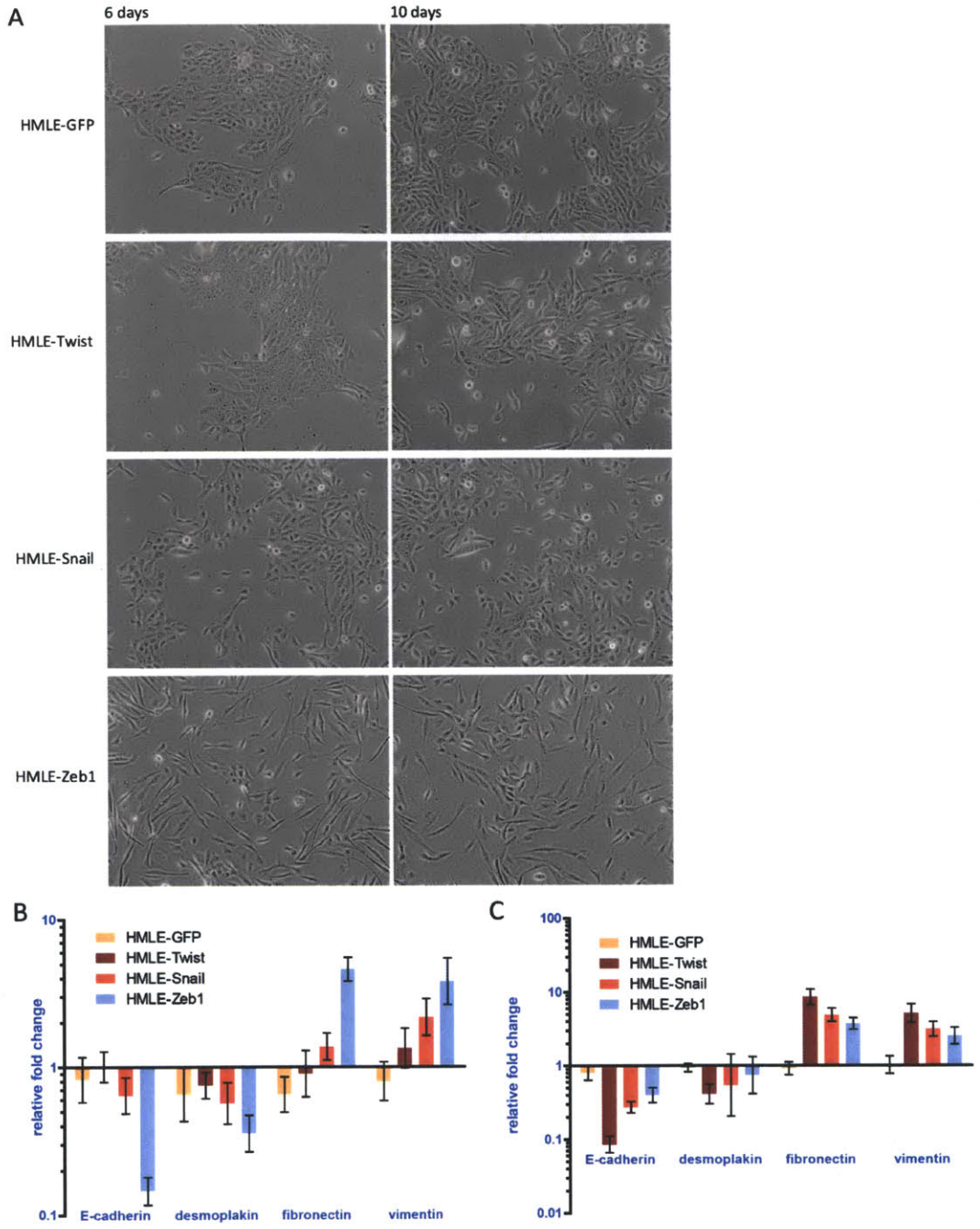


Figure 4. (A) Morphology of HMLE cells after infection with constitutive-expressing lentivirus, 6 and 10 days after infection. (B and C) Realtime PCR of EMT markers at 5 and 22 days, respectively, post-infection showing HMLE-Twist and HMLE-Snail eventually do achieve EMT.

It was possible that despite the identical vector context, the HMLE-ZEB1 cells underwent EMT more quickly because there was a higher multiple of infection, or more efficient transcription of the transduced gene, leading to more abundant vector-encoded mRNA. However, the lentivirus used in these experiments also encoded the mStrawberry fluorescent protein; its mRNA was co-transcribed and then co-translated together with the gene of interest (Ryan and Drew, 1994). Accordingly, I measured the levels of mStrawberry transcript in all of the vector-infected cells (Figure 5). The same target sequence was amplified in all cases, so the relative abundance of the mStrawberry signal was directly related to the number of transcripts expressed, even if the transcripts were of different sizes in the different cells. HMLE-Twist cells in fact exhibited the highest level of exogenous transcript, while HMLE-ZEB1 cells had the lowest, contrasting with the fact that the ZEB1 vector induced an EMT far more rapidly than did the Twist vector. Since these levels did not correlate with the speed of EMT induction in these cells, I could exclude the simple explanation of gene dosage effects to rationalize these differences in response. This strengthened the case that HMLE-ZEB1 cells underwent a more rapid EMT because of the intrinsic properties of ZEB1.

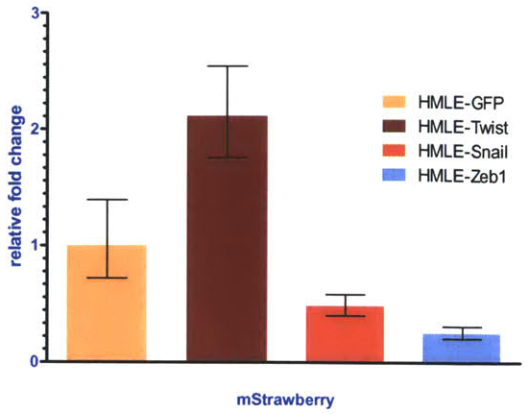


Figure 5. Levels of mStrawberry fluorescent protein mRNA, relative to HMLE-GFP. Levels do not correlate with speed of EMT in HMLE cells.

3.2.4 Twist-induced EMT displays characteristics of selective outgrowth

I observed that infection of the Twist-encoding virus led to more cell death compared with the GFP, Snail and ZEB1-encoding viruses. This occurred during the process of puromycin selection, even though there was no discernible difference in mStrawberry positivity. HMLE cells are known to be heterogeneous and contain a small basal subpopulation of mesenchymal cells (Mani et al., 2008). Carcinoma cells with mesenchymal characteristics are more resistant to chemo- and radio-toxicity (Creighton et al., 2009; Gupta et al., 2009). Twist overexpression combined with puromycin exposure might lead to a state of stress that mesenchymal cells could better survive than epithelial cells. It is not unusual for cells to become stressed by exposure to an EMT-inducing factor. The best example is NMuMG cells being exposed to TGF-β (Gal et al., 2008), resulting in anoikis and the emergence of mesenchymal cells. These observations, taken together, suggested the possibility that the emergence of mesenchymal cells from HMLE-Twist was due in part to the

selective outgrowth of existing mesenchymal cells rather than the conversion of epithelial cells to mesenchymal cells.

To further characterize this phenomenon, HMLE cells were infected with Snail and Twist virus separately, then FACS-sorted for only the mStrawberry-positive cells. In this way, all the cells under subsequent study were guaranteed to be infected and should have contained the puromycin resistance marker. Each sorted population was then observed with and without puromycin selection over time. Over the course of 25 days, the HMLE-Snail cells underwent EMT, progressively downregulating E-cadherin while upregulating fibronectin and ZEB1 (Figure 6). There was no difference in the behavior of E-cadherin, fibronectin and ZEB1 between cells that were exposed to puromycin and cells that were not. In the HMLE-Twist cells, however, the absence of puromycin selection (and the attendant cell death) resulted in cells that displayed marked differences in behavior. With puromycin selection, the downregulation of E-cadherin as well as the upregulation of fibronectin and ZEB1 occurred at a rate on par with or more quickly than the HMLE-Snail cells. Without puromycin selection, the expression changes of all three genes occurred more slowly than in HMLE-Snail cells (Figure 6). This correlation between speed of EMT and presence of cell death suggests that Twist-induced EMT in HMLE cells may be due in part to selective outgrowth of existing mesenchymal cells, although a role for Twist in actively inducing EMT could not be ruled out. This phenomenon was not pursued further, as Twist was not deemed central to the present research, but did suggest that Twist is not as capable of inducing an EMT in HMLE cells as Snail.

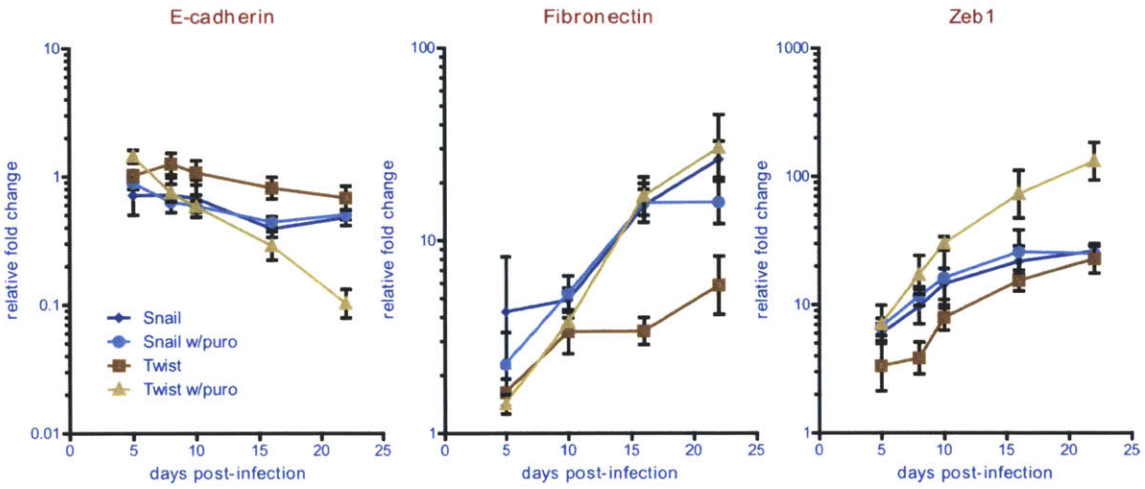


Figure 6. Realtime PCR of selected EMT markers over time for HMLE cells infected and treated with or without puromycin selection as indicated. All cells were RFP-sorted prior to day 5. HMLE-Snail cells behave no differently in the presence or absence of puromycin, but HMLE-Twist cells acquired EMT characteristics more quickly with puromycin selection than without, despite all the cells already possessing the selectable marker.

3.2.5 ZEB1-induced EMT is apparently not stable

During long term culture of HMLE-ZEB1 cells, epithelial cells invariably appeared in the culture dish and would eventually dominate the population. This appearance of epithelial cells could represent a Mesenchymal-Epithelial Transition (MET) triggered by some feedback mechanism, or an outgrowth of a minor population of epithelial cells that never underwent EMT in the first place. Indeed, other work with similar mammary cell lines has shown that epithelial cells can proliferate more rapidly than their mesenchymal counterparts, explaining their ability to overgrow mesenchymal populations (Chaffer et al., 2011). A measurement of the mStrawberry transcript levels over time showed that while the levels were relatively stable in HMLE-GFP, HMLE-Twist and HMLE-Snail, they continually declined in HMLE-ZEB1 cells (Figure 7). This supported the scenario

describing outgrowth of a minor, pre-existing population, but could not exclude the additional contribution of an MET.

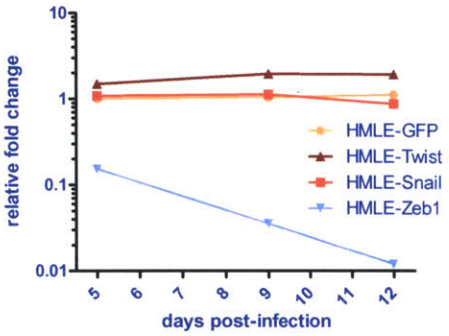


Figure 7. Levels of mStrawberry mRNA over time, relative to HMLE-GFP at day 5. Levels are relatively stable in HMLE-GFP, HMLE-Twist and HMLE-Snail cells over this timeframe, but drop more than tenfold in HMLE-ZEB1 cells.

Despite continuous selection with puromycin and fluorescent cell sorting based on mStrawberry fluorescence, the ZEB1 vector-infected mesenchymal cells could not be maintained from a polyclonally infected population, being repeatedly overgrown by epithelial cells (Figure 8). The FACS profile of HMLE-GFP cells had shown that a very small proportion of infected cells could express the post-2A polypeptide without the pre-2A polypeptide (Figure 2), so it is possible that in the ZEB1 vector-infected population, there were some cells that were mStrawberry-positive and ZEB1-negative. These would remain epithelial but pass through both selection mechanisms. For whatever reasons, the inability to keep cells stably mesenchymal made it difficult to test conditions that would actively induce an MET, since the appearance of epithelial cells could be due to this spontaneous process rather than the condition under study.

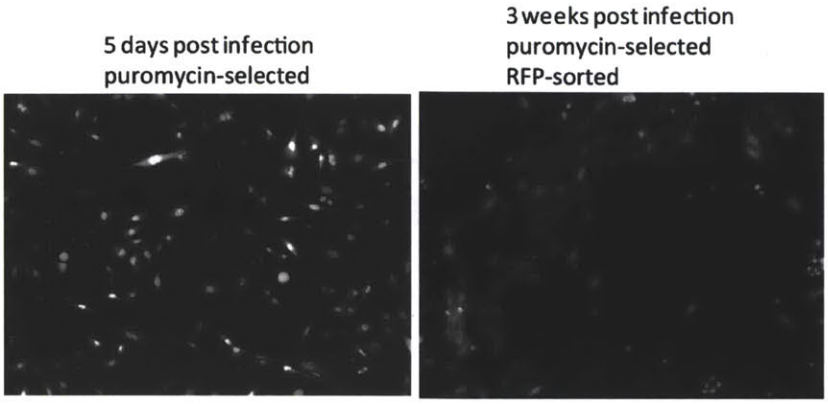


Figure 8. Morphology of HMLE-ZEB1 cells by RFP fluorescence. Left, cells were infected and placed on puromycin selection. Mesenchymal appearance is evident. Right, the same cells, after further sorting for positive mStrawberry fluorescence and culturing for more than two weeks. Epithelial islands form on the culture dish. Despite two-factor selection and confirmation of infection by fluorescence, cells could not maintain mesenchymal phenotype over long periods of culture.

