Title
An Ovol2-Zeb1 EMT-Regulatory Circuit Governs Mammary Basal-Luminal Binary Differentiation

Permalink
https://escholarship.org/uc/item/5b8693wq

Author
Villarreal Ponce, Alvaro P.

Publication Date
2017

Peer reviewed|Thesis/dissertation
UNIVERSITY OF CALIFORNIA,
IRVINE

An Ovol2-Zeb1 EMT-Regulatory Circuit Governs Mammary Basal-Luminal Binary Differentiation

DISSERTATION

submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in Biomedical Sciences

by

Alvaro Patricio Villarreal-Ponce

Dissertation Committee:
Professor Xing Dai, Chair
Professor Bogi Anderson
Professor Peter Donovan
Professor Klemens Hertel
Professor Kyoko Yokomori

2017
DEDICATION

To

my family and friends
# TABLE OF CONTENTS

| LIST OF FIGURES                           | iv  |
| LIST OF TABLES                           | vi  |
| LIST OF ABBREVIATIONS                    | vii |
| ACKNOWLEDGEMENTS                         | viii|
| CURRICULUM VITAE                         | xi  |
| ABSTRACT OF DISSERTATION                 | xiv |
| CHAPTER 1: Introduction                  | 1   |
| CHAPTER 2: Materials and Methods         | 30  |
| CHAPTER 3: Genome-Wide Assessment of Ovol2 Binding Targets Reveals its Function as a Gatekeeper of Epithelial Identity | 47  |
| CHAPTER 4: An Ovol2-Zeb1 Regulatory Circuit Governs Mammary Basal-Luminal Binary Differentiation | 65  |
| CHAPTER 5: Discussion and Future Perspectives | 100 |
| APPENDIX A Uncovering the Cellular Heterogeneity Existing Within the Basal Compartment of the Adult Mouse Mammary Gland | 103 |
| REFERENCES                               |     |
## LIST OF FIGURES

<p>| FIGURE 1.1 | Introductory figure 1 | 22 |
| FIGURE 1.2 | Introductory figure 2 | 24 |
| FIGURE 1.3 | Introductory figure 3 | 25 |
| FIGURE 1.4 | Introductory figure 4 | 26 |
| FIGURE 1.5 | Introductory figure 5 | 27 |
| FIGURE 3.1 | The Ovol2 isoform with identified transcriptional repressor function is the exclusive isoform produced in HC11 mammary epithelial cells. | 58 |
| FIGURE 3.2 | Genome-wide analysis identifies Ovol2 to function as a transcriptional repressor that binds to the promoter-proximal regions with high fidelity | 59 |
| FIGURE 3.3 | Genome-wide analysis of Ovol2 binding identifies a myriad of EMT-related to be direct binding targets of Ovol2 | 60 |
| FIGURE 3.4 | Ovol2 directly represses multiple downstream target genes involved with EMT and stemness in keratinocytes | 63 |
| FIGURE 4.1 | Zeb1 is expressed in the basal compartment of the mammary epithelium and its expression is inversely correlated with expression of MET-TF Ovol2 | 85 |
| FIGURE 4.2 | Zeb1 is expressed in the basal compartment of the mammary epithelium and its protein is readily detected during times of active mammary morphogenesis | 87 |
| FIGURE 4.3 | Zeb1 depletion in unsorted mammary epithelial cells results in an imbalance of basal and luminal/alveolar cell-derived primary organoid colonies | 89 |
| FIGURE 4.4 | Zeb1 depletion in basal cells influences regenerative ductal branching | 91 |
| FIGURE 4.5 | Depletion of MET-TF Ovol2 results in biased fate determination toward basal fate in mammary epithelial cells | 93 |</p>
<table>
<thead>
<tr>
<th>FIGURE</th>
<th>Description</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>FIGURE 4.6</td>
<td>Depletion of Zeb1 in sorted basal cells results in reduced clonogenicity and altered gene expression</td>
<td>94</td>
</tr>
<tr>
<td>FIGURE S4.1</td>
<td>Supplemental to Figure 4.1</td>
<td>95</td>
</tr>
<tr>
<td>FIGURE S4.2</td>
<td>Supplemental to Figure 4.2</td>
<td>96</td>
</tr>
<tr>
<td>FIGURE S4.3</td>
<td>Supplemental to Figure 4.3</td>
<td>97</td>
</tr>
<tr>
<td>FIGURE S4.4</td>
<td>Supplemental to Figure 4.4</td>
<td>98</td>
</tr>
<tr>
<td>FIGURE S4.5</td>
<td>Supplemental to Figure 4.5</td>
<td>99</td>
</tr>
</tbody>
</table>
# LIST OF TABLES

<table>
<thead>
<tr>
<th>TABLE</th>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>TABLE 1.1</td>
<td>Introductory table 1</td>
<td>29</td>
</tr>
<tr>
<td>TABLE 3.1</td>
<td>GO analysis reveals top candidate processes associated with Ovol2-bound genes</td>
<td>62</td>
</tr>
</tbody>
</table>
# LIST OF ABBREVIATIONS