3.2.6 ZEB1-inducible single-cell clones can stably maintain a mesenchymal state

I chose to use single-cell cloning to resolve the issue of phenotypic instability in the ZEB1 vector-infected cell populations. However, single-cell cloning of cells infected with the constitutively expressing vector was not ideal. HMLE cells are known to be intrinsically heterogeneous and to contain a small basal subpopulation of mesenchymal cells (Mani et al., 2008). I could therefore not exclude the possibility that any mesenchymal clone obtained was mesenchymal to start with, rather than having been converted by the ectopic expression of ZEB1.

This logic indicated that an inducible system was required in order to address these issues in a definitive manner. Cells infected with an inducible ZEB1 expression vector would, in principle, remain epithelial so long as ZEB1 expression was not induced. Single-cell clones could be isolated while they were still epithelial and uninduced, and tested afterwards for their respective abilities

to undergo a ZEB1-induced EMT. Thus, I would eliminate the possibility that the mesenchymal cells obtained were derived from the pre-existing mesenchymal subpopulation. Moreover, if a single-cell clone underwent a spontaneous MET following the initial EMT, doing so even under continuous induced expression of ZEB1, this would support the model of a transient, reversible EMT in response to prolonged exposure to ZEB1. On the other hand, if a single-cell clone could be stably maintained in the mesenchymal state, whether in the presence or absence of the inducing signal, it could be used to test conditions that might induce MET.

The first attempt to create an inducible system focused on fusing ZEB1 with a modified estrogen-receptor (ER) fragment (Littlewood et al., 1995), to enable induction by 4-hydroxytamoxifen (OHT). This system would allow rapid induction kinetics, as there would be no time-lag from transcription of new mRNA and translation of new protein. Unfortunately, HMLE cells infected with a viral vector encoding ZEB1 fused with ER on the C-terminus underwent EMT even before OHT was introduced. The same occurred with an N-terminal fusion of ER (Supplementary Figure 1A). Further addition of OHT did not appear to bring about any additional change (Supplementary Figure 1B). The tamoxifen-inducible system was hence deemed to have too high a level of uninduced background activity for practical use.

A doxycycline-inducible expression vector was therefore employed instead. The ubiquitin promoter of the FUW virus was replaced with multiple TetO response elements (Gossen, 1992). This new construct (Figure 9), along with another virus encoding the reverse tetracycline transactivator (rtTA), could be introduced into HMLE cells without immediately triggering an EMT. As I then observed, exposure to doxycycline resulted in an EMT as anticipated. Epithelial single-

cell clones (SSC) that could express ZEB1 under doxycycline control were then derived (ZEB1-Inducible Clones, ZIC).

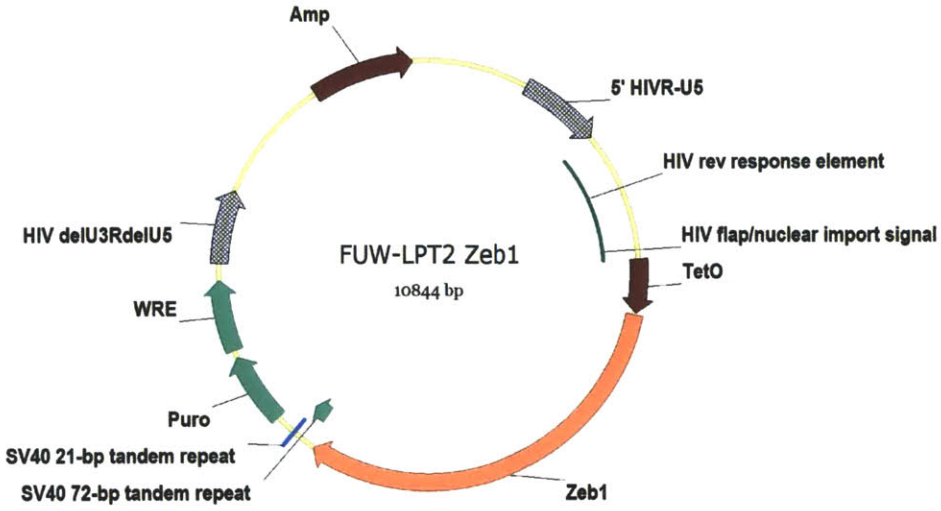


Figure 9. Vector map of modified lentivirus encoding doxycycline-inducible ZEB1.

Under constant exposure to doxycycline for over 1 month, ZIC cells underwent EMT and retained mesenchymal morphology (Supplementary Figure 2), no longer giving rise to epithelial cells like the constitutively expressing polyclonal HMLE-ZEB1 population described in Section 3.2.5. The cells still underwent EMT in a rapid fashion, such that by day 8 post-induction, they had already fully taken on the appearance of mesenchymal cells. Of note, subsequent withdrawal of doxycycline after 8 days did not result in a reversion to the epithelial phenotype, at least for the one month duration of the experiment performed (Supplementary Figure 2).

In fact, doxycycline could be withdrawn even before the appearance of mesenchymal morphology, and the cells would still go on to undergo EMT. I attempted to find the minimum exposure time to doxycycline required to induce EMT. Cells were exposed transiently to doxycycline for between 6

and 24 hours, then passaged as normal in the absence of doxycycline for seven days. Exposure to doxycycline for 24 hours was sufficient to convert all cells to the mesenchymal state at the end of the experiment. Lower exposure times gave rise to an increasingly mixed population of epithelial and mesenchymal cells (Supplementary Figure 3). Subsequently, I found that elevated protein levels of ZEB1 could persist for more than 24 hours after withdrawal of doxycycline (Supplementary Figure 4). This suggested that finding the minimum exposure time to doxycycline would not be a meaningful pursuit, since it would not bear a straightforward relationship to the actual exposure time of the cells to ectopically expressed ZEB1.

3.2.7 Endogenous ZEB1 is necessary for maintaining the mesenchymal state

Having established lines that would not spontaneously revert to epithelial cells after EMT, it became possible to investigate the mechanisms that kept them in their mesenchymal state, and what could cause them to revert to an epithelial state. As previously stated, ZIC cells could maintain a mesenchymal phenotype after 8 days of doxycycline exposure followed by withdrawal of doxycycline for one month(Supplementary Figure 2). Under these conditions, the continued expression of vector-transduced ZEB1 mRNA was unlikely following withdrawal of doxycycline, leaving unexplained why and how the transiently doxycycline-exposed cells continued to reside in the mesenchymal state.

I recalled the double-negative feedback loop existing between ZEB1 and the miR-200 miRNA family (Burk et al., 2008; Bracken et al., 2008), and noted its potential to behave as a bistable switch. I thus hypothesized that exogenous ZEB1 encoded by the viral vector could spur the

upregulation of endogenous ZEB1 encoded by the cellular genome, perhaps by repressing miR-200 expression and in turn lifting its repressive effect on endogenous ZEB1. The elevated levels of endogenous ZEB1 could then take the place of exogenous ZEB1 in maintaining the mesenchymal phenotype. Indeed, the levels of endogenous ZEB1 transcript were elevated in cells that were exposed to doxycycline for 8 days, and these levels remain high after doxycycline was removed for a further 8 days (Figure 10A). ZEB1 protein levels were elevated after doxycycline exposure for 24 hours, fell over the next 48 hours following doxycycline withdrawal, then climbed again over the next five days, indicating the buildup of endogenous ZEB1 protein after withdrawal of exogenous ZEB1 (Supplementary Figure 4). These results supported the hypothesis that the ectopically expressed ZEB1 was inducing expression of its endogenous counterpart, which thereafter maintained its own expression through some type of positive feedback loop.

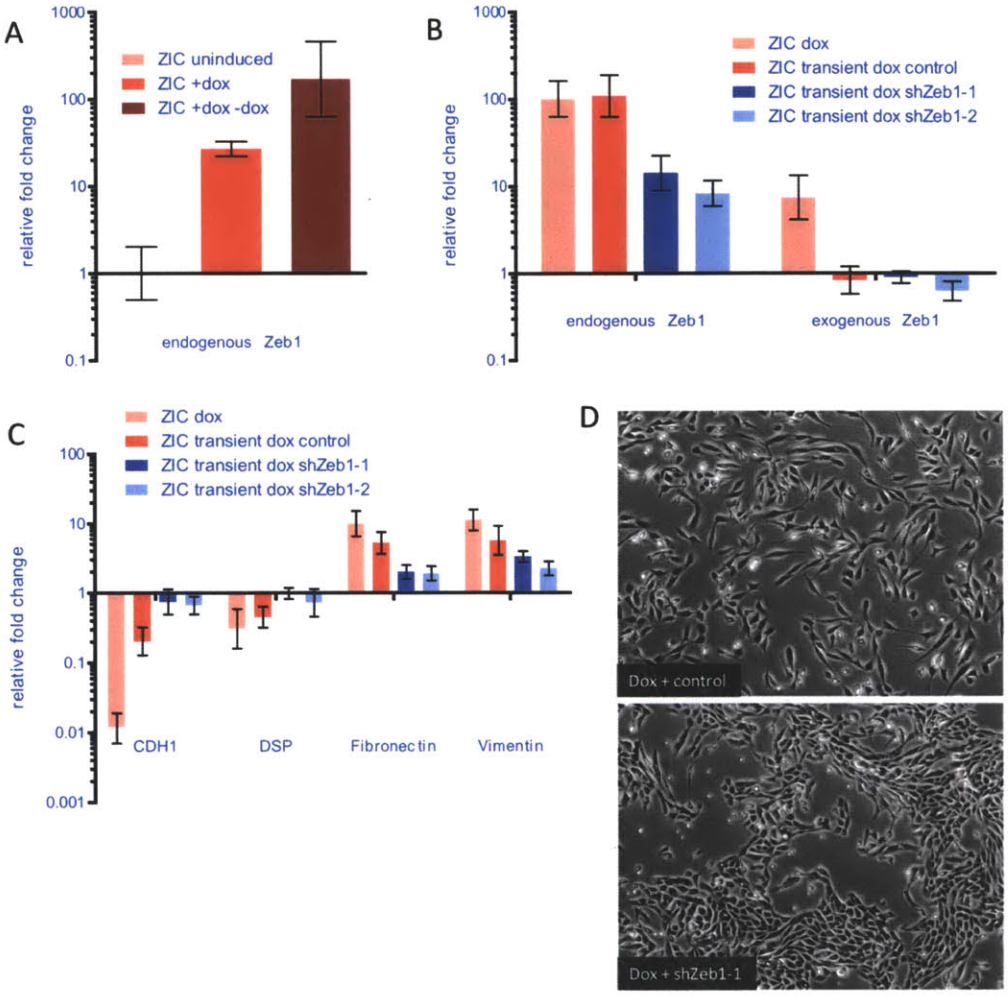


Figure 10. (A) Levels of endogenous ZEB1 mRNA in uninduced ZIC, cells that were exposed to 8 days of doxycycline (ZIC + dox), and cells that had been exposed to 8 days of doxycycline and a further 8 days without doxycycline (ZIC +dox –dox). Endogenous ZEB1 mRNA stays elevated even after doxycycline is withdrawn. (B) Levels of ZEB1 mRNA relative to uninduced ZIC, showing knockdown of endogenous ZEB1 by two shRNA hairpins. Exogenous ZEB1 mRNA from the lentiviral construct falls to basal levels after doxycycline is withdrawn, regardless of knockdown. (C) EMT markers relative to uninduced ZIC, showing that ZEB1 knockdown shifts the profile from mesenchymal to epithelial. (D) Morphology of ZIC after transient doxycycline followed by infection of control virus (top), or virus coding for shRNA against ZEB1 (bottom).

To directly address the issue of whether ZEB1 was necessary to maintain the mesenchymal phenotype, I constructed two lentiviral vectors, each expressing a different shRNA directed against ZEB1. ZIC cells that had been transiently exposed to 8 days of doxycycline and thus already

undergone EMT were further infected with the shRNA vectors, in order to force the downregulation of endogenous ZEB1 in these cells. Both vectors could reduce endogenous ZEB1 levels by an order of magnitude, but could not bring levels back to pre-doxycycline-exposure levels (Figure 10B). Nevertheless, this resulted in cells that upregulated epithelial genes such as E-cadherin and downregulated mesenchymal genes such as fibronectin (Figure 10C). Knockdown of ZEB1 additionally caused the cells to reacquire an epithelial morphology (Figure 10D). To my knowledge, this is the first definitive demonstration that HMLE cells that had undergone an EMT could further undergo an MET. These findings indicated that ZEB1 was not only sufficient to induce the EMT in HMLE cells, but its ongoing expression, driven by the endogenous ZEB1 gene, was necessary to maintain the mesenchymal state. The fact that the EMT status of a cell could be determined by manipulation of a single factor (ZEB1) hinted at its importance in the EMT regulatory circuit.

3.2.8 ZEB1 overrides miR-200c in determining EMT status

As concluded above, the ability of ZIC cells to undergo stably maintain the mesenchymal state after transient doxycycline exposure suggested that exogenous ZEB1 can trigger the upregulation of endogenous ZEB1, which is then maintained in the upregulated state even after the initial trigger is removed. The ZEB1-miR-200 control loop, with its predicted ability to act as a bistable switch, offered a perfect explanation for the observed effects. The loop dictates that elevated ZEB1 and elevated miR-200 cannot coexist in the same cells, and that once either member of the loop dominates, it will retain its dominance and suppress the other member until some external factor intervenes.

Most studies have assumed that within this double-negative-feedback loop, miR-200 influences the EMT primarily through its antagonism of ZEB1. However, this has never been rigorously demonstrated. ZEB1 is known to bind the promoters of several epithelial genes (Eger et al., 2005; Aigner et al., 2007a, 2007b), but the mechanistic connection between ZEB1 and the mesenchymal genes is unknown. In order to dissect the respective roles of ZEB1 and miR-200, it became necessary to decouple the regulation of both factors by one another.

The ZIC cells expressed exogenous ZEB1 mRNA that did not encode the native 3'-UTR and was therefore not subject to the repressive effects of miR-200. However, the miR-200 levels still remain affected by ZEB1 transcriptional repressive activity, as seen by the reduction in the levels of miR-200c upon doxycycline induction of ZEB1 (Figure 11). Indeed, the levels of miR-200c continued to remain low long after doxycycline was withdrawn, as expected if elevated endogenous ZEB1 continued to repress its expression.

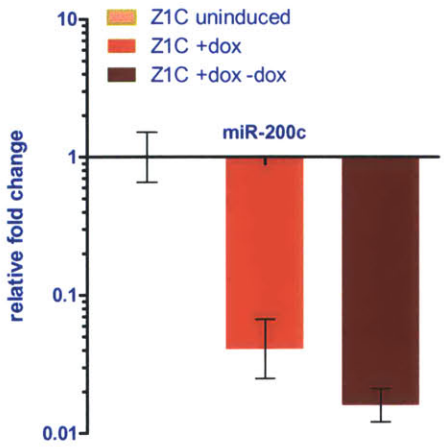


Figure 11. Levels of miR-200c in ZIC cells upon addition and subsequent withdrawal of doxycycline. miR-200c levels fall upon doxycycline exposure and stay depressed afterwards.

In order to decouple miR-200c expression from ZEB1 transcriptional repressive activity, the miR-200c genomic region was cloned into a constitutively active lentiviral construct, which was infected onto ZIC cells that had not previously been exposed to doxycycline. This exogenous, vector-driven miR-200c expression was driven by a CMV promoter and was thus not subject to repression by ZEB1, allowing miR-200c levels to remain elevated regardless of ZEB1 levels. In this way, I engineered cells that, upon doxycycline exposure, were able to simultaneously maintain high levels of ZEB1 and miR-200c. Under these circumstances, if the epithelial markers directly under ZEB1 control are repressed, but the mesenchymal markers are unchanged, this would strongly suggest that control of the mesenchymal genes lies downstream of miR-200c, and not of ZEB1.

In fact, keeping miR-200c levels elevated did not prevent EMT from taking place under doxycycline induction, as evidenced by the change in morphology (Figure 12C) as well as both the downregulation of epithelial markers (Figure 12A) and concomitant upregulation of mesenchymal markers (Figure 12B). Importantly, total miR-200c levels in the ZIC-miR-200c cells exceeded that of epithelial ZIC cells at all times (Figure 12A), demonstrating that whatever effects ZEB1 had on the cells, downregulation of miR-200c was not a critical, obligatory step in this process. The fact that EMT occurs when miR-200c levels remained elevated and essentially unchanged indicated that the levels of ZEB1, not miR-200c, primarily governed the EMT program.

A known direct target of miR-200c is fibronectin, which has been posited to be involved in cell motility and anoikis (Howe et al., 2011). This knockdown effect of miR-200c on fibronectin expression was indeed observed in the ZIC-miR-200c cells relative to ZIC-control cells exposed to

doxycycline (Figure 12B). However, the presence of high miR-200c did not, on its own, prevent the upregulation of fibronectin mRNA by exogenous ZEB1, indicating that ZEB1 could exert an effect on fibronectin expression through a miR-200c-independent pathway.

As a control, I demonstrated the miR-200c lentiviral vector to be functional by removing doxycycline in the cells described above. Without exogenously expressed, vector-driven ZEB1 expression, the exogenous miR-200c, which remained elevated, dominated the control loop and acted to repress endogenous ZEB1 (Figure 13B), causing an MET to occur. The cells re-expressed E-cadherin and downregulated fibronectin (Figure 13A and C). This validated the use of the miR-200c expression vector and its effectiveness in inducing an MET when not impeded by exogenous ZEB1.

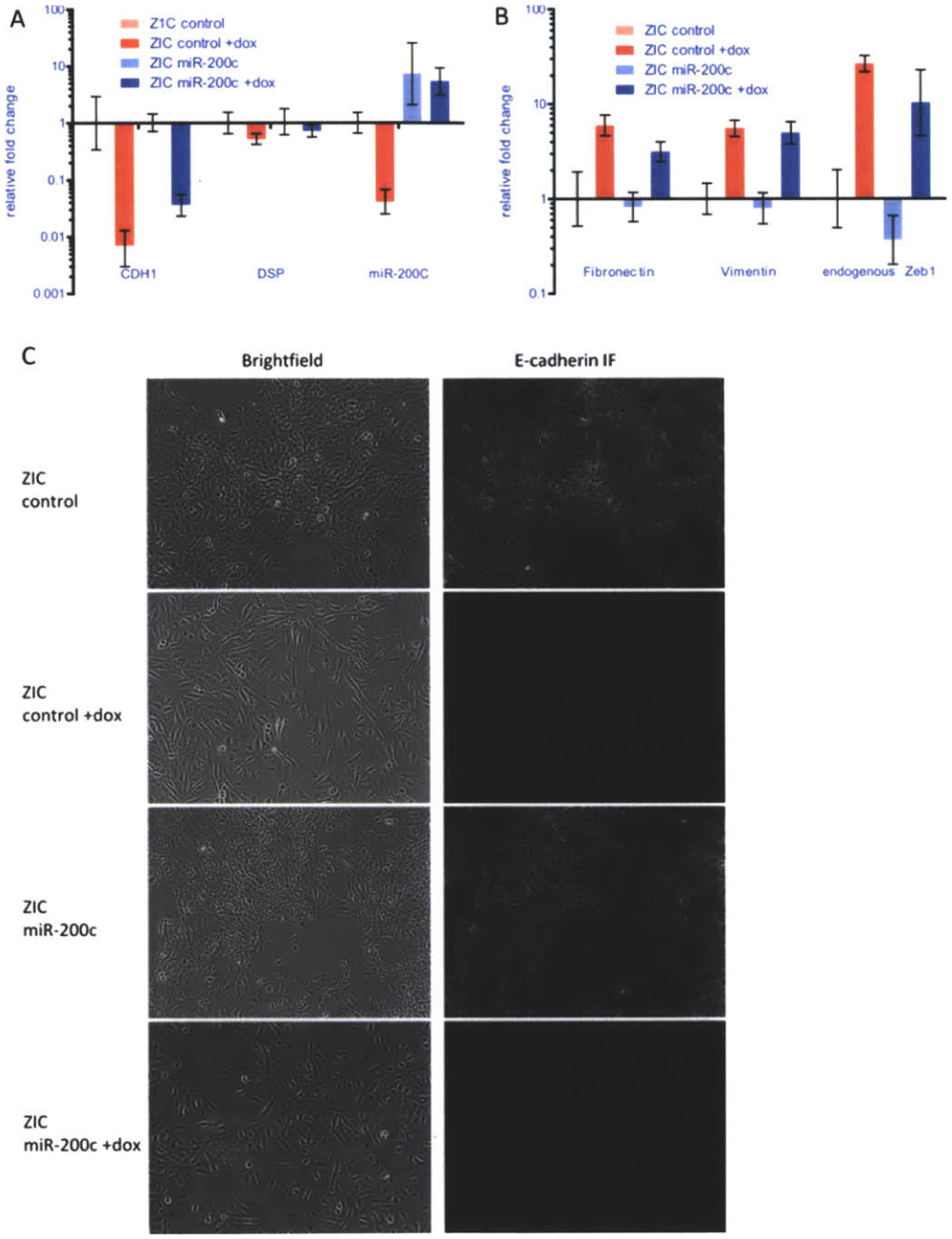


Figure 12. (A) Epithelial EMT markers in ZIC cells overexpressing exogenous miR-200c. Cells overexpressing miR-200c downregulate epithelial markers but continue to retain high levels of the microRNA even upon doxycycline induction of ZEB1. (B) Mesenchymal EMT markers in the same cells, shown to be elevated upon doxycycline exposure. Endogenous ZEB1 mRNA is detected at a lower level in ZIC miR-200c cells, but is still upregulated upon doxycycline induction. (C) Morphology of ZIC miR-200c cells upon exposure to doxycycline. Both morphology and immunofluorescence of E-cadherin agree with realtime PCR data on the EMT status of the cells.

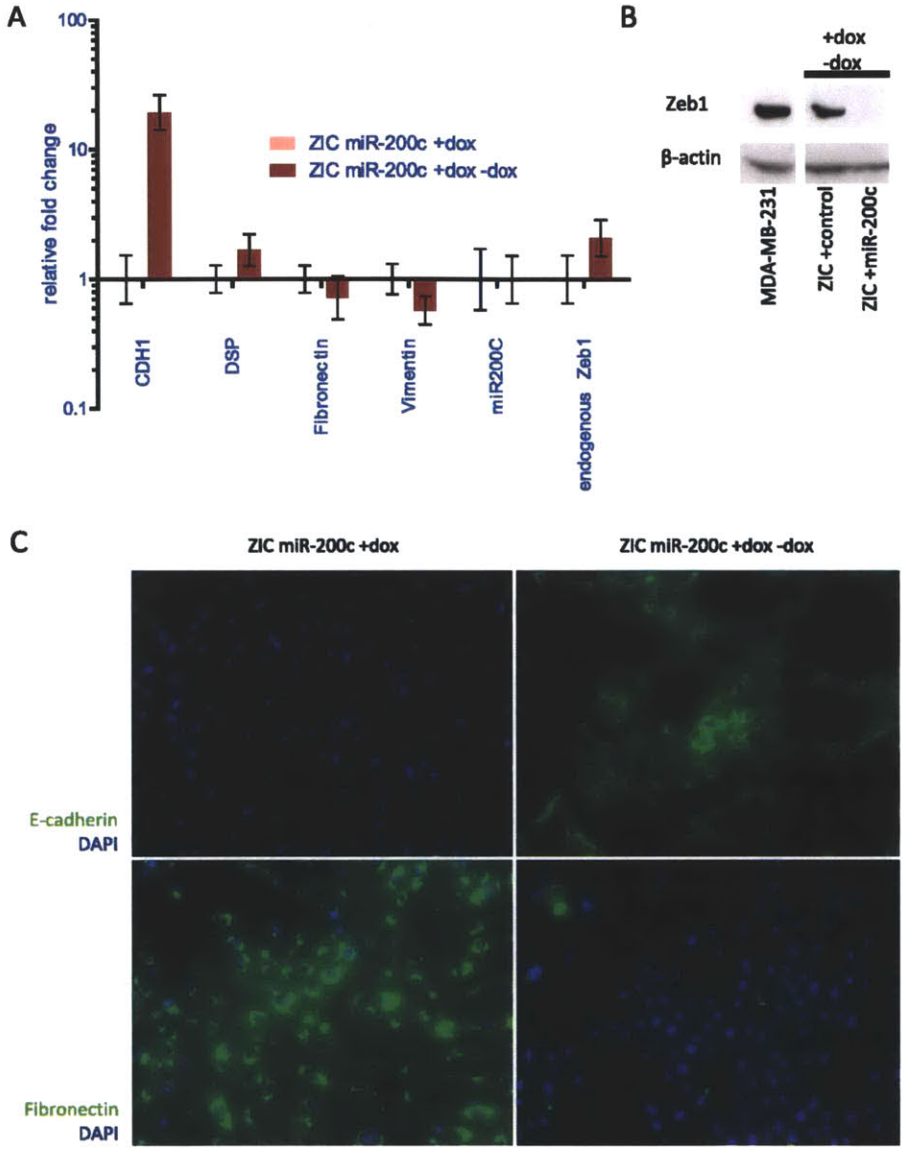


Figure 13. (A) Realtime PCR of EMT markers, relative to ZIC mIR-200c cells exposed to doxycycline. Upon withdrawal of doxycycline, epithelial markers are upregulated while mesenchymal markers are downregulated, indicating the occurrence of an MET. miR-200c levels remain unchanged, while endogenous ZEB1 transcript is not downregulated. (B) Western blot confirming the downregulation of ZEB1 protein in ZIC miR-200c cells after doxycycline is withdrawn. (C) Immunofluorescence of ZIC miR-200c cells after exposure to and withdrawal of doxycycline. In agreement with realtime PCR data, epithelial E-cadherin is re-expressed, while mesenchymal fibronectin in downregulated upon withdrawal of doxycycline.

While miR-200c can reduce endogenous ZEB1 mRNA and protein levels through RNAi, it is not thought to directly affect transcription initiation of the ZEB1 gene. If exogenous ZEB1 regulates endogenous ZEB1 mRNA levels solely through the miR-200 RNAi mechanism, then holding miR-200c at a constant elevated level should render the endogenous ZEB1 mRNA levels unresponsive to exogenous ZEB1. However, this was not found to be the case. Overexpressing miR-200c reduced overall levels of endogenous ZEB1 mRNA compared to control cells, but did not prevent endogenous ZEB1 mRNA from being upregulated upon doxycycline exposure (Figure 12B), even though miR-200c levels were not significantly altered (Figure 12A). In fact, the fold-change of ZEB1 mRNA in response to doxycycline was almost identical whether the cells were overexpressing miR-200c or not. This indicated that in addition to influencing endogenous ZEB1 expression through the known ZEB1-miR-200 control loop, exogenous ZEB1 could be upregulating endogenous ZEB1 expression through an unknown miR-200c-independent mechanism.

3.2.9 ZEB1 is a mediator of Snail-induced EMT

Multiple factors can induce an EMT, but there have been few attempts to arrange the factors into a hierarchy or explore their interdependence. In one study, forced re-expression of E-cadherin in cells that had undergone Twist-induced EMT failed to reverse cell morphology or any of the EMT markers, other than E-cadherin itself (Yang et al., 2004). This suggested that Twist does not act through E-cadherin downregulation to bring about the other changes involved in an EMT. In another, TGF-β was shown to upregulate both Snail and ZEB1 in NMuMG cells, while knockdown of Snail or Twist reduced ZEB1 levels in SW-620 cells (Dave et al., 2011). However, this study did not explore whether TGF-β required the action of Snail or ZEB1 to mediate the EMT, or whether

ZEB1 downregulation was a necessary step for MET triggered by Snail or Twist knockdown. I hypothesized that ZEB1 is not just upregulated by the other EMT-inducing signals, it is actually an essential component of their ability to mediate an EMT or maintain the mesenchymal state.

I wanted to test this hypothesis by examining the role of ZEB1 in Snail-induced EMT. In a manner analogous to my isolating the ZIC cells, I created a doxycycline-inducible Snail expression construct and obtained Snail-inducible single-cell clones (SIC). Unfortunately, I was unable to observe any change in cell morphology after exposure to doxycycline for more than one month (data not shown), even though Snail expression was confirmed (Supplementary Figure 5). The inability of an inducible Snail system to trigger EMT was corroborated by other members of the lab working with an independent inducible construct. The difference in the phenotype obtained between the constitutive expression construct and the inducible construct is probably due to the level of expression. The ZEB1 constitutive expression construct produced higher levels of protein than the equivalent inducible construct, and the relationship between the two Snail constructs is probably similar. This would appear to suggest that Snail can only induce EMT in HMLE cells when it is expressed under the control of a very strong promoter.