<table>
<thead>
<tr>
<th>ABBREVIATION</th>
<th>DEFINITION</th>
</tr>
</thead>
<tbody>
<tr>
<td>EMT</td>
<td>Epithelial-to-Mesenchymal Transition</td>
</tr>
<tr>
<td>SC</td>
<td>Stem Cell</td>
</tr>
<tr>
<td>TF</td>
<td>Transcription Factor</td>
</tr>
<tr>
<td>miRNA</td>
<td>microRNA</td>
</tr>
<tr>
<td>MG</td>
<td>Mammary Gland</td>
</tr>
<tr>
<td>GSEA</td>
<td>Gene Set Enrichment Analysis</td>
</tr>
<tr>
<td>FACS</td>
<td>Fluorescence-Activated Cell Sorting</td>
</tr>
<tr>
<td>H3K4me3</td>
<td>Histone 3 Lysine residue 4, tri-methylation</td>
</tr>
<tr>
<td>H3K27me3</td>
<td>Histone 3 Lysine residue 27, tri-methylation</td>
</tr>
<tr>
<td>DPC</td>
<td>Days-Post Coitus</td>
</tr>
<tr>
<td>ChIP</td>
<td>Chromatin Immunoprecipitation</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase Chain Reaction</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger Ribonucleic Acid</td>
</tr>
<tr>
<td>shRNA</td>
<td>Short Hairpin Ribonucleic Acid</td>
</tr>
<tr>
<td>3D</td>
<td>Three-Dimensional</td>
</tr>
<tr>
<td>qRT-PCR</td>
<td>Quantitative Real-Time PCR</td>
</tr>
<tr>
<td>GFP</td>
<td>Green Fluorescence Protein</td>
</tr>
<tr>
<td>WT</td>
<td>Wild-Type</td>
</tr>
<tr>
<td>SSKO</td>
<td>Skin and Mammary Epithelial-Specific Knockout</td>
</tr>
<tr>
<td>GO</td>
<td>Gene Ontology</td>
</tr>
<tr>
<td>MET</td>
<td>Mesenchymal-to-Epithelial Transition</td>
</tr>
<tr>
<td>MECs</td>
<td>Mammary Epithelial Cells</td>
</tr>
<tr>
<td>E (number)</td>
<td>Embryonic Day (number)</td>
</tr>
<tr>
<td>MaSCs</td>
<td>Mammary Stem Cells</td>
</tr>
<tr>
<td>TEB</td>
<td>Terminal End Bud</td>
</tr>
<tr>
<td>ECM</td>
<td>Extra-Cellular Matrix</td>
</tr>
<tr>
<td>MMP</td>
<td>Matrix Metallo-Protease</td>
</tr>
<tr>
<td>RNA-seq</td>
<td>RNA (High-throughput) Sequencing</td>
</tr>
<tr>
<td>ChIP-seq</td>
<td>ChIP (High-throughput) Sequencing</td>
</tr>
<tr>
<td>MOI</td>
<td>Multiplicity of Infection</td>
</tr>
<tr>
<td>MACS</td>
<td>Model-Based Analysis for ChIP-seq</td>
</tr>
<tr>
<td>CTC</td>
<td>Circulating Tumor Cell</td>
</tr>
<tr>
<td>iMMEC</td>
<td>Immortalized Mammary Epithelial Cell</td>
</tr>
<tr>
<td>FDR</td>
<td>False Discovery Rate</td>
</tr>
<tr>
<td>cDNA</td>
<td>Complementary Deoxyribonucleic Acid</td>
</tr>
<tr>
<td>FPKM</td>
<td>Fragments per Kilo-base &amp; Million Mapped Reads</td>
</tr>
<tr>
<td>t-SNE</td>
<td>t-Distributed Stochastic Neighbor Embedding</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal Component Analysis</td>
</tr>
</tbody>
</table>
ACKNOWLEDGMENTS

Completion of a doctorate degree is a significant academic achievement that is not only guided by one’s own scientific curiosity, but also through a dedicated community that fosters mentorship and incessant support. Throughout my graduate career, I have been fortunate to have been surrounded by many loving and inspiring individuals who have encouraged and stimulated my growth tremendously in various ways. The work I describe through this dissertation represents the fruit of my labor, which was impassioned and influenced by generous individuals whom I am eternally grateful for.

Primarily, I would like to express my greatest appreciation to my thesis mentor, Dr. Xing Dai, for her invaluable guidance and commitment to training me as a scientist. Her wisdom and contagious enthusiasm for scientific inquiry invigorated my passion for science and inspired me to engage myself scientifically to develop the traits of an independent critical thinker and successful investigator. She also challenged me to defend my ideas and taught me the importance of strategy. I thank you for your encouragement, support, and unfathomable kindness, which has been central to enriching my academic experience.

I am also very grateful to the rest of my thesis committee members, Dr. Bogi Andersen, Dr. Peter Donovan, Dr. Klemens Hertel, and Dr. Kyoko Yokomori, for their constructive criticism, suggestions, and incessant devotion to helping me thrive throughout my graduate career. The advice they provided during our meetings stimulated me intellectually and allowed me to approach my scientific questions in novel ways.

I would also like to extend an additional heartfelt thanks to Dr. Peter Donovan, program director of the California Institute of Regenerative Medicine (CIRM) at UC Irvine, for believing in my potential to succeed scientifically and for his advocacy to fund my research throughout a
significant portion of my graduate career. Similarly, I would like to thank Dr. Luis Mota-Bravo and Dr. Marlene De La Cruz, directors of the NIH Minority Biomedical Research Support (MBRS) Program, for their financial support during my first year as a graduate student. I stand by your vision in preparing underrepresented minorities with training in biomedical research, and look forward to contributing to your vision in a similar manner in the future.

I am also indebted to both past and current members of the Dai laboratory for creating an auspicious environment that inculcated scientific growth and many friendships. My most sincere gratitude goes to Dr. Kazuhide Watanabe for his many constructive conversations and suggestions. His endless knowledge and amazing work ethic inspired me to strive further and to overcome any obstacles I may encounter throughout my journey as a graduate student. I treasure the insight he gave me on various scientific topics and even the introduction on rock-climbing. I wish him the best of luck on his future endeavors as a scientist at the RIKEN Center. I also want to express my appreciation to Dr. Briana Lee-Altman, who helped me adjust to my new surrounding upon joining the Dai Laboratory by being a great friend and colleague. I am honored to have gotten the opportunity to contribute to the work of these two scientists early during my graduate career.

Within the current Dai laboratory, I have been fortunate to have crossed paths with a great group of colleagues. I thank Daniel Haensel for the wonderfully stimulating conversations we had discussing our scientific curiosity and for the interest he showed when I shared my work. He has a brilliant mind and I am excited to see his continuous growth as a scientist. I thank Dr. Yingying Han for her assistance in investigating how perturbations to the Zeb1-Ovol2 regulatory circuit influence basal-luminal differentiation with the mammary epithelium. Working with her during these months has been an amazing experience. Also, I thank Yuan Zhou and Dr. Peng
Sun for their help in managing my mouse colonies and for the mentorship they provided. Both of their contributions have been indispensable to my scientific training and success as a researcher.

Outside of the laboratory setting, my family and friends have been a vital source of encouragement and support, which continue to inspire me and bring light into my life. I value every moment we’ve had to share, learn, and grow together. I express my warmest gratitude to my parents, Martha Villarreal and Alvaro Villarreal, Sr., for their ever-lasting love and for their dedication in creating an auspicious environment for my siblings and me to grow in. Your teachings and hard work instilled in me a strong passion for learning and to strive continuously to become a better individual. I thank my siblings, Pavel Villarreal and Ivana Villarreal, for their support, advice, and motivation. They are both incredible role models to me that inspire me to persist and continue pushing forward.

Additionally, I give special thanks to Christian Guerrero-Juarez and Ghaidaa Kashgari, two amazing life-long friends that have played a pivotal role during my time at UC Irvine. We have experienced so many valuable memories together and feel we have grown so much from them as well. I treasure the engaging conversations and laughter we shared over the years. I am excited for the brilliant future that is ahead for both of them.