I wished nonetheless to test directly whether the EMT induced by Snail depended on ZEB1 function. Since I was unable to establish a usable system where Snail could be turned on and off at will, I simply infected HMLE cells with the constitutively active Snail-expressing lentiviral vector described in Section 3.2.3 and allowed the cells to undergo EMT. I then further infected these cells with the shRNA lentiviral vectors directed against ZEB1, described in Section 3.2.7.

The resulting cells did not fully reacquire the epithelial phenotype, but did express higher levels of E-cadherin and lower levels of fibronectin (Figure 14). This partial reversion from the mesenchymal state could be a result of incomplete knockdown of ZEB1, but the possibility that Snail can act independently of ZEB1 to enact some components of the EMT program could not be ruled out. Nevertheless, the results indicated that Snail-induced EMT in HMLE cells was maintained to some extent by ongoing actions of ZEB1. Of particular note was the fact that despite Snail and ZEB1 both being able to transcriptionally repress E-cadherin, it appeared that endogenous ZEB1 was still required for E-cadherin repression in this model, even in the presence of highly overexpressed Snail. This raised the question of whether Snail expressed at physiological levels ever directly targets E-cadherin, or whether it regulates E-cadherin expression indirectly through another factor, such as ZEB1.

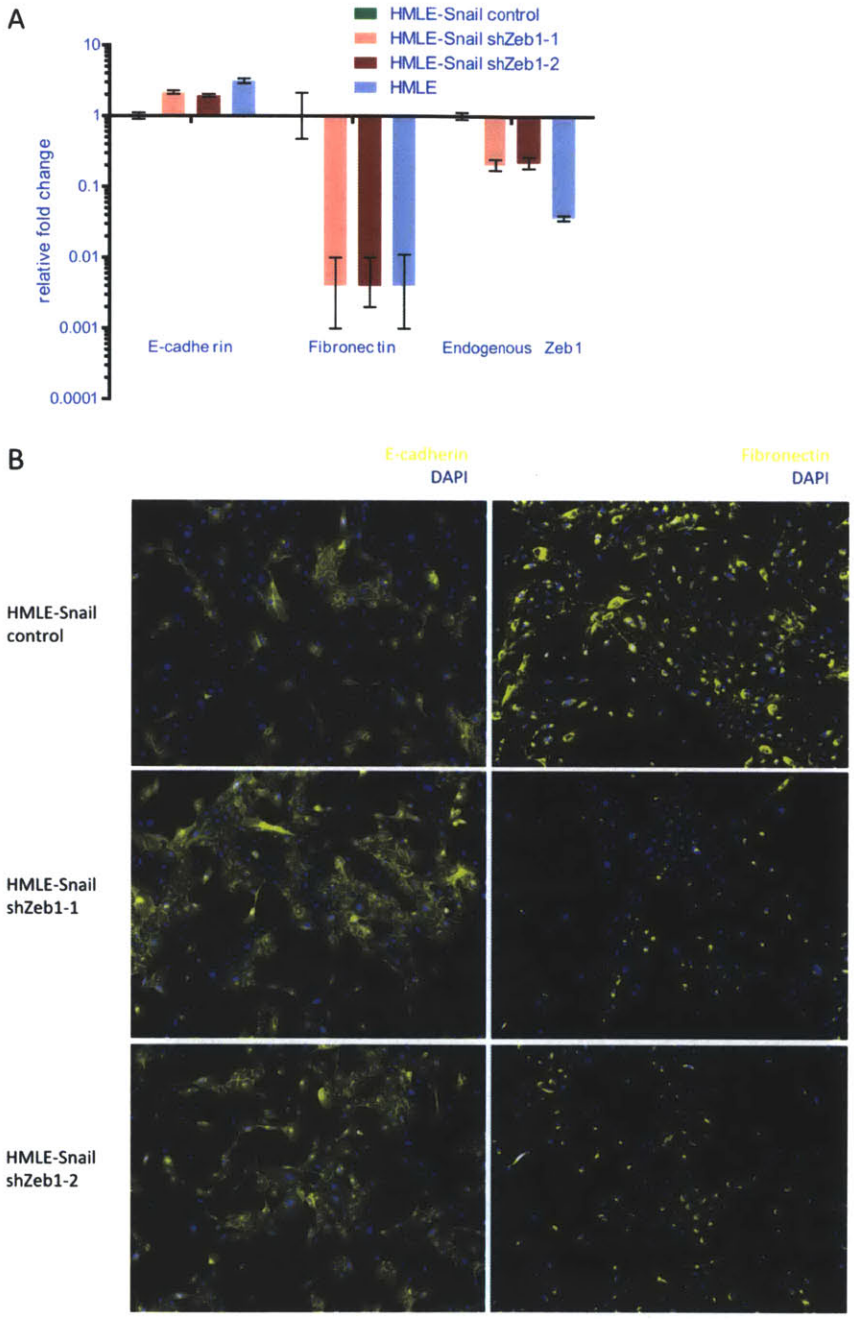


Figure 14. (A) EMT markers in HMLE-Snail cells expressing shRNA against ZEB1, relative to cells infected with control shRNA virus. ZEB1 knockdown causes re-expression of E-cadherin and downregulation of fibronectin, to levels approaching that of unmodified epithelial HMLE cells. (B) Immunofluorescence showed that not all the polyclonally infected HMLE-Snail cells downregulate E-cadherin, but a higher proportion of cells express E-cadherin after ZEB1 knockdown. Fibronectin expression, while not completely eliminated at the protein level, is still visibly reduced upon ZEB1 knockdown.

3.2.10 Interfering with the mesenchymal state of ZEB1-induced cells

As described in section 3.2.6, ZIC cells transiently exposed to doxycycline could maintain their mesenchymal state in the absence of further ectopically expressed ZEB1. This state was maintained by endogenous factors, among them endogenous ZEB1. The induction of MET was possible in these cells, as demonstrated earlier by responses observed upon knockdown of the endogenously expressed ZEB1 and by the forced overexpression of miR-200c. For these reasons, I believed that cells that had been exposed to a brief period of ZEB1 would be suitable for studying various signals that might perturb the resulting mesenchymal state and thus cause a reversion to the epithelial state. Snail did not appear to be upregulated, or indeed detectable at all, in ZEB1-induced EMT (Supplementary Figure 5), making it unlikely that Snail was involved in this maintenance. I did attempt Twist knockdown in these cells, but three different shRNAs directed against Twist merely killed cells (data not shown). Hence, other than the fact that Twist appeared to be an essential gene for the survival of these cells, I was unable to draw any further conclusions about the role of Twist in ZEB1-induced EMT or the maintenance of the ZEB1-induced mesenchymal state.

I looked at the role of TGF-β signaling in the maintenance of the mesenchymal state by exposing the mesenchymal cells to the small molecule inhibitor SB431542, which selectively targets the kinase activity of the ALK5 TGF-β type I receptor and to a lesser extent the kinase domains of other ALKs (Inman, 2002; Mordasky Markell et al., 2010). The results were mixed, as cells with epithelial morphology appeared on the culture plate, but no apparent changes in the levels of E-cadherin, fibronectin or ZEB1 were detected (Supplementary Figure 6).

In summary, I was unable to demonstrate a reversal of ZEB1-induced EMT by perturbing anything other than ZEB1 itself. ZEB1 may lie downstream of other EMT-inducing signals such as TGF-β and Snail, and as such may not depend on their activity to trigger EMT.

3.2.11 Downstream target genes of ZEB1

ZEB1 appeared to more directly impinge on the EMT program in HMLE cells than several of the other known EMT-inducing factors. When choosing between ZEB1 and miR-200c, I had concluded that the levels of ZEB1 mattered more in determining EMT status; this suggests, in turn, that the important mediators of EMT lie downstream of ZEB1 rather than miR-200c. For this reason, I deemed it useful to identify direct downstream targets of ZEB1.

In principle, downstream targets of a transcription factor can be identified biochemically or kinetically. Biochemically, the binding of the transcription factor protein to its target promoter regions can be detected using a chromatin immunoprecipitation procedure. Kinetically, gene expression profiling can be performed at short time intervals after induction of the transcription factor activity, in order to find the first genes that respond to the change in transcription factor activity. Each approach has its own drawbacks. The binding of a transcription factor to a region of DNA does not necessarily have functional significance, particularly if the transcription factor is overexpressed beyond physiological levels. The chromatin immunoprecipitation procedure also depends critically on the specificity of the antibody used. The kinetic approach shows functional changes in levels of mRNA, but does not necessarily reflect a direct action of the transcription factor on the gene target. This procedure requires a strong, synchronous and rapid activation of

the transcription factor activity in order to minimize the time interval between initiating the transcription factor activity and collecting the samples. The longer the time interval, the more likely the changes detected were due to secondary, indirect effects rather than direct action of the transcription factor. I chose the kinetic approach, performing a microarray analysis of the change in transcript levels in ZIC cells shortly after doxycycline exposure. In fact, the doxycycline-inducible system was not ideal for this, as there is a time-lag between the induced transcription and accumulation of ZEB1 protein. However, given that the tamoxifen-inducible system, which in principle could trigger a more rapid response, did not behave as hoped, the doxycycline system was used instead.

Samples were taken at 14 and 24 hours post-doxycycline exposure, and compared with time-matched, no-doxycycline controls; both sets of transcripts were then analyzed on Agilent GE 60k arrays. While there are undoubtedly many ways to produce a list of candidate genes that were differentially regulated at these time points, based on varying fold-change or p-value cutoffs, I chose to isolate a high-confidence set of genes using the adjusted p-value cutoff of 0.05, thereby ignoring certain candidates lost in the cutoff process. Using this cutoff criterion, 36 genes behaved consistently across the two time points. The rVISTA tool described in Section 2.2.3 (Loots et al., 2002) recognized 34 of these genes and called 20 genes with conserved AREB6 (ZEB1) binding sites in their promoters (Table 1). This high proportion indicated the experiment was broadly successful at enriching for ZEB1 downstream targets. The failure to call E-cadherin (CDH1), which is a known direct target with conserved ZEB1 binding sites in its promoter, suggests that yet other genes were

missed as well. Nevertheless, this list brought up several genes of interest, subdivided into genes that had previous association with ZEB1 or EMT, and genes with no previous association.

Gene Symbol	Log2 fold change at 14 hours dox	Log2 fold change at 24 hours dox
ACOT11	-0.577	-0.699
AGR2	-0.588	-1.40
AP1M2	-0.675	-1.59
CBX6	0.539	0.722
CDC42BPG	-0.702	-0.756
CRB3	-1.10	-1.53
CXADR	-0.620	-0.846
ELOVL7	-0.548	-0.638
EPHA1	-0.875	-0.941
ESRP1	-0.943	-0.899
GNG11	0.732	0.988
GRHL2	-1.39	-1.45
ITGB6	-1.08	-0.866
KIAA1324	0.511	0.820
LOXL3	-0.949	-0.984
MARVELD3	-1.47	-1.35
OVOL2	-0.943	-1.19
PHIP	1.73	1.82
TFAP2C	-0.576	-1.06
TMEM125	-1.92	-2.17

Table 1. Genes containing conserved ZEB1 binding sites on their promoters, responding consistently in ZIC cells between 14 and 24 hours after exposure to doxycycline.

Among these genes, the cell polarity gene Crumb3 (CRB3) plays a role in the formation of tight junctions in epithelial cells (Roh et al., 2003; Lemmers et al., 2004; Fogg et al., 2005). It was previously identified as a direct target of ZEB1 in MDA-MB-231 cells. CRB3 was upregulated in an MET triggered by knockdown of ZEB1, and ZEB1 was found to bind directly to its promoter (Aigner

et al., 2007a). Coxsackie virus and adenovirus receptor (CXADR) is a component of tight junctions in epithelial cells, and previously found to be downregulated in EMT (Lacher et al., 2006). CXADR was subsequently identified as a direct target of ZEB1, again by direct physical interaction of ZEB1 protein with the CXADR promoter (Lacher et al., 2011). Epithelial splicing regulatory protein (ESRP1) was previously associated with the EMT, being present in epithelial HMLE cells and absent in cells that had undergone Twist-induced EMT (Shapiro et al., 2011). Overexpression of ESRP1 inhibited EMT by Twist, while knockdown of ESRP1 enhanced TGF-β-induced EMT (Brown et al., 2011). Furthermore, ESRP1 was identified as a ZEB1-responsive gene in non-small cell lung cancer (Gemmill et al., 2011), and was also present in my own EMT core signature. The presence of a conserved ZEB1 binding site on the promoter of ESRP1, as found through rVISTA, supports the notion that ZEB1 directly represses ESRP1 to bring about necessary changes in the alternative splicing program during EMT.

Two additional genes with no previous association with ZEB1 or EMT were identified from the list based on their mutant phenotypes in mice. These were Grainyhead-like 2 (GRHL2), a transcription factor, and Ovo-like 2 (OVOL2), a putative transcription factor. Both displayed limbs/digits/tail and craniofacial phenotypes, reminiscent of the phenotype of Twist knockout mice. In OVOL2 loss-of-function mutants, neural crest cells formed, but failed to migrate and underwent apoptosis instead (Mackay et al., 2006). A gain-of-function mutation of GRHL2 caused embryonic lethality and failure to close the neural tube, echoing phenotypes associated with EMT failure (Brouns et al., 2011). A chemically induced loss-of-function mutation of GRHL2 resulted in loss of E-cadherin and other epithelial genes, including CDH3, EPCAM and BCAM (Pyrgaki et al., 2011), which are

represented in the EMT core signature. After I had identified GRHL2 as a potential candidate, it was reported that GRHL2 could indeed suppress the EMT if overexpressed (Cieply et al., 2012). In their report, Cieply et al. showed that GRHL2 could directly regulate ZEB1. If ZEB1 and GRHL2 can directly repress each other, this sets up the potential for another double-negative feedback loop akin to the loop between ZEB1 and miR-200c.

3.3 Conclusions

When I first came upon ZEB1 in my own work, ZEB1 had only just been identified as an EMT-inducing factor, the newest among a list of others such as Snail and Twist. There was little to suggest that ZEB1 was any different from the other factors in its ability to induce EMT or its importance in any putative core EMT regulatory circuitry. ZEB1 became more widely studied during the course of my research, and was demonstrated to be sufficient to induce EMT (Eger et al., 2005) and necessary to maintain the mesenchymal phenotype (Burk et al., 2008) in different cell lines. I demonstrated these two qualities in a single cell line, changing the EMT status of HMLE cells back and forth solely by manipulating ZEB1 levels. Until recently, the relationship between ZEB1 and other EMT-inducing signals was not heavily explored. I provided evidence to suggest that ZEB1 may be qualitatively different from other EMT-inducing signals: ZEB1 can induce EMT on its own more quickly than Snail or Twist and it is necessary for maintenance of the mesenchymal phenotype in Snail-induced EMT. This study suggests that the known EMT-inducing signals may interact with one another in complex ways, and may converge on a common denominator. In HMLE cells, the common denominator appears to be ZEB1.

3.4 Materials and Methods

3.4.1 Cell culture and reagents

Immortalized human breast epithelial cells (HMLE), generated through the introduction of the SV40 early region and hTERT, were maintained as described (Elenbaas et al., 2001). Doxycycline (Sigma) was used to induce at a concentration of 1.0 μg/ml.

3.4.2 mRNA and miRNA extraction and detection

Total RNA containing small RNAs was extracted from cells using the RNeasy Plus Mini Kit with the miRNA supplementary protocol 1 (Qiagen). Reverse transcription was performed using the High Capacity RNA-to-cDNA Kit (Applied Biosystems) for mRNA and the miScript Reverse Transcription Kit (Qiagen) for miRNA. Quantitation of mRNA by realtime PCR was performed on an Applied Biosystems 7900HT using SYBR Green PCR Master Mix (Invitrogen). Quantitation of miRNA was performed with miScript SYBR Green PCR Kit (Qiagen) and microRNA-specific miScript Primer Assays (Qiagen). Microarray quantitation of mRNA was performed using SurePrint G3 Human GE 8x60k microarrays (Agilent).

3.4.3 Immunofluorescence

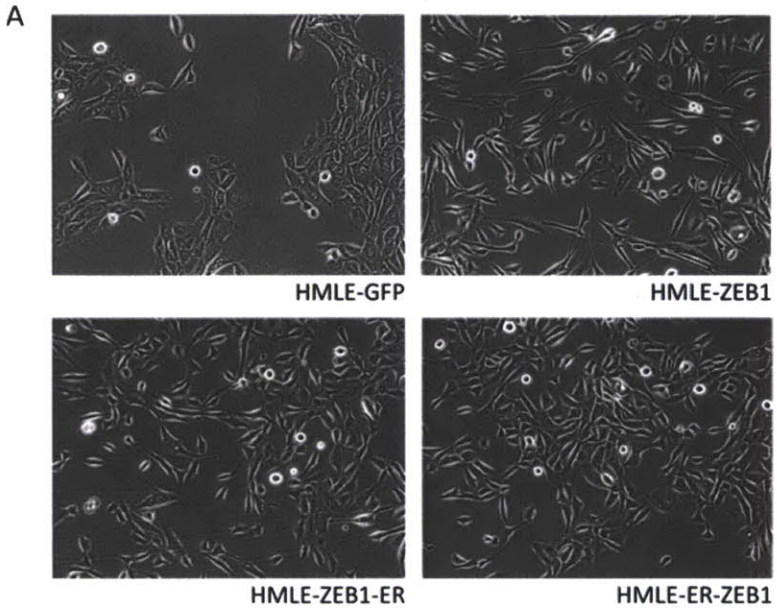
Cells were grown on Chamber Slide System 8-well glass slides (Fisher Scientific), fixed and permeabilized with methanol prior to blocking with 4% BSA (Sigma) in PBS. Slides were mounted with ProLong Gold antifade reagent with DAPI (Invitrogen) which allowed visualization of cell

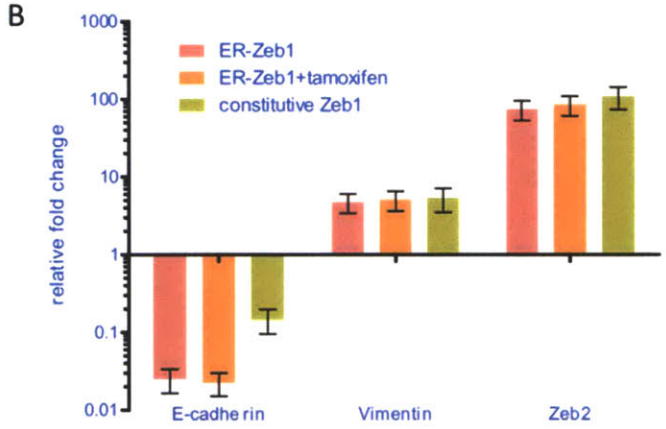
nuclei. Secondary antibodies were goat anti-rabbit or –mouse coupled to Alexa-546 or -647 (Invitrogen).

3.4.4 Transcription factor binding site analysis

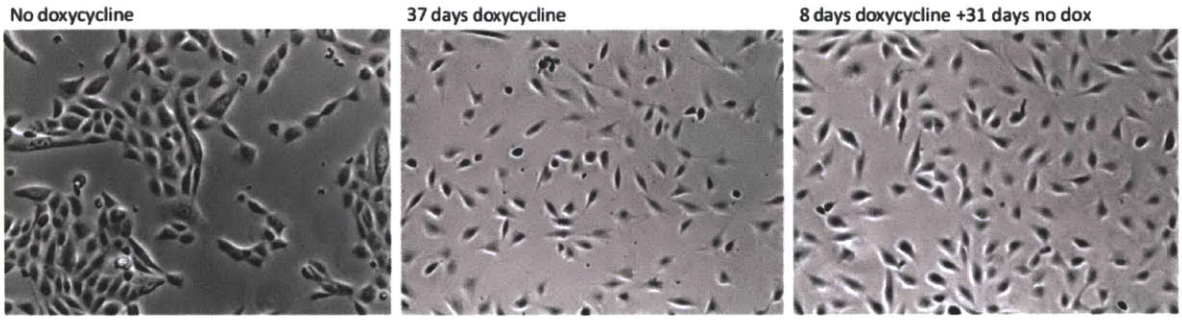
The VISTA suite of online tools (http://genome.lbl.gov/vista/index.shtml) was used for TFBS analysis. Whole genome rVISTA with the precomputed alignments between the Human March 2006 (hg18) assembly and the Mouse February 2006 (mm8) assembly was selected. The promoter region of genes was designated as the 3000 bp upstream of the coding start site. Consistent overlapping probes were converted into unique gene locus link ids before being submitted to analysis.

3.5 Supplementary Data

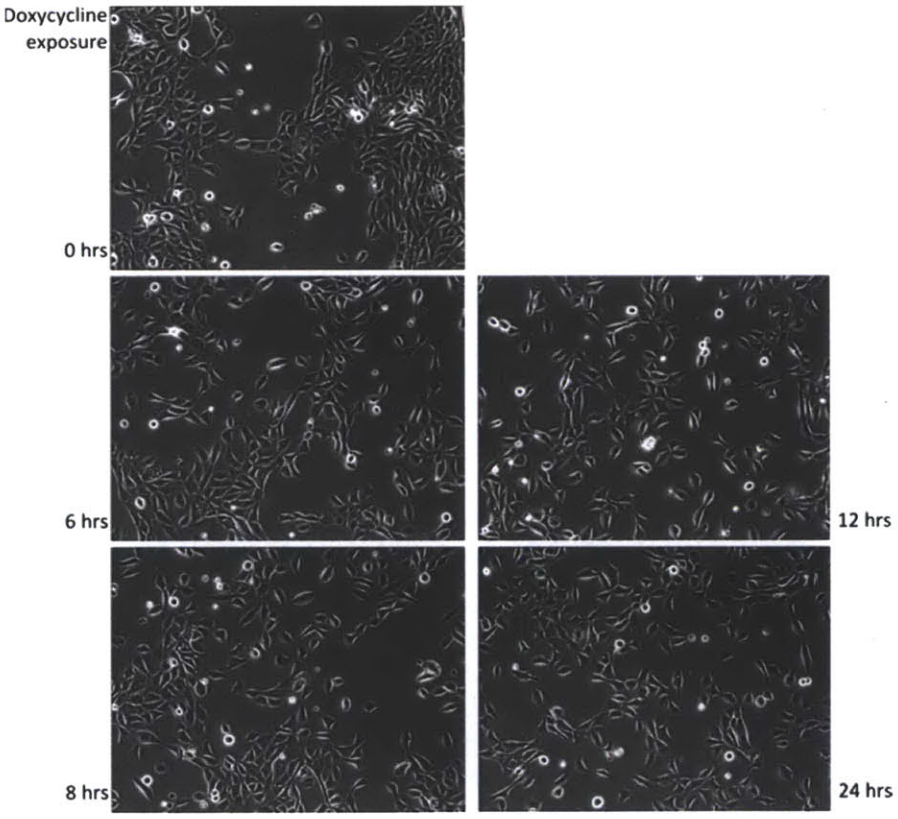




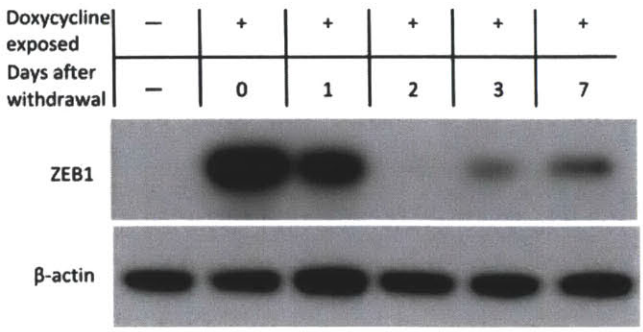
Supplementary Figure 1. (A) Morphology of HMLE cells infected with virus encoding ZEB1-ER or ER-ZEB1, in the absence of OHT, compared to the epithelial HMLE-GFP and mesenchymal HMLE-ZEB1 cells. Even without induction, cells undergo EMT, indicating leakiness in the inducible system. (B) Realtime PCR of selected markers in HMLE-ER-ZEB1 cells, in the presence or absence of OHT, normalized to uninfected HMLE cells. The ER-ZEB1 virus cause changes in EMT markers as effectively as the constitutively active ZEB1 virus, if not more, and additional 4-OHT does not make a difference.



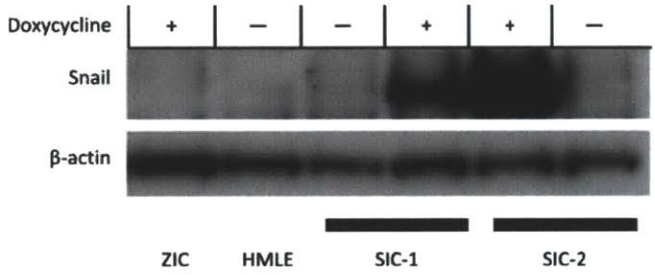
Supplementary Figure 2. Morphology of doxycycline-inducible ZEB1-expressing HMLE single-cell clone, ZIC, under various conditions. Cells are stably epithelial in the absence of doxycycline, undergo EMT and remain mesenchymal for extended periods under constant exposure to doxycycline, and also stay mesenchymal with transient exposure to doxycycline.



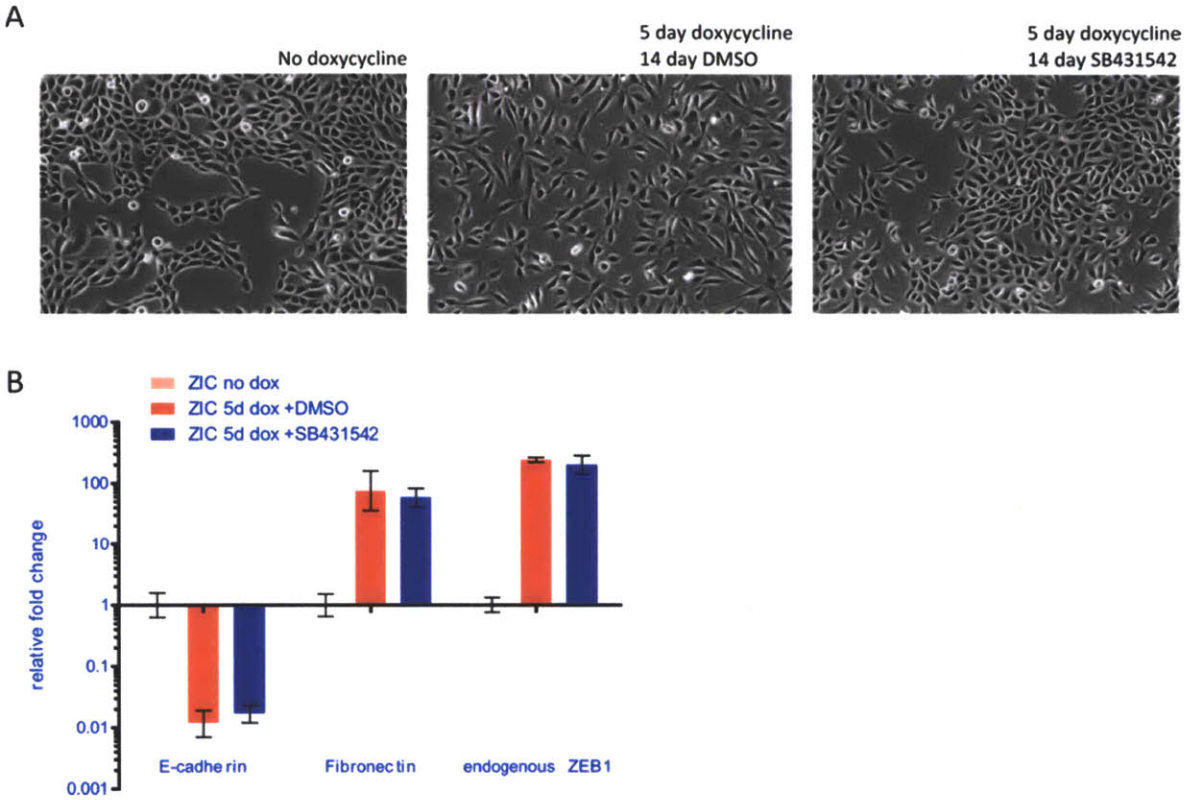
Supplementary Figure 3. Short exposure of ZIC cells to doxycycline, followed by 7 days passage without doxycycline. Shorter exposures give rise to mixed populations of epithelial and mesenchymal cells, with no definite time point below which no EMT occurs, and above which all cells undergo EMT.



Supplementary Figure 4. ZEB1 protein levels in ZIC cells upon 24 hour exposure to doxycycline and subsequent withdrawal. Initial pulse of ZEB1 disappears within two days of doxycycline withdrawal, but levels build up again, suggesting that this latter signal comprises endogenous ZEB1.



Supplementary Figure 5. Western blot confirming expression of Snail in Snail-inducible clones (SIC) upon exposure to doxycycline. ZIC cells in their mesenchymal state after doxycycline exposure also did not appear to upregulate Snail.