Finally, I would like to thank my girlfriend, Celeste Lanuza, for her immense love, care, and encouragement. She is an inspiring individual with a kind heart that motivates me to bring out my best self always. I treasure our endless experiences that initiated as graduate students at UC Irvine that have allowed us to flourish together.
CURRICULUM VITAE

Alvaro P. Villarreal-Ponce, PhD candidate
Department of Biological Chemistry
William J. Cheney Hall - Medical Sciences 1, Building D Wing
University of California, Irvine
Irvine, CA 92697-4716, USA
Laboratory: +1 (949) 824-7409 | Mobile: +1 (714) 423-4487 | E-mail: apvillar@uci.edu

PERSONAL INFORMATION
Full Name: Villarreal-Ponce, Alvaro Patricio
Place of Birth: Mexico City, Mexico | Date of Birth: March 14, 1985
Citizenship(s): United Mexican States (Native), United States of America (Naturalized)
Languages(s): Castilian Spanish (Native), English (Fluent)

OBJECTIVE
Highly motivated cross-disciplinary biomedical scientist, currently studying molecular and epigenetic mechanisms that regulate cellular differentiation in mammalian epithelial regeneration. My graduate experience involves training in a combination of areas, as they pertain to mammary epithelial biology.

RESEARCH INTEREST(S)
Epithelial stem cells; Mammary Gland; Epithelial Development; Epithelial Regeneration; Epithelial-to-Mesenchymal Transition (EMT); Cancer Stem Cells; Skin, Wound Healing

TECHNIQUES
Mouse Genetics, Mouse Surgery, Tissue Culture, Bioinformatics, Viral Work, Fluorescence and Confocal Microscopy, ChIPseq/ChIP-PCR, scRNAseq/RNAseq, PCR/qPCR, Flow Cytometry/FACS, Immunocytochemistry/ Immunofluorescence staining, UCSC Genome Browser, Broad Institute GSEA

EDUCATION AND TRAINING
University of California, Irvine Doctor of Philosophy, Medical Sciences 06/2011-05/2017
California State University, Fullerton Bachelor of Science, Cellular and Molecular Biology 08/2006-06/2009

GRANTS AND RESEARCH FELLOWSHIPS
California Institute of Regenerative Medicine pre-doctoral trainee Grant No. TG2-01152
University of California, Irvine 05/2013-12/2015
National Institute of Health MBRS-IMSD Fellowship Program Grant No. R25GM055246 08/2012-05/2013
University of California, Irvine

National Science Foundation LSAMP Research Training Program 01/2010-12-2010
Grant No. HRD-1302873
California State University, Fullerton

Grant No. T37MD001485
California State University, Fullerton
Instituto Leloir, Buenos Aires, Argentina

National Institute of Health Bridges Fellowship Program 05/2006-08/2006
Grant No. GM-55246
Fullerton College, Fullerton
University of California, Irvine

AWARDS AND HONORABLE MENTIONS
UC Irvine School of Medicine Travel Award 05/2016
University of California, Irvine

University Honors – Dean’s List 08/2006-05/2007
California State University, Fullerton

PRESENTATIONS, SEMINARS, AND POSTER SESSIONS


TEACHING EXPERIENCE
Advanced Human Genetics 01/2012-03/2012
Graduate Teaching Assistant
University of California, Irvine

STEM Subjects 01/2010-06/2012
Academic Tutor
Nicholas Academic Center, Santa Ana, CA

Biochemistry / Organic Chemistry 09/2010-12/2010
Lecturer
California State University, Fullerton

General Chemistry Lab
Lecturer
California State University, Fullerton

Biochemistry
Teaching Assistant
California State University, Fullerton

STEM Subjects
Undergraduate Academic Tutor
California State University, Fullerton

PUBLICATIONS


URL to full list of published work:
The capacity of epithelial cells to acquire enhanced lineage plasticity could depend on their ability to undergo EMT. Investigations performed on cultured epithelial cells support a link between EMT and bestowment of stem cell (SC)-like properties, raising the possibility that regulators of EMT may be responsible for producing an intermediate cellular identity between epithelial and mesenchymal states and is compatible with SC potential. The goal of my thesis project is to identify and characterize key transcriptional regulators of the dynamic EMT process that facilitate the production and maintenance of epithelial SCs, using the MG as a model system. The Dai laboratory identified Ovol2 as a TF that is required for mammary and epidermal development. My work contributed to the discovery of Ovol2 as a master negative regulator of EMT that directly represses the expression of various EMT-related genes, the most important being Zeb1, a critical mediator of Ovol2 loss-of-function effects. Zeb1 is a potent EMT-TF implicated in conferring SC-like traits to differentiated cells in mammary epithelial tumors. However, its in vivo role within normal mammary epithelia has not been studied. I found that Zeb1 also directly represses Ovol2, leading to the identification of an Ovol2-Zeb1 cross-
repression circuit, which is shown by mathematic modeling to support intermediate cellular states between terminal epithelial and mesenchymal identities. Additionally, my data shows that Zeb1 expression is activated during early pregnancy in the basal cells of the mammary epithelium, which are known to gain multipotency upon pregnancy or transplantation. Using in vivo and ex vivo approaches to determine how perturbations to the Ovol2-Zeb1 circuit regulate stemness, I found this circuit to be important in modulating mammary SC basal-luminal differentiation. In addition to protecting basal cells from precocious differentiation toward a luminal fate, Zeb1 functions in regulating SC self-renewal/proliferative activity. Both mechanisms may contribute to the observed, Zeb1 loss-induced defect in ductal branching during mammary regeneration. My findings uncover a previously unknown role of Zeb1 and its associated molecular circuit in regulating mammary SC activity and basal/luminal differentiation, offering new insights into how epithelial plasticity contributes to stemness and identify novel transcriptional regulators of epithelial SCs.
INTRODUCTION

1.1. THE MAMMARY GLAND

THE MAMMARY GLAND AS A SYSTEM TO STUDY EPITHELIAL STEM CELL REGULATION

The mammary gland (MG) is a unique mammalian organ whose development and tissue organization occur primarily after birth. Changes to female hormones levels during the stages of puberty and pregnancy induce growth and dynamic architectural remodeling to a simple bilayered epithelium that produces an intricate arboreal-like network of ducts and alveoli that are responsible for providing a source of nourishment to newborn offspring. Anatomically, the bilayered mammary epithelium is composed of an inner, luminal population of cells that are capable of terminally differentiating into milk-producing cells, and a unique, mesenchymal-like outer epithelial population of basal/myoepithelial cells, that produce contractile forces to eject the milk (Watson & Khaled, 2008). Importantly, the basal epithelial population is also the resident site of a subset population of multipotent mammary stem cells (SCs or MaSCs) that are responsible for manifesting critical events that pertain to mammary epithelial development, homeostasis, and regeneration (Prater et al., 2014; Wang et al., 2014; Zeng et al., 2010). Cells from this population are also proposed to be the tumorigenic origin of a subset of breast cancers (Shackleton et al., 2006; Van Amerongen et al., 2012).