Supplementary Figure 6. (A) Morphology of ZIC cells exposed to doxycycline and TGF-β signaling inhibitor SB431542, compared with epithelial and mesenchymal controls. SB431542 appears to promote the appearance of epithelial cells. (B) Realtime PCR of selected EMT markers showing that E-cadherin, fibronectin and endogenous ZEB1 levels appear unaltered, despite the visible change in morphology.

3.6 References

Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., Mikulits, W., Brabletz, T., Strand, D., Obrist, P., Sommergruber, W., Schweifer, N., Wernitznig, A., Beug, H., Foisner, R., and Eger, A. (2007a). The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. Oncogene.

Aigner, K., Descovich, L., Mikula, M., Sultan, A., Dampier, B., Bonne, S., van Roy, F., Mikulits, W., Schreiber, M., Brabletz, T., Sommergruber, W., Schweifer, N., Wernitznig, A., Beug, H., Foisner, R., and Eger, A. (2007b). The transcription factor ZEB1 (deltaEF1) represses Plakophilin 3 during human cancer progression. FEBS Lett *581*, 1617–1624.

Bracken, C. P., Gregory, P. A., Kolesnikoff, N., Bert, A. G., Wang, J., Shannon, M. F., and Goodall, G. J. (2008). A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. Cancer research *68*, 7846–7854.

Brouns, M. R., De Castro, S. C. P., Terwindt-Rouwenhorst, E. A., Massa, V., Hekking, J. W., Hirst, C. S., Savery, D., Munts, C., Partridge, D., Lamers, W., Köhler, E., van Straaten, H. W., Copp, A. J., and Greene, N. D. E. (2011). Over-expression of Grhl2 causes spina bifida in the Axial defects mutant mouse. Human molecular genetics *20*, 1536–1546.

Brown, R. L., Reinke, L. M., Damerow, M. S., Perez, D., Chodosh, L. A., Yang, J., and Cheng, C. (2011). CD44 splice isoform switching in human and mouse epithelium is essential for epithelial-mesenchymal transition and breast cancer progression. The Journal of clinical investigation *121*, 1064–1074.

Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S., and Brabletz, T. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep *9*, 582–589.

Casas, E., Kim, J., Bendesky, A., Ohno-Machado, L., Wolfe, C. J., and Yang, J. (2011). Snail2 is an essential mediator of Twist1-induced epithelial mesenchymal transition and metastasis. Cancer research *71*, 245–254.

Chaffer, C. L., Brueckmann, I., Scheel, C., Kaestli, A. J., Wiggins, P. A., Rodrigues, L. O., Brooks, M., Reinhardt, F., Su, Y., Polyak, K., Arendt, L. M., Kuperwasser, C., Bierie, B., and Weinberg, R. A. (2011). Normal and neoplastic nonstem cells can spontaneously convert to a stem-like state. Proceedings of the National Academy of Sciences of the United States of America *108*, 7950–7955.

Cieply, B., Riley, P., Pifer, P. M., Widmeyer, J., Addison, J. B., Ivanov, A. V., Denvir, J., and Frisch, S. M. (2012). Suppression of the Epithelial-Mesenchymal Transition by Granyhead-like 2. Cancer research.

Creighton, C. J., Li, X., Landis, M., Dixon, J. M., Neumeister, V. M., Sjolund, A., Rimm, D. L., Wong, H., Rodriguez, A., Herschkowitz, J. I., Fan, C., Zhang, X., He, X., Pavlick, A., Gutierrez, M. C., Renshaw, L., Larionov, A. A., Faratian, D., Hilsenbeck, S. G., Perou, C. M., Lewis, M. T., Rosen, J. M., and Chang, J. C. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proceedings of the National Academy of Sciences of the United States of America *106*, 13820–13825.

Dave, N., Guaita-Esteruelas, S., Gutarra, S., Frias, À., Beltran, M., Peiró, S., and de Herreros, A. G. (2011). Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition. The Journal of biological chemistry *286*, 12024–12032.

Donnelly, M. L., Luke, G., Mehrotra, A., Li, X., Hughes, L. E., Gani, D., and Ryan, M. D. (2001). Analysis of the aphthovirus 2A/2B polyprotein "cleavage" mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal "skip". The Journal of general virology *82*, 1013–1025.

Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., Berx, G., Cano, A., Beug, H., and Foisner, R. (2005). DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. Oncogene *24*, 2375–2385.

Elenbaas, B., Spirio, L., Koerner, F., Fleming, M. D., Zimonjic, D. B., Donaher, J. L., Popescu, N. C., Hahn, W. C., and Weinberg, R. A. (2001). Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. Genes & development *15*, 50–65.

Fogg, V. C., Liu, C.-J., and Margolis, B. (2005). Multiple regions of Crumbs3 are required for tight junction formation in MCF10A cells. Journal of cell science *118*, 2859–2869.

Gal, A., Sjöblom, T., Fedorova, L., Imreh, S., Beug, H., and Moustakas, A. (2008). Sustained TGF beta exposure suppresses Smad and non-Smad signalling in mammary epithelial cells, leading to EMT and inhibition of growth arrest and apoptosis. Oncogene *27*, 1218–1230.

Gemmill, R. M., Roche, J., Potiron, V. A., Nasarre, P., Mitas, M., Coldren, C. D., Helfrich, B. A., Garrett-Mayer, E., Bunn, P. A., and Drabkin, H. A. (2011). ZEB1-responsive genes in non-small cell lung cancer. Cancer letters *300*, 66–78.

Gossen, M. (1992). Tight Control of Gene Expression in Mammalian Cells by Tetracycline-Responsive Promoters. Proceedings of the National Academy of Sciences *89*, 5547–5551.

Gupta, P. B., Onder, T. T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R. A., and Lander, E. S. (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell *138*, 645–659.

Hartwell, K. A., Muir, B., Reinhardt, F., Carpenter, A. E., Sgroi, D. C., and Weinberg, R. A. (2006). The Spemann organizer gene, Goosecoid, promotes tumor metastasis. Proc Natl Acad Sci U S A *103*, 18969–18974.

Howe, E. N., Cochrane, D. R., and Richer, J. K. (2011). Targets of miR-200c mediate suppression of cell motility and anoikis resistance. Breast Cancer Res *13*, R45.

Inman, G. J. (2002). SB-431542 Is a Potent and Specific Inhibitor of Transforming Growth Factor-beta Superfamily Type I Activin Receptor-Like Kinase (ALK) Receptors ALK4, ALK5, and ALK7. Molecular Pharmacology *62*, 65–74.

Lacher, M. D., Shiina, M., Chang, P., Keller, D., Tiirikainen, M. I., and Korn, W. M. (2011). ZEB1 limits adenoviral infectability by transcriptionally repressing the coxsackie virus and adenovirus receptor. Molecular cancer *10*, 91.

Lacher, M. D., Tiirikainen, M. I., Saunier, E. F., Christian, C., Anders, M., Oft, M., Balmain, A., Akhurst, R. J., and Korn, W. M. (2006). Transforming growth factor-beta receptor inhibition enhances adenoviral infectability of carcinoma cells via up-regulation of Coxsackie and Adenovirus Receptor in conjunction with reversal of epithelial-mesenchymal transition. Cancer research *66*, 1648–1657.

Lemmers, C., Michel, D., Lane-Guermonprez, L., Delgrossi, M.-H., Médina, E., Arsanto, J.-P., and Le Bivic, A. (2004). CRB3 binds directly to Par6 and regulates the morphogenesis of the tight junctions in mammalian epithelial cells. Molecular biology of the cell *15*, 1324–1333.

Littlewood, T. D., Hancock, D. C., Danielian, P. S., Parker, M. G., and Evan, G. I. (1995). A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. Nucleic Acids Res *23*, 1686–1690.

Lois, C., Hong, E. J., Pease, S., Brown, E. J., and Baltimore, D. (2002). Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. Science (New York, N.Y.) *295*, 868–872.

Loots, G. G., Ovcharenko, I., Pachter, L., Dubchak, I., and Rubin, E. M. (2002). rVista for comparative sequence-based discovery of functional transcription factor binding sites. Genome research *12*, 832–839.

Mackay, D. R., Hu, M., Li, B., Rhéaume, C., and Dai, X. (2006). The mouse Ovol2 gene is required for cranial neural tube development. Developmental biology *291*, 38–52.

Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J., and Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715.

Mordasky Markell, L., Pérez-Lorenzo, R., Masiuk, K. E., Kennett, M. J., and Glick, A. B. (2010). Use of a TGFbeta type I receptor inhibitor in mouse skin carcinogenesis reveals a dual role for TGFbeta signaling in tumor promotion and progression. Carcinogenesis *31*, 2127–2135.

Onder, T. T., Gupta, P. B., Mani, S. A., Yang, J., Lander, E. S., and Weinberg, R. A. (2008). Loss of E-cadherin promotes metastasis via multiple downstream transcriptional pathways. Cancer Res *68*, 3645–3654.

Pyrgaki, C., Liu, A., and Niswander, L. (2011). Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. Developmental biology *353*, 38–49.

Roh, M. H., Fan, S., Liu, C.-J., and Margolis, B. (2003). The Crumbs3-Pals1 complex participates in the establishment of polarity in mammalian epithelial cells. Journal of cell science *116*, 2895–2906.

Ryan, M. D., and Drew, J. (1994). Foot-and-mouth disease virus 2A oligopeptide mediated cleavage of an artificial polyprotein. The EMBO journal *13*, 928–933.

Shapiro, I. M., Cheng, A. W., Flytzanis, N. C., Balsamo, M., Condeelis, J. S., Oktay, M. H., Burge, C. B., and Gertler, F. B. (2011). An EMT-Driven Alternative Splicing Program Occurs in Human Breast Cancer and Modulates Cellular Phenotype. PLoS genetics *7*, e1002218.

Shu, X., Shaner, N. C., Yarbrough, C. A., Tsien, R. Y., and Remington, S. J. (2006). Novel chromophores and buried charges control color in mFruits. Biochemistry *45*, 9639–9647.

Whipple, R. A., Matrone, M. A., Cho, E. H., Balzer, E. M., Vitolo, M. I., Yoon, J. R., Ioffe, O. B., Tuttle, K. C., Yang, J., and Martin, S. S. (2010). Epithelial-to-mesenchymal transition promotes tubulin detyrosination and microtentacles that enhance endothelial engagement. Cancer research *70*, 8127–8137.

Yang, J., Mani, S. A., Donaher, J. L., Ramaswamy, S., Itzykson, R. A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R. A. (2004). Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. Cell *117*, 927–939.

Chapter 4:

Conclusions and Future Directions

4.1 Significance and utility of the EMT core signature

In Section 2.2.1, I demonstrated that, using one parental cell line and five different methods of inducing the EMT, it was possible to find a large microarray signature (624 probes) that behaved commonly across the five different treatments. Each treatment induced many changes in the cells, and not every change was necessarily related to the EMT. Moreover, the inducing factors were all expressed by constitutively active constructs, and these cells were stably maintained over a long period in the mesenchymal state, so that secondary effects of the overexpressed factor would be evident, including secondary effects of changes unrelated to the EMT. Hence, many of the changes observed following induction of the EMT might well be these secondarily-induced genes that are not integral components of the EMT program *per se*.

Comparing any single mesenchymal line to its epithelial control, one is confronted with thousands of gene expression changes with no straightforward method to distinguish those related directly to the EMT program and those that are not. Simply shrinking the list to a manageable size by picking the probes with the highest fold-change or lowest p-value would mean discarding a large amount of useful information without fundamentally improving the quality of the analysis, since, as detailed above, many of the genes that are significantly changed may not be integral to the EMT program. However, a second mesenchymal line, induced to undergo EMT by a different factor but starting from the same baseline state, would exhibit its own spectrum of gene expression changes unrelated to EMT but would share some changes with the first line, assuming the EMT process in

the two cell populations used the same regulatory mechanisms. Extending this further, each additional, independently-induced mesenchymal line would further refine the signature.

This was the basis for the design of the microarray experiment described in Section 2.2.1, and its success was contingent on the assumption that all five lines, one way or another, ended up using a common mechanism to enter into the mesenchymal state. Accordingly, the presence of even a single line that did not use the same mechanism would drastically reduce the size of the overlapping gene signature. In fact, this was not observed, providing assurance that the five mesenchymal lines arrived at similar phenotypes using components of a common EMT program.

Beyond the mathematical treatment described in Section 2.2.1 showing that over 98% of the probes in the EMT core signature were not due to chance, it is difficult to quantify the significance of the size of the signature. Taube et al. used the same dataset and different criteria to obtain a signature of 246 genes by employing a fold-change cutoff (Taube et al., 2010). They were more concerned with obtaining a signature that could be correlated with clinical breast cancer subtypes and therefore had different requirements. My aim was to discern mechanistic details from my signature using bioinformatics methods, which favored a larger, less conservative signature than a smaller, higher-confidence one. In this way, I would not prematurely exclude genes from downstream analyses.

At the beginning of this study, HMLE cells were believed to exist in the epithelial state, and the assumption was that the five treatments were actively converting cells from the epithelial to the mesenchymal state. The heterogeneity of the HMLE cells, in particular the significance of the minor subpopulation of pre-existing mesenchymal cells, was not appreciated until much later

(Mani et al., 2008). The observation that the mesenchymal cells were generally more resistant to killing by various means (Creighton et al., 2009; Gupta et al., 2009), combined with this pre-existing mesenchymal population in HMLE cells, gave rise to the formal possibility that any treatment that stresses the cells could, over time, give rise to a large proportion of mesenchymal cells purely by means of selective outgrowth, rather than such a treatment having any role in actively converting individual epithelial cells into mesenchymal ones.

All previous protein overexpression constructs used in the lab, including the ones used to generate the samples for the EMT core signature (Goosecoid, Snail, Twist and TGF-β1), were based on an MMLV retrovirus vector, which by all accounts had a low infectivity. This meant that, in obtaining the mesenchymal cells, a large proportion of the cells were killed off during the antibiotic selection process. In this particular overexpression system, there was no way to distinguish cells killed by the antibiotic from those killed by the overexpressed gene of interest. The lentiviral expression construct I described in Section 3.2.1 gave me the ability to select infected cells by fluorescence-activated cell sorting (FACS), without resorting to antibiotic selection and the accompanying cell death. This novel property allowed me to observe the anomalous cell death of cell-sorted Twist-infected cells in the presence of puromycin, when all the cells should in principle already be puromycin-resistant. This suggested that Twist overexpression combined with puromycin exposure may lead to a state of stress that could be killing cells, representing a selection pressure that could lead to the preferential outgrowth of mesenchymal cells in the manner described earlier.