Substantial progress has been made in resolving the cellular composition of each population of the MG, and has enabled investigators to utilize this system to address questions pertinent to epithelial SC regulation. Combinatorial labeling of these cells with lineage-specific (CD45, CD31, TER119) and epithelial surface markers permits the isolation and profiling of both luminal cells (Lin\textsuperscript{−}CD29\textsuperscript{low}CD24\textsuperscript{+} or Lin\textsuperscript{−}CD49\textsuperscript{low}CD24\textsuperscript{+} or Lin\textsuperscript{−}CD49\textsuperscript{high}EpCAM\textsuperscript{low}) and
SC-containing basal population (LinCD29\textsuperscript{high}CD24\textsuperscript{+} or LinCD49f\textsuperscript{high}CD24\textsuperscript{+} or Lin\textsuperscript{-}CD49\textsuperscript{high}EpCAM\textsuperscript{low}) through Fluorescence-Activated Cell Sorting (FACS) (Prater et al., 2014; Visvader et al., 2014). Moreover, lineage tracing and fat pad transplantation assays are functional tools that are instrumental to allowing for the investigation of genes/factors that are responsible for regulating SC potential at various stages of development and regeneration (De Ome et al., 1959; Shackleton et al., 2006; Stingl et al., 2006; Van Keymeulen et al., 2011; Rios et al., 2014). Studies that implement these tools stipulate that cells of the basal population display distinct lineage potential under transplantation and physiological conditions (Van Keymeulen et al., 2011; Rios et al., 2014), underscoring the intrinsic lineage plasticity of mammary basal cells and the influence environmental factors have on SC fate decisions. Integration of these techniques can offer insight into SC function and the mechanisms that regulate cell fate specification to properly produce this glandular organ.

OVERVIEW OF MAMMARY GLAND DEVELOPMENT

The MG is a glandular organ present in mammalian organisms with a function in providing newborn offspring with a source of nutrients and anti-microbial protection (Peaker, 2002; Watson and Khaled, 2008). The mature complex secretory organ constitutes an extensive, functional network of epithelial ducts and milk-producing alveolar structures that develop through distinct developmental stages taking place predominantly after birth. Structurally, the mammary epithelium is composed of a cellular bi-layer that is maintained by two major mammary lineages that are derived from a single embryonic cellular origin (FIGURE 1.1) (Watson et al., 2008). The hollow lumen of the ductal architecture is lined with an inner luminal population of epithelial cells that have ability to onset terminal differentiation programs to
become milk-producing cells (Visvader et al., 2014). Luminal epithelial cells display traits that are characteristic of true epithelial cells, including maintenance of apicobasal polar orientation, and expression of genes that encode various cell adhesion proteins that help form a network of tightly inter-connected cells. The outer layer of the mammary epithelium serves a more supportive role, and is composed of a unique population of basal/myoepithelial cells possessing an epithelioid identity between epithelial and mesenchymal states (Prater et al., 2014), which are responsible for producing the contractile forces that eject the nutrient-rich milk during lactation (Raymond et al., 2011). Importantly, within the basal population, lineage tracing and transplantation studies reveal the existence of a multi/bipotent mammary epithelial stem cell (MaSC) population that can differentiate to produce cells of both basal and luminal epithelial lineages (DeOme et al., 1959; Prater et al., 2014; Shackleton et al., 2006; Van Keymeulen et al., 2011; Wang et al., 2015). While the differentiation hierarchical map of MaSCs remains largely presumptive and speculative (FIGURE 1.2), advances have been made to identify unique cytokeratin and surface marker combinations for each population that have enabled us to gather information on the cellular and molecular properties of each cell lineage.

In mice, development of the MG initiates independently of hormonal stimulation shortly after mid-gestation (embryonic day E10.5), with the formation of ectoderm-derived milk lines that extend ventrally from anterior to posterior limbs (Henninghausen et al., 2001; Hens and Wysolmerski, 2005). Five pairs of disc-shaped placodes then form at specific sites along these mammary lines, and undergo morphogenic alterations to produce bulbous structures that sink into the dermis and begin developing into the primitive epithelial bud (Robinson, 2007). By E15.5, the nascent epithelial bud reaches the mesenchymal precursor of the mammary fat pad, and invaginates within this tissue to form a rudimentary ductal network by E18.5 (Hens and
After birth, systemic hormonal cues couple with mammotrophic factors to promote vast, dynamic architectural changes to the MG that can be categorized into three differentially regulated developmental stages: active morphogenesis (pubertal and reproductive), homeostasis (mature adult), and regression (involution) (FIGURE 1.3) (Visvader et al., 2014; Macias et al., 2013; Henninghausen et al., 2001). This work will focus on stages pertinent to active morphogenesis and regeneration of the mammary epithelium.

At puberty, the primary mammary anlage reinstates its developmental program by proceeding with ductal morphogenesis, a biological program that promotes proliferation, ductal elongation and branching, to produce an intricate network of epithelial ducts that penetrate and fill the surrounding stromal fat pad (Sternlicht, 2005). This program initiates in response to the release of estrogens at adolescence (approximately 3 weeks of age in mice) and reaches completion at approximately ten to twelve weeks of age (adulthood) (Sternlicht et al., 2005). Central to pubertal morphogenesis are the terminal end buds (TEBs), bulbous structures at the ends of the emergent epithelial ducts that are responsible for the dramatic expansion of the cellular constituents of the mammary epithelium (Hinck and Silberstein, 2005). Basal cap cells of the TEB are indispensable for proper ductal elongation, in part through the enriched abundance of mammary SCs that catalyze this proliferative program (Visvader et al., 2014), but also by supporting the cohesive migration of the epithelium through their partial mesenchymal phenotype that is engendered by partial onset of Epithelial-to-Mesenchymal Transition (EMT) (REFER TO SECTION 1.2 EPITHELIAL-TO-MESENCHYMAL TRANSITION). Partial EMT regulates inter-cellular adhesion and polarity (Nanba et al., 2001; Hinck and Silberstein, 2005; Kouros-Mehr and Werb, 2006; Ewald et al., 2012) and regulates changes to the surrounding extracellular matrix (ECM), specifically through up-regulation of genes encoding
matrix metallo-proteinases (MMPs) and matrix glycosaminoglycan-degrading glycolytic enzymes, that create space for these cells to migrate (Hinck and Silberstein, 2005).

Interestingly, in vitro findings associate transient expression of EMT regulators in producing a cellular identity that is characterized by enhanced epithelial plasticity and a SC-trait (REFER TO SECTION 1.3 PARTIAL EMT AND STEMNESS). While enriched expression of EMT markers is consistently documented in several postulated SC populations (Wang et al., 2015; Guo et al., 2012), the physiological relevance of this phenomenon is to be determined. Moreover, very little is known of the regulatory mechanisms that control the extent of EMT within these cells. Elucidating how EMT is regulated has major implications to cancer research, as aberrant activation of EMT-inducing transcription factors (EMT-TFs) is seen to be a major contributor of tumorigenesis (Smalley et al., 2003, Mani et al, 2008; Wang et al., 2015).

Independent of the influence TEB biology has on pubertal morphogenesis, ductal elongation and epithelial branching will contribute to mammary morphogenesis by expanding the surface area of the mammary epithelium to maximize the number of milk-producing alveolar structures that will be generated upon induction of pregnancy. Much like the dynamic signaling that is involved in ductal elongation, hormonal and local controls regulate branching morphogenesis (Sternlicht et al., 2006), as does the stromal components of the surrounding mammary fat pad (Gjorevski et al., 2011). In addition, basal and luminal epithelial populations have been found to communicate via paracrine signaling to regulate basal cell proliferation, which in turn has a direct effect on governing the extent of mammary branching (Macias et al., 2011). Specifically, regulation of mammary branching is mediated by paracrine Slit/Robo signaling between luminal and basal cells, where activation of this pathway inhibits basal cell...
proliferation by negatively regulating expression of canonical Wnt signaling targets and increasing the cytoplasmic pool of β-catenin (Macias et al., 2011).

Upon reaching adulthood, SC activity will cease and enter a stage of quiescence that ends at the reproductive stage (Macias et al., 2013). Upon onset of pregnancy, the MG will again undergo vast dynamic changes in preparation for lactation. During pregnancy, changes in hormone levels (progesterone and prolactin) drive adult MaSC expansion to support secondary and tertiary branching (Joshi et al., 2010; Asselin-Labat et al., 2010), as well as the initiation of alveologenesis-a program where proliferating luminal epithelial cells generate alveolar buds that differentiate into alveoli to generate milk-producing lobules (Macias et al., 2013). These findings suggest that adult mammary basal cells can be activated to expand and adopt multiple plastic states when properly stimulated with proper environmental stimuli (Prater et al., 2014). The molecular mechanisms that control SC activation and its underlying plasticity are yet to be elucidated and are the focus of this study.

MAMMARY STEM CELLS

Although the molecular mechanisms that regulate epithelial plasticity and the epithelial SC state remain poorly understood, significant advancements have been made toward identifying the properties of mammary SCs. Since the pioneering regenerative experiments describing multi/bi-potent MaSCs with the capacity to produce a mammary tree consisting of both mammary lineages (DeOme et al., 1959), monumental studies have uncovered the physiological existence of both multi/bi-potent mammary epithelial SCs (MaSCs) and long-lived unipotent SC/progenitors in basal compartment of the MG (Visvader et al., 2014; Rios et al., 2014; Van Keymeulen et al., 2011). Despite the controversy surrounding which population is pertinent to
postnatal morphogenesis, both SC populations are likely to exist, with environmental conditions and developmental stages largely governing which cell type to be most active.

Several morphologically distinct basal sub-populations have been associated with capturing the mammary SC phenotype (Visvader et al., 2014). Expression of genes that have been found to be functionally relevant in the identification of SC populations in other tissues have been integrated in the search of the SC population within the MG, but with only marginal success. Among these genes, expression of Ship was initially identified to reside in embryonic and hematopoietic SCs, but not within differentiated cell types (Tu et al., 2001). Within the MG, expression of Ship is localized to cap cells of TEBs and basal alveolar bud cells, two populations exhibiting enriched SC activity (Visvader et al., 2014), and Ship+ basal cells also exhibited enrich for active SCs in the pubertal and pregnant gland over Ship− basal cells (Bai et al., 2010). However, basal cells that are negative for Ship expression also exhibit capacity to regenerate a mammary tree, albeit at a lower frequency. Additionally, Wnt targets Lgr5 and Axin2 have received significant attention as potential SC markers through their ability to denote SC populations in other epithelial tissues. Indeed, Lgr5+ and Axin2+ cells have been identified in the MG and shown to have SC potential and are considered self-renewal factors for mammary stem cells (Barker et al., 2013; Zeng et al., 2010). Recently, Wnt target protein C receptor (Procr) was identified as a cell surface receptor that labels a unique basal population of cells that display characteristics of EMT and exhibit characteristics of multipotent adult mammary SCs (Wang et al., 2015) (REFER TO SECTION 1.3 PARTIAL EMT AND STEMNESS). Within the regenerative realm, these cells have high regenerative capacity and can differentiate into both basal and luminal lineages of the mammary epithelium (Wang et al., 2015). This said, existing
evidence also suggests that the MaSC compartment is heterogeneous, and additional stem cell subpopulations are yet to be identified.

Evidence from the aforementioned studies, describe the difficulty in truly separating the SC population from the differentiated myoepithelial and basal progenitor subsets. Surprisingly, a finding from Prater et al., 2014, revealed that terminally differentiated myoepithelial cells in the basal compartment can interconvert to give rise to cells exhibiting capacity to repopulate the mammary gland upon transplantation (MaSC state) (Prater et al., 2014). Such finding suggests that the SC state is not a fixed cellular entity, but rather is a plastic one that responds to environmental or pathological stimuli. Not much is known about the molecular mechanisms that underlie this remarkable lineage plasticity. As briefly described above, regulatory mechanisms that control EMT provide interesting and viable gene candidates. While some data in the literature seem to support this notion (Guo et al., 2012), whether and how an EMT-regulatory network governs the intrinsic plasticity of MaSCs has yet to be elucidated. Moreover, the physiological significance of EMT control in normal MG homeostasis (as opposed to transplantation) remains unknown (REFER TO SECTION 1.3 PARTIAL EMT AND STEMNESS).