In Section 3.2.4, I showed how the absence or presence of puromycin-induced cell death correlated with the speed of apparent EMT in the Twist-infected cells. Snail-infected cells treated the same way did not experience the anomalous cell death and did not behave differently in the absence or presence of puromycin. Combined with the earlier observations of pre-existing mesenchymal cells in the HMLE parental pool and the hardiness of mesenchymal cells compared to their epithelial counterparts, this suggested that selective outgrowth, rather than the direct conversion of epithelial cells, may have played a role in how the original Twist-infected mesenchymal cells were obtained.

The possibility of selective outgrowth of pre-existing mesenchymal cells following activation of an EMT-inducing signal has an impact on the utility of the EMT core signature. There is a distinction between the regulatory program required to actively convert an epithelial cell into a mesenchymal one, and the regulatory program required to maintain a mesenchymal cell in its present state. If even one of the mesenchymal lines was obtained largely by a process of selective outgrowth, then the signature may only encompass the set of genes that describes the latter program rather than the former. In the case of an overexpressed factor actively inducing an EMT, we can be fairly confident that the signals responsible for the EMT are still expressed at the time the samples were collected, since the factor is constitutively expressed and presumably still acting on its downstream targets. This would not be the case if the overexpressed factor was merely selecting for the outgrowth of pre-existing mesenchymal cells, because it is not known if the pre-existing cells are expressing the program for active conversion as opposed to maintenance. The genes in the two programs are not mutually exclusive, and in this light, the identification of ZEB1 as a

candidate was fortuitous, because it both promotes the EMT and maintains the mesenchymal state. Going forward, it is safer to assume that candidates found within the signature have a role in maintaining the mesenchymal state, as opposed to an active involvement in the transition.

4.2 Relationship between EMT core signature and published gene signatures

The four previously published cancer gene signatures used for comparison to the EMT core signature in Section 2.2.2 represented different aspects of metastatic development (O'Donnell et al., 2005; Provenzani et al., 2006; Schuetz et al., 2006; Jaeger et al., 2007). The signature of O'Donnel et al. describes genes that are different between primary tumors that give rise to metastases and primary tumors that do not, using unmatched samples. Provenzani et al. generated a signature that describes the difference between an already-invasive tumor and a matched metastasis from the same individual, although both samples had undergone the process to convert them into cell lines. Schuetz et al. used matched samples of non-invasive DCIS and invasive ductal carcinoma to generate their signature, exploiting the heterogeneity observed even within single primary tumors. Finally, Jaeger et al. directly compared unmatched melanoma metastases to primary tumors, with no reference to whether the primary tumors were potentially invasive or not. The fact that the EMT core signature showed a high overlap with the latter three signatures suggests that parts of the EMT process may be involved in converting localized tumor regions into invasive ones, and also show up as differences between primary tumors and their derived metastases. The weak overlap with the first signature may indicate the difficulty in searching for EMT by looking at bulk, unfractionated primary tumors, lending credence to the idea

that EMT in clinical cases is a localized phenomenon occurring only in some parts of the tumor rather than across the whole tumor.

Previously, attempts to create "metastasis signatures" or "prognosis signatures" used bulk tumor samples (Ramaswamy et al., 2003; van t Veer et al., 2002; Wang et al., 2005) to generate their RNA for microarray analysis. Bulk tumors contain various proportions of non-neoplastic stromal cells, which may include fibroblasts and myofibroblasts, which are mesenchymal in nature. Poor prognosis samples may contain higher amounts of stroma, and hence give a higher mesenchymal gene signature even if all the cancer cells themselves remained epithelial. For diagnostic and prognostic purposes, this may not matter, but this issue means that gene signatures generated from bulk tumor samples cannot easily be used to dissect EMT mechanisms. It is perhaps no coincidence that the three signatures showing significant overlap are derived from either laser-capture microdissected samples or cell line samples, all of which avoid the issue of stromal contamination.

The high signature overlap found in comparison with the melanoma metastasis signature of Jaeger et al. came as a surprise. The signature was one of the earliest generated from samples extracted using laser-capture microdissection. Even though melanoma is not a carcinoma, I was drawn to the possibility that this may represent a very clean signature of metastases versus primary tumors. Previous work in the lab showed that the transcription factor Slug was essential in a melanoma metastasis model (Gupta et al., 2005), and more recently, Slug has been implicated in transcriptionally activating ZEB1 to produce an EMT-like phenotype in melanoma cells (Wels et al., 2011). Melanocytes can trace their developmental lineage back to neural crest cells (Dupin and Le

Douarin, 2003), meaning they have undergone at least one EMT during the development of their cell lineage. Perhaps it is not so surprising that transformed melanocytes reuse some of the EMT program to acquire the ability to disseminate once again and form metastases. My analysis of the overlapping genes reinforces the idea that ZEB1 may play a key role in the metastasis of melanoma, in relation to EMT-like changes.

Even though the overlap between the EMT core signature and the breast carcinoma signature of Schuetz et al. was not pursued in this study, the high percentage signal of that overlap warrants further investigation. The possible role of LEF1 in promoting the invasive phenotype in breast cancer is supported by work in lung adenocarcinoma showing that LEF1 is a mediator of chemotactic invasion and metastasis (Nguyen et al., 2009), as well as correlative evidence that LEF1 overexpression is a prognostic factor for poor survival and increased liver metastasis in colorectal cancer (Lin et al., 2011).

The EMT core signature was optimized by me for research purposes, specifically for understanding the mechanism of the EMT, and will probably be of limited utility as a clinical diagnostic tool. A useful clinical diagnostic microarray gene signature would be one that could be used to recommend specific treatment options or inform about the risk of further progression. Schuetz et al. used matched samples of DCIS and IDC within the same tumor to generate their signature, but such a procedure to distinguish the heterogeneity within individual tumors would not be common practice in the clinic for diagnostic purposes. Since the EMT is likely to appear only in a subset of primary tumor cells, a microarray gene signature for the EMT would be difficult to detect, even if the stromal cells were excluded. Nevertheless, the success at finding meaningful overlaps

between the EMT core signature and the published signatures described above suggests that processes involved in an EMT may be at work in a broad range of cancer progression processes. As newer data based on laser-capture microdissected tumor samples become publicly available, there is potential for discovering yet more links between the EMT and cancer progression.

4.3 ZEB1 is qualitatively different from other EMT-inducing factors

The first indication that ZEB1 may be different from the other EMT-inducing factors was the fact that ZEB1 itself was represented as an EMT core signature gene, while the others were not. In a simplified model, ZEB1 would be placed downstream of the other factors, one or several of which would induce ZEB1 in addition to a number of other EMT-associated genes. The ability of *Drosophila* Twist and Snail to bind to the promoter of zfh1, the fly ortholog of ZEB1, lends further support to this idea. More recently, others have found Snail and Twist to impinge on the regulation of ZEB1 in NMuMG cells (Dave et al., 2011), with Twist binding directly to the ZEB1 promoter.

However, the actual mechanisms of these gene interactions may not be so straightforward. In Section 3.2.6, I demonstrated that ZEB1 overexpression resulted in a rapid EMT that is stably maintained, once experimental artifacts are accounted for. Twist and Snail overexpression can also produce mesenchymal HMLE cells, but on a far longer timescale. If Twist or Snail can directly upregulate ZEB1, why would there be a major difference in the schedule of EMT induction? The upregulation of ZEB1 could already be detected in day 5 post-infection with the Snail or Twist lentivirus, but the appearance of mesenchymal morphology, particularly of the HMLE-Twist cells,

happened more than five days after that, long after the equivalent response was observed in the HMLE-ZEB1 cells. The difference might be due to ZEB1 levels induced by transcription of its endogenous locus versus its expression using a ubiquitin or doxycycline-inducible promoter. Alternatively, the increased levels of ZEB1 in the HMLE-Snail and HMLE-Twist cells could be the result of an increased proportion of pre-existing mesenchymal cells which already have higher levels of ZEB1, rather than increased levels of ZEB1 induced in individual cells; the data that were available to me could not distinguish between the two. Regardless, ZEB1 was a demonstrably better EMT-inducing factor than was Twist or Snail, at least in the HMLE cells studied here.

ZEB1 is not only sufficient to induce the EMT in HMLE cells, it is also necessary for the maintenance of the resulting mesenchymal state. This dual property has not been demonstrated with the other factors in HMLE cells. By manipulating the levels of a single transcription factor (ZEB1), I was able to move cells from the epithelial state to the mesenchymal state and back again. This had recently also been demonstrated in the H358 non-small cell lung cancer cell line using a doxycycline-inducible ZEB1 system (Thomson et al., 2011). Intriguingly, their results did not show a self-sustaining upregulation of endogenous ZEB1; once doxycycline was removed, their cells spontaneously reverted to the epithelial phenotype. The ability of ZEB1 to stably upregulate its own expression, which I described in Section 3.2.7, may therefore not be a universal phenomenon. Since the other study did not attempt to distinguish endogenous and exogenous ZEB1 mRNA transcripts, it was not known if endogenous ZEB1 was upregulated at all by the exogenous ZEB1.

Dave et al. have shown that Snail and Twist can influence ZEB1 expression, but they did not show whether Snail or Twist require ZEB1 to carry out the EMT. I demonstrated that in the presence of

overexpressed Snail, the knockdown of ZEB1 caused a partial MET, in particular increasing the proportion of cells that express E-cadherin. Despite evidence that Snail binds to the E-cadherin promoter (Batlle et al., 2000) and has a repressive effect (Cano et al., 2000), overexpressed levels of Snail in HMLE cells still relied on endogenous levels of ZEB1 to repress E-cadherin. Without a complete knockdown or knockout of ZEB1, the hypothesis that ZEB1 is necessary for the EMT triggered by other factors could not be properly tested. With the recent developments in TALE nuclease technology, it is now possible to efficiently edit the genome of somatic cells, including cell lines (Miller et al., 2010; Hockemeyer et al., 2011). For future studies, a useful reagent to create using this technology would be HMLE cells with ZEB1 knocked out, or ZEB1 modified for CRE-mediated knockout. In this way, one can be certain of the total absence of ZEB1, and properly test the ability of Snail or Twist or other factors to induce EMT and sustain the mesenchymal state without the involvement of ZEB1.

4.4 ZEB1 can override miR-200c

Even though miR-200 is known to target ZEB1, there have not been any reports demonstrating the pre-eminent role of one factor or another in controlling EMT status. In the clinic, this question may be moot, since the ZEB1 and miR-200 are never observed to be upregulated in the same cells and at the same time. But while ZEB1 is known to directly repress several epithelial genes (Eger et al., 2005; Aigner et al., 2007a, 2007b), the role of miR-200 in the EMT beyond its ability to influence ZEB1 expression is unknown. In Section 3.2.8, I demonstrated that elevated ZEB1 was able to override high miR-200c levels to induce an EMT. This disproved the formal possibility that

miR-200c is the main actor in EMT induction, and that an important role of ZEB1 is to repress miR-200c in order to reverse the RNAi effects of miR-200c on downstream EMT mediators.

Fibronectin was previously demonstrated to be a direct target of miR-200c (Howe et al., 2011), and one would assume that ZEB1 affects fibronectin levels mainly through its action on miR-200c. However, a perturbation of miR-200c is known to cause an inverse change in ZEB1 levels (Park et al., 2008; Burk et al., 2008), so the possibility that fibronectin is also regulated by another pathway downstream of ZEB1 could not be ruled out in that study. Indeed, with the double perturbation experiment in Section 3.2.8, I show that ZEB1 could upregulate fibronectin despite unchanging levels of miR-200c. This observation pointed to the existence of a miR-200c-independent pathway for ZEB1 to regulate fibronectin expression. In the same experiment, I found that ZEB1 could upregulate its own transcription through a miR-200c-independent mechanism, adding a further layer of complexity to the ZEB1-miR-200c control loop.

The relative importance of ZEB1 over miR-200c with respect to EMT, combined with the possible downstream position of ZEB1 in relation to other EMT-inducing factors, means that the constituency of genes that are eventually found to lie downstream of ZEB1 should provide a rich source of genes involved in the EMT core regulatory circuit.

4.5 ZEB1 downstream targets

In fact, the microarray expression profiling of ZIC cells described in Section 3.2.11, combined with rVISTA promoter analysis, provided a list of potential direct downstream target genes of ZEB1. E-cadherin, the best candidate for a positive control because of the extensive knowledge

accumulated regarding its regulation by ZEB1 (Eger et al., 2005), failed to be identified as a target, calling into question the sensitivity of my microarray experiment. On the other hand, the list of genes included Crumb3 (CRB3) and Coxsackie virus and adenovirus receptor (CXADR), epithelial genes previously identified to be direct targets of ZEB1 (Aigner et al., 2007a; Lacher et al., 2011). Epithelial splicing regulatory protein 1 (ESRP1), also on the list, was previously associated with the EMT (Shapiro et al., 2011) and identified as a ZEB1-responsive gene (Gemmill et al., 2011). The presence of a conserved ZEB1 binding site on the promoter of ESRP1, as indicated by rVISTA analysis, supports the notion that ESRP1 is also a direct target of ZEB1.

Besides the above genes that had previous associations with ZEB1 or EMT, I identified two other genes on the list with no previous association that may nevertheless be of interest. Ovo-like 2 (OVOL2) and Grainyhead-like 2 (GRHL2) are two transcription factors whose mutant phenotypes have been characterized in mice. Both displayed limbs/digits/tail and craniofacial abnormalities, similar to Twist knockout mice. GRHL2 appears to be involved in the maintenance of epithelial genes, since a loss-of-function mutation of GRHL2 results in the loss of expression of several epithelial genes, including E-cadherin (Pyrgaki et al., 2011). Conversely, a gain-of-function mutation of GRHL2 causes embryonic lethality and failure to close the neural tube, echoing phenotypes associated with EMT failure (Brouns et al., 2011). OVOL2 is less well-characterized, but loss-of-function mutants have defective neural crest cells that fail to migrate (Mackay et al., 2006). OVOL2 and GRHL2 made compelling targets for further characterization. Since both appear to be repressed by ZEB1, it would be interesting to study the effects on EMT of overexpressing ZEB1 and either OVOL2 or GRHL2 at the same time. After I had identified GRHL2 as

a potential candidate, it was reported that GRHL2 could indeed suppress the EMT if overexpressed (Cieply et al., 2012), providing support for this approach to identifying candidates.