1.2. EPITHELIAL-TO-MESENCHYMAL TRANSITION (EMT)

INTRODUCTION

Induction of pluripotency in differentiated cell types highlights the lineage plasticity of somatic cells. While this form of cellular plasticity offers a multitude of opportunities for regenerative medicine, it raises questions as to how plasticity is restricted during the dynamic processes of tissue morphogenesis, homeostasis, and regeneration. EMT, and its reverse process,
mesenchymal-to-epithelial transition (MET), represent cellular trans-differentiation programs that highlight the intrinsic and remarkable plasticity of epithelial cells. EMT is defined as a naturally occurring biological process that is driven by growth factors or signaling proteins such as TGF-β, Wnt/beta-catenin, and sonic hedgehog (Shh) (Nieto et al., 2013), which instructs cells of epithelial identity to carry out biochemical and morphological changes that promote shedding of traits associated with differentiated epithelial cells and acquisition of a more mesenchymal phenotype (Kalluri et al., 2009). Principal cellular features affected by progression of EMT include cell junctions, apical-basal polarity, reorganization of cytoskeletal filaments, and ECM deposition, all of which confer the cells with migratory properties that are necessary in various biological contexts (FIGURE 1.4) (Nieto et al., 2016).

The functional context by which EMT occurs and the extent of its reversibility are predominating characteristics that classify different types of EMT. While EMT events taking place during early embryogenesis are regarded largely as complete and transformative, new evidence suggests greater flexibility in this transitional program to occur in the various contexts; most notably in those associated with malignancy and most recently stemness (Nieto et al., 2016). Our expanding knowledge now appreciates the existence of a spectrum of intermediary phases that correspond to partial-EMT states that play important roles in normal and neoplastic settings, including active epithelial morphogenesis, regenerative wound healing, fibrosis, and cancer metastasis (Kalluri et al., 2009; Nieto et al., 2016; Smalley et al., 2003) (TABLE 1.1). The primary differences between complete and partial-EMT rely largely on the extent of biochemical changes resulting in dissolution of the cohesive sheets of epithelial cells to produce a collection of motile and independent cells. Cells bearing a partial-EMT state co-express both transcriptional activators and repressors of EMT, which act antagonistically to maintain a hybrid,
“metastable” phenotype that exhibits flexibility to induce or reverse the process (Nieto et al., 2016;).

Complete progression of EMT is thoroughly documented to be active during early stages of embryogenesis. Its relevance in producing the primary mesenchyme has been demonstrated through studies on gastrulation using various species (Alberga et al., 1991; Thiery et al., 2006). During gastrulation, the epithelial cells at the primitive streak undergo EMT to produce the mesoderm; one of the three primordial germ layers that gives rise to the mesenchyme, mesothelium, and non-epithelial blood cells in developing organisms. Genetic studies uncover EMT-TF Snail (Snai1) as a transcriptional repressor with a critical role in orchestrating progression of EMT in cells of the mouse primitive streak. Snai1-deficient mutant embryos lose viability, exhibiting defects during gastrulation that result from morphological malformations of the mesoderm (Carver et al., 2001). Detailed analysis of these mutants reveals that loss of Snai1 leads to failure to down-regulate expression of epithelial adhesion gene Cdh1 (E-cadherin), which prohibits cells of the emerging mesoderm to lose epithelial identity and engender the desired mesenchymal state.

Unlike the complete trans-differentiation of epithelial cells to a mesenchymal identity described above, cutaneous wound healing and ductal elongation during MG morphogenesis represent a transient event of cellular reprogramming that depends on a reversible partial progression of EMT. During cutaneous wound healing, epithelial cells at the wound edge engender a partial and reversible EMT to move into the damaged region and begin restoring the cutaneous tissue (Nieto et al., 2016; Savagner et al., 2008). These cells then revert to an epithelial identity to reconstitute epithelial integrity and its protective barrier function. A partial-EMT in the wounded epithelium is found to be dependent by up-regulation of EMT-TF Slug.
That is dependent upon EGFR signaling, as loss of this transcriptional regulator results in failure to properly heal (Arnoux et al., 2008; Hudson et al., 2013). Similar findings have been attributed to EMT-TFs Snail (Snai1) and Twist (Twist1) (Terao et al., 2011).

Simultaneously, aberrant activation of the EMT program is closely associated with cancer progression, both through promoting dissociation of carcinoma cells from the primary tumor and into the bloodstream (extravasation), and through mis-regulated bestowment of SC-like traits that engender a CSC state (see below) (Thiery et al., 2009; Nieto et al., 2016). Heterotypic signals from the malignant tissue functionally activate expression of EMT-TFs Snail, Slug, Twist1, or Zeb1, that function to disrupt adhesion junctions between cells and the ECM to engender a more invasive and metastatic cellular phenotype that is believed to play a role in dissemination of the primary tumor (Kalluri et al., 2009). Pleiotropic activity of EMT-TFs also bestows these cells with enhanced resistance to apoptosis and resilience to drug therapy, thus highlighting the importance in identifying the diverse functional roles of these EMT-TFs in the normal setting and how these pathways become mis-regulated in cancer progression (Settleman et al., 2010).

1.3. PARTIAL EMT AND STEMNESS

Existing evidence suggests that a SC state is not necessarily a fixed cellular entity, but rather a transient state that responds to external stimuli and is in equilibrium with non-SCs. Bi-directional transition between these two states has recently been demonstrated utilizing cultured epithelial cells, which uncover an interesting link between a transient, partial onset of EMT and production of SC-like properties. These findings also underscore the capacity of EMT to enhance the self-renewal and lineage plasticity of epithelial cells (Mani et al. 2008, Morel et al. 2008). In vitro cell culture studies have made a tantalizing finding, where both normal and neoplastic
differentiated epithelial cells that progress through a partial EMT acquire SC-like traits, such as multipotency and capacity to self-renew (Chaffer et al., 2013; Guo et al., 2012; Morel et al., 2012). In particular, studies on cultured human breast epithelial cells reveal that transitions between SC and non-SC states can be induced through transient transcriptional activation of potent EMT-TF, ZEB1, by the conversion its promoter from a poised/bivalent epigenetic configuration to a transcriptionally active one (Chaffer et al., 2013). As will be further described below, the finding suggests the importance in tightly regulating the expression of ZEB1 in restricting lineage plasticity and is a transcriptional regulator with a central role in the questions posed in this work.

1.4. TRANSCRIPTIONAL REGULATION OF EMT

A complex network of signaling pathways and downstream effectors work together to ensure proper execution of the EMT. Multi-functional cytokine, transforming growth factor beta (TGF-β), and related members from the TGF-β superfamily of signaling proteins are best recognized for their ability to initiate a signaling cascade that induces progression of EMT (Xu et al., 2009). Simultaneously however, many growth factors and morphogens from Wnt, Notch, and Hedgehog signaling pathways can also participate in activating the EMT program (Lamouille et al., 2014). Extracellular signals that induce progression of EMT stimulate expression of a collection of pleiotropic EMT-TFs that directly regulate expression of genes characteristic of epithelial and mesenchymal cells. Extensive studies have identified many EMT-TFs and key inhibitors that protect epithelial identity and inhibit aberrant progression of this program (TABLE 1.2). Master transcriptional drivers of EMT include Zeb1, Zeb2, Snail, Slug, and Twist1, which govern the expression of effectors responsible for mediating changes to the
cytoskeleton as well as cellular machinery involved in adhesion and migration. Hallmark gene expression changes associated with progression of EMT include down-regulation of epithelial adhesion molecule E-cadherin (Cdh1) and up-regulation of mesenchymal surface marker N-cadherin (Cdh2) and vimentin (Vim) (Lamouille et al., 2014; Kalluri et al., 2009)

In the following section, transcriptional regulators of EMT belonging the Ovol and Zeb families will be described as their important roles in regulating this process are relevant to this study.

TRANSCRIPTIONAL REGULATORS OF EMT: OVOL PROTEINS

The *Ovo* gene family represents an evolutionarily conserved set of genes encoding proteins belonging to the classic class of Cys$_2$His$_2$ zinc-finger-containing TFs (Dai et al., 1998; Gand et al., 1998; Mevel-Ninio et al, 1991, 1995; Garfinkel et al., 1992, 1994) that have a role in regulating expression of genes in various epithelial tissues (Watanabe et al., 2014; Lee et al., 2014; Taniguchi et al., 2017; Kitawaza et al., 2016; Chizaki et al., 2012; Mackay et al., 2006). Advancements to our understanding on the function of these proteins arises from investigations using various model systems, which uncover the conservation of genomic elements and biochemical functions that consequently allow us to make inferences on the activities Ovo proteins partake. Early studies on the family’s prototype gene, *Ovo/svb*, highlight the critical role this factor has in mediating cell fate decisions during epidermal morphogenesis in *Drosophila melanogaster* (Payre et al., 1998). Through this study, *Ovo/svb* was found to act as a downstream mediator of Wingless (Wg)/Wnt and DER-mediated signaling to control epidermal differentiation and denticle formation (Payre et al., 1998). In addition, activity of *Ovo/svb* is
involved in other developmental processes such as maintenance of the female germline (Oliver et al., 1987) and oogenesis (Dai et al., 1998).

While a single *Ovo/svb* gene is present in *Drosophila melanogaster*, mammalian systems express three orthologous *Ovol* (*Ovo*-like) genes, identified as *Ovol1*, *Ovol2*, and *Ovol3* (Dai et al., 1998; Li et al., 2002; Masu et al., 1998). Of these genes, *Ovol1* and *Ovol2* exhibit the greatest sequence identity and identical DNA-binding specificity (Wells et al., 2009; Nair et al., 2007), raising the likeliness of partial functional redundancy of these transcription factors. *Ovol3* was discovered as an expressed sequence tag (EST) and its function remains poorly understood.

Like its predecessors, Ovol proteins are structurally characterized by the presence of four Cys2His2 zinc-finger domains residing on the carboxy (C)-terminus of the polypeptide chain, that recognize and bind to specific DNA sequences on the promoter and enhancer regions of target genes (Nair et al., 2006; Watanabe et al., 2014). Investigations using techniques such as cyclic amplification of selected targets (CAST) and chromatin immuno-precipitation coupled with sequencing (ChIP-seq) have been instrumental in identifying the nucleotide sequence 5’-CCGTTA as the consensus binding motif recognized by Ovol1 and Ovol2 (Wells et al., 2009 and discussed in Chapter 3). Ovol TFs can then act as either activators or repressors of transcription; a phenomenon that is dictated by the resultant protein isoform generated from alternative promoter usage and alternative splicing. The genetic locus of both mouse *Ovol1* and *Ovol2* can produce multiple transcripts from the usage of distinct promoters (Wells et al., 2009; Nair et al., 2007), which differ depending on the presence or absence of a sequence encoding the SNAG (Snail1/GFI) repressor domain on the amino (N)-terminus of the polypeptide chain (Wells et al., 2009; Nair et al., 2007, Lin et al., 2010). Expression of the longest transcriptional isoform of
Ovol2, encoding the repressor isoform, predominates in epithelial tissues that include the epidermis and mammary epithelium.

**OVOL PROTEINS AS GUARDIANS OF EPITHELIAL IDENTITY**

Recent publications from within the Dai Laboratory show Ovol homologs Ovol1 and Ovol2 to play a prominent role as guardians of epithelial identity by negatively regulating progression of EMT in skin and mammary epithelial development (Lee et al., 2014; Watanabe et al., 2014).

In the developing skin, epithelial cells of the epidermis are responsible for creating an effective barrier that protects the organism from desiccation, bacterial infection and other insults that are routinely encountered. Production of this vital protective organ is dependent on the activity of a basal stem/progenitor population of keratinocytes, which self-renew, proliferate, and enter terminal differentiation programs to produce the suprabasal layers (stratum spinosum, stratum granulosum, and stratum lucidum) and the stratum corneum (outermost layers of epidermis). During skin development, proper suppression of EMT is necessary to maintain epithelial identity and differentiation competence of embryonic epidermal stem/progenitor cells (Lee et al., 2014). Ovol1 and Ovol2 share a redundant and compensatory function in repressing expression of TFs important for promoting SC-like traits in skin as well as proliferation, namely Snai2, Trp63, and Myc (Lee et al., 2014), as Ovol1/Ovol2-deficient epidermal cells engender a proliferative progenitor state leading to expanded basal compartment.