4.6 Final Perspective

In this study, I sought to identify elements of a core regulatory circuit for the Epithelial-Mesenchymal Transition. In the HMLE cell model that I used, ZEB1 appeared time and again to act as a central mediator for the EMT, being the conduit through which other known EMT-inducing factors flow. The ZEB1-inducible single-cell clones provide a tractable system for exploring further downstream of ZEB1 in the EMT regulatory circuit. The dependency relationship found in this study between ZEB1 and Snail are somewhat surprising, considering that ZEB1-null mice can develop almost to term (Miyoshi et al., 2006), whereas Snail-nulls die at E8.5 with phenotypes widely indicative of defective mesoderm formation and EMT failure (Carver et al., 2001). ZEB1 is apparently not crucial to developmental EMT, yet played a critical role in my cells. The apparent discrepancy could be explained by the presence of related family members that may play equivalent roles but are only available in specific tissue or developmental contexts.

The transcription factor Slug, belonging to the same family as Snail, is necessary for the formation of mesoderm and neural crest migration in chick embryos (Nieto et al., 1994), but Slug-null mouse embryos can develop to term (Jiang et al., 1998), unlike their Snail-null counterparts, which die early. And whereas ZEB1-null mice can survive until birth, mice null for the related family member ZEB2 arrest at E8.5 with failure in neural tube closure and neural crest migration (Van de Putte et al., 2003; Miyoshi et al., 2006). I speculate that ZEB1 and ZEB2 may perform equivalent but non-

redundant functions in mediating EMT and occupy equivalent positions in the EMT core circuitry, downstream of Snail or Slug which similarly act in an equivalent but non-redundant manner, depending on the situation. While ZEB1 alone may not be indispensable for EMT in every situation, ZEB1 and ZEB2 collectively could serve as the gatekeepers for determining EMT status.

This modular concept of the EMT regulatory circuit could be extended further, whereby at every step, one or another of a related family of genes plays the role in transducing the signal to undergo an EMT. The choice of which member is active at each step would vary from one EMT program to another, but the structure of the different EMT programs would be similar. This has implications in diagnosis and treatments targeting the EMT, as an EMT signature that only contains single members of each class is unlikely to be broadly applicable in different types of tissue. However, thanks to gene homology, a detailed characterization of the EMT program in one system, for example the HMLE cells, could be used to rapidly identify similar programs in other systems by searching not just for the specific gene identified, but its closest homologs as well.

4.7 References

Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., Mikulits, W., Brabletz, T., Strand, D., Obrist, P., Sommergruber, W., Schweifer, N., Wernitznig, A., Beug, H., Foisner, R., and Eger, A. (2007a). The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. Oncogene.

Aigner, K., Descovich, L., Mikula, M., Sultan, A., Dampier, B., Bonne, S., van Roy, F., Mikulits, W., Schreiber, M., Brabletz, T., Sommergruber, W., Schweifer, N., Wernitznig, A., Beug, H., Foisner, R., and Eger, A. (2007b). The transcription factor ZEB1 (deltaEF1) represses Plakophilin 3 during human cancer progression. FEBS Lett *581*, 1617–1624.

Batlle, E., Sancho, E., Francí, C., Domínguez, D., Monfar, M., Baulida, J., and García De Herreros, A. (2000). The transcription factor snail is a repressor of E-cadherin gene expression in epithelial tumour cells. Nature cell biology *2*, 84–89.

Brouns, M. R., De Castro, S. C. P., Terwindt-Rouwenhorst, E. A., Massa, V., Hekking, J. W., Hirst, C. S., Savery, D., Munts, C., Partridge, D., Lamers, W., Köhler, E., van Straaten, H. W., Copp, A. J., and Greene, N. D. E. (2011). Over-expression of Grhl2 causes spina bifida in the Axial defects mutant mouse. Human molecular genetics *20*, 1536–1546.

Burk, U., Schubert, J., Wellner, U., Schmalhofer, O., Vincan, E., Spaderna, S., and Brabletz, T. (2008). A reciprocal repression between ZEB1 and members of the miR-200 family promotes EMT and invasion in cancer cells. EMBO Rep *9*, 582–589.

Cano, A., Perez-Moreno, M. A., Rodrigo, I., Locascio, A., Blanco, M. J., del Barrio, M. G., Portillo, F., and Nieto, M. A. (2000). The transcription factor snail controls epithelial-mesenchymal transitions by repressing E-cadherin expression. Nat Cell Biol *2*, 76–83.

Carver, E. A., Jiang, R., Lan, Y., Oram, K. F., and Gridley, T. (2001). The mouse snail gene encodes a key regulator of the epithelial-mesenchymal transition. Molecular and cellular biology *21*, 8184–8188.

Cieply, B., Riley, P., Pifer, P. M., Widmeyer, J., Addison, J. B., Ivanov, A. V., Denvir, J., and Frisch, S. M. (2012). Suppression of the Epithelial-Mesenchymal Transition by Granyhead-like 2. Cancer research.

Creighton, C. J., Li, X., Landis, M., Dixon, J. M., Neumeister, V. M., Sjolund, A., Rimm, D. L., Wong, H., Rodriguez, A., Herschkowitz, J. I., Fan, C., Zhang, X., He, X., Pavlick, A., Gutierrez, M. C., Renshaw, L., Larionov, A. A., Faratian, D., Hilsenbeck, S. G., Perou, C. M., Lewis, M. T., Rosen, J. M., and Chang, J. C. (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proceedings of the National Academy of Sciences of the United States of America *106*, 13820–13825.

Dave, N., Guaita-Esteruelas, S., Gutarra, S., Frias, À., Beltran, M., Peiró, S., and de Herreros, A. G. (2011). Functional cooperation between Snail1 and twist in the regulation of ZEB1 expression during epithelial to mesenchymal transition. The Journal of biological chemistry *286*, 12024–12032.

Dupin, E., and Le Douarin, N. M. (2003). Development of melanocyte precursors from the vertebrate neural crest. Oncogene *22*, 3016–3023.

Eger, A., Aigner, K., Sonderegger, S., Dampier, B., Oehler, S., Schreiber, M., Berx, G., Cano, A., Beug, H., and Foisner, R. (2005). DeltaEF1 is a transcriptional repressor of E-cadherin and regulates epithelial plasticity in breast cancer cells. Oncogene *24*, 2375–2385.

Gemmill, R. M., Roche, J., Potiron, V. A., Nasarre, P., Mitas, M., Coldren, C. D., Helfrich, B. A., Garrett-Mayer, E., Bunn, P. A., and Drabkin, H. A. (2011). ZEB1-responsive genes in non-small cell lung cancer. Cancer letters *300*, 66–78.

Gupta, P. B., Kuperwasser, C., Brunet, J.-P. P., Ramaswamy, S., Kuo, W.-L. L., Gray, J. W., Naber, S. P., and Weinberg, R. A. (2005). The melanocyte differentiation program predisposes to metastasis after neoplastic transformation. Nat Genet *37*, 1047–1054.

Gupta, P. B., Onder, T. T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R. A., and Lander, E. S. (2009). Identification of selective inhibitors of cancer stem cells by high-throughput screening. Cell *138*, 645–659.

Hockemeyer, D., Wang, H., Kiani, S., Lai, C. S., Gao, Q., Cassady, J. P., Cost, G. J., Zhang, L., Santiago, Y., Miller, J. C., Zeitler, B., Cherone, J. M., Meng, X., Hinkley, S. J., Rebar, E. J., Gregory, P. D., Urnov, F. D., and Jaenisch, R. (2011). Genetic engineering of human pluripotent cells using TALE nucleases. Nature biotechnology *29*, 731–734.

Howe, E. N., Cochrane, D. R., and Richer, J. K. (2011). Targets of miR-200c mediate suppression of cell motility and anoikis resistance. Breast Cancer Res *13*, R45.

Jaeger, J., Koczan, D., Thiesen, H. J., Ibrahim, S. M., Gross, G., Spang, R., and Kunz, M. (2007). Gene expression signatures for tumor progression, tumor subtype, and tumor thickness in laser-microdissected melanoma tissues. Clin Cancer Res *13*, 806–815.

Jiang, R., Lan, Y., Norton, C. R., Sundberg, J. P., and Gridley, T. (1998). The Slug gene is not essential for mesoderm or neural crest development in mice. Developmental biology *198*, 277–285.

Lacher, M. D., Shiina, M., Chang, P., Keller, D., Tiirikainen, M. I., and Korn, W. M. (2011). ZEB1 limits adenoviral infectability by transcriptionally repressing the coxsackie virus and adenovirus receptor. Molecular cancer *10*, 91.

Lin, A. Y., Chua, M.-S., Choi, Y.-L., Yeh, W., Kim, Y. H., Azzi, R., Adams, G. A., Sainani, K., van de Rijn, M., So, S. K., and Pollack, J. R. (2011). Comparative profiling of primary colorectal carcinomas and liver metastases identifies LEF1 as a prognostic biomarker. PloS one *6*, e16636.

Mackay, D. R., Hu, M., Li, B., Rhéaume, C., and Dai, X. (2006). The mouse Ovol2 gene is required for cranial neural tube development. Developmental biology *291*, 38–52.

Mani, S. A., Guo, W., Liao, M. J., Eaton, E. N., Ayyanan, A., Zhou, A. Y., Brooks, M., Reinhard, F., Zhang, C. C., Shipitsin, M., Campbell, L. L., Polyak, K., Brisken, C., Yang, J., and Weinberg, R. A. (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704–715.

Miller, J. C., Tan, S., Qiao, G., Barlow, K. A., Wang, J., Xia, D. F., Meng, X., Paschon, D. E., Leung, E., Hinkley, S. J., Dulay, G. P., Hua, K. L., Ankoudinova, I., Cost, G. J., Urnov, F. D., Zhang, H. S., Holmes, M. C., Zhang, L., Gregory, P. D., and Rebar, E. J. (2010). A TALE nuclease architecture for efficient genome editing. Nature Biotechnology *29*, 143–148.

Miyoshi, T., Maruhashi, M., Van De Putte, T., Kondoh, H., Huylebroeck, D., and Higashi, Y. (2006). Complementary expression pattern of Zfhx1 genes Sip1 and deltaEF1 in the mouse embryo and their genetic interaction revealed by compound mutants. Developmental dynamics : an official publication of the American Association of Anatomists *235*, 1941–1952.

Nguyen, D. X., Chiang, A. C., Zhang, X. H.-F., Kim, J. Y., Kris, M. G., Ladanyi, M., Gerald, W. L., and Massagué, J. (2009). WNT/TCF signaling through LEF1 and HOXB9 mediates lung adenocarcinoma metastasis. Cell *138*, 51–62.

Nieto, M. A., Sargent, M. G., Wilkinson, D. G., and Cooke, J. (1994). Control of cell behavior during vertebrate development by Slug, a zinc finger gene. Science (New York, N.Y.) *264*, 835–839.

O'Donnell, R. K., Kupferman, M., Wei, S. J., Singhal, S., Weber, R., O'Malley, B., Cheng, Y., Putt, M., Feldman, M., Ziober, B., Muschel, R. J., O'Donnell, R. K., and O'Malley, B. (2005). Gene expression signature predicts lymphatic metastasis in squamous cell carcinoma of the oral cavity. Oncogene *24*, 1244–1251.

Park, S.-M., Gaur, A. B., Lengyel, E., and Peter, M. E. (2008). The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. Genes & development *22*, 894–907.

Provenzani, A., Fronza, R., Loreni, F., Pascale, A., Amadio, M., and Quattrone, A. (2006). Global alterations in mRNA polysomal recruitment in a cell model of colorectal cancer progression to metastasis. Carcinogenesis *27*, 1323–1333.

Van de Putte, T., Maruhashi, M., Francis, A., Nelles, L., Kondoh, H., Huylebroeck, D., and Higashi, Y. (2003). Mice lacking ZFHX1B, the gene that codes for Smad-interacting protein-1, reveal a role for multiple neural crest cell defects in the etiology of Hirschsprung disease-mental retardation syndrome. American journal of human genetics *72*, 465–470.

Pyrgaki, C., Liu, A., and Niswander, L. (2011). Grainyhead-like 2 regulates neural tube closure and adhesion molecule expression during neural fold fusion. Developmental biology *353*, 38–49.

Ramaswamy, S., Ross, K. N., Lander, E. S., and Golub, T. R. (2003). A molecular signature of metastasis in primary solid tumors. Nat Genet *33*, 49–54.

Schuetz, C. S., Bonin, M., Clare, S. E., Nieselt, K., Sotlar, K., Walter, M., Fehm, T., Solomayer, E., Riess, O., Wallwiener, D., Kurek, R., and Neubauer, H. J. (2006). Progression-specific genes identified by expression profiling of matched ductal carcinomas in situ and invasive breast tumors, combining laser capture microdissection and oligonucleotide microarray analysis. Cancer Res *66*, 5278–5286.

Shapiro, I. M., Cheng, A. W., Flytzanis, N. C., Balsamo, M., Condeelis, J. S., Oktay, M. H., Burge, C. B., and Gertler, F. B. (2011). An EMT-Driven Alternative Splicing Program Occurs in Human Breast Cancer and Modulates Cellular Phenotype. PLoS genetics *7*, e1002218.

Taube, J. H., Herschkowitz, J. I., Komurov, K., Zhou, A. Y., Gupta, S., Yang, J., Hartwell, K., Onder, T. T., Gupta, P. B., Evans, K. W., Hollier, B. G., Ram, P. T., Lander, E. S., Rosen, J. M., Weinberg, R. A., and Mani, S. A. (2010). Core epithelial-to-mesenchymal transition interactome gene-expression signature is associated with claudin-low and metaplastic breast cancer subtypes. Proc Natl Acad Sci U S A *107*, 15449–15454.

Thomson, S., Petti, F., Sujka-Kwok, I., Mercado, P., Bean, J., Monaghan, M., Seymour, S. L., Argast, G. M., Epstein, D. M., and Haley, J. D. (2011). A systems view of epithelial-mesenchymal transition signaling states. Clinical & experimental metastasis *28*, 137–155.

van t Veer, L. J., Dai, H., van de Vijver, M. J., He, Y. D., Hart, A. A., Mao, M., Peterse, H. L., van der Kooy, K., Marton, M. J., Witteveen, A. T., Schreiber, G. J., Kerkhoven, R. M., Roberts, C., Linsley, P. S., Bernards, R., and Friend, S. H. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530–536.

Wang, Y., Klijn, J. G., Zhang, Y., Sieuwerts, A. M., Look, M. P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M. E., Yu, J., Jatkoe, T., Berns, E. M., Atkins, D., and Foekens, J. A. (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671–679.

Wels, C., Joshi, S., Koefinger, P., Bergler, H., and Schaider, H. (2011). Transcriptional Activation of ZEB1 by Slug Leads to Cooperative Regulation of the Epithelial-Mesenchymal Transition-Like Phenotype in Melanoma. The Journal of investigative dermatology *131*, 1877–1885.