Simultaneously, Ovol proteins are involved in regulating progression of EMT, primarily through the transcriptional regulation of Zeb1, to allow for proper epithelial differentiation of the developing tissue (Lee et al., 2014). During epidermal differentiation, basal keratinocytes
delaminate and migrate upwards and enter differentiation programs that give rises to the protective suprabasal layers (Fuchs et al., 2007). Loss of Ovol results defective terminal differentiation that results from dramatic up-regulation of Zeb1, aberrantly promotes EMT by down-regulating epithelial markers Cdh1 and Ctnna1 that impact cellular adhesion. The findings demonstrate Ovol proteins to function in promoting terminal differentiation within the epidermis, and simultaneously restricting SC-like traits, such as proliferative potential.

In support of these findings, additional recent studies uncover the important role of Ovol proteins in maintaining epithelial identity within the mammary epithelium (Watanabe et al., 2014). The studies indicate Ovol2 as being critically important for the development of the MG and implicate how mis-regulation of this TF may support production of aggressive forms of cancer (Watanabe et al., 2014). Within the MG, deletion of Ovol2 in both basal and luminal mammary epithelial compartments results in severe defects to ductal elongation (Watanabe et al., 2014). In vivo lineage tracing and ex vivo explant cultures demonstrate that Ovol2-deficient epithelial cells transdifferentiate into mesenchymal-like cells (Watanabe et al., 2014). Molecular analysis reveals unregulated expression of myriad genes known to be involved in EMT, and identifies many such genes (e.g., Zeb1 and vimentin) as direct targets of Ovol2’s transcriptional repressor activity (Watanabe et al., 2014) (also described in CHAPTER 3: GENOME-WIDE ASSESSMENT OF OVOL2 BINDING TARGETS REVEALS ITS FUNCTION AS A GATEKEEPER OF EPITHELIAL IDENTITY). Collectively, these findings suggest that Ovol2 functions as a molecular barrier that restricts identity switching from epithelial to mesenchymal states.

Although multiple EMT genes are regulated by Ovol2, Zeb1 surfaces as a functionally relevant one not only because it is the most affected by Ovol2 loss and most tightly bound by
Ovol2, but also because its depletion via lentivirally-delivered shRNA partially rescues the ductal regeneration defect of Ovol2-deficient basal cells in a mammary fat pad transplantation assay (Watanabe et al., 2014). Further strengthening the critical importance of Ovol2 repression of Zeb1, 1) Zeb1 depletion also rescues the Ovol loss of function phenotypes in epidermal progenitor cells; and 2) a spontaneous mutation (Twirler), which resides in the first intron of Zeb1 disrupting an Ovol-binding consensus, affects multiple epithelial tissues (Kurima et al., 2011). Collectively, these findings highlight an important role of Ovol1/2 in restricting EMT-like events during morphogenesis of two different epithelial tissues, and identify Zeb1 as an important downstream target and functional mediator.

**TRANSCRIPTIONAL REGULATORS OF EMT: ZEB PROTEINS**

Zeb1 (also identified as δ-EF1, TCF8, Nil-2a, Bzp, Areb6, Meb1, Zfhx1a, and Zfhep) and Zeb2 (also identified as SIP1 and Zfhx1b) are vertebrate homologs of transcriptional regulators belonging to the Zinc Finger E-box-binding homeobox 1 (Zeb1) family that have a prominent role in vertebrate development and disease. Primarily, members from this family are classified as potent hallmark drivers of EMT, but their influence expands into embryonic development and cell differentiation, as well as its heavy involvement in cancer progression.

Structurally, proteins from this family function as complex transcriptional regulators that carry out their function directly and/or through interaction with co-regulatory proteins (Vanderwalle et al., 2009). Zeb proteins are characterized by the presence of two Cys2His2 zinc-finger domain clusters that are on opposite ends of the polypeptide chain; four residing on the N-terminus and three on the C-terminus, which bind to target genes by recognizing 5’-CACCT(G) and 5’-CAGGT(G) motif sequences (Verschueren et al., 1999). Between Zeb1 and Zeb2,
greatest degree of sequence identity is at the zinc-finger domain clusters (Verschueren et al., 1999), suggesting these two transcriptional regulators may bind to and regulate similar gene targets. Additionally, Zeb proteins possess a POU-like homeodomain sequence that is located at the center of the proteins but the ability of this domain to bind DNA has, to date, only been described in *Drosophila melanogaster* (Su et al., 1999).

Early investigations uncovering the regulatory function of Zeb1, provide evidence that Zeb1 functions as a direct transcriptional inhibitor of delta-crystallin, specifically through its repressor domain located at the N-terminus. Alternatively, Zeb1 can function as either transcriptional activator or inhibitor depending on its association with co-regulatory proteins (Vanderwalle et al. 2008). Specifically, Zeb1 has been identified to form a co-repressor complex through its association with CtBP (C-terminal-binding protein 1) at the CtBP interaction domain (CID), which further interacts with histone deacetylases, histone demethylases, and coREST proteins, to manifest vast alterations to gene expression and the epigenetic landscape (Shi et al., 2003). Moreover, Zeb1’s repressor function through association with NC2 and TIP60 has also been described (Vanderwalle et al., 2008). Interestingly, Zeb1 can function as a transcriptional activator through its association with Smad proteins, as well as histone acetyl transferase p300, to facilitate transcriptional activation of genes acting downstream of the TGF-β/Smad signaling pathway (Liu et al., 2008; Postigo et al., 2003). Recent studies also describe how ZEB1 interacts with YAP1 to positively regulate transcriptional activation of many genes that are downstream of the HIPPO signaling pathway (Lehmann et al., 2015). This context-dependent regulatory function of Zeb1 suggests the pleiotropic roles of this potent EMT-TF.

**ZEB1 AS A REGULATOR OF A CANCER STEM CELL-STATE**
Accumulating evidence suggests that a SC state is not necessarily a fixed cellular entity, but rather a transient state that responds to external stimuli and is in equilibrium with non-SCs. Bi-directional transition between these two states has been demonstrated utilizing both normal and neoplastic cultured epithelial cells, which uncover an interesting link between a transient, partial onset of EMT and production of SC-like properties, that are not only restricted to expression of cell surface markers, but functional traits as well. Transcriptional activation of ZEB1 within a breast cancer cell line, was recently observed to enhance both the molecular profile (CD44\textsuperscript{low} to CD44\textsuperscript{high}) and SC-like traits of these cells, supporting the interesting link between a transient, partial onset of EMT and production of SC-like properties. These findings also underscore the capacity of EMT to enhance the self-renewal and lineage plasticity of epithelial cells (Mani et al. 2008, Morel et al. 2008), and suggest the importance of tightly controlling the expression of ZEB1 in lineage plasticity regulation. The role of Zeb1 in epithelial SC within the normal conditions in vivo has not been studied as is an important topic of this work.

1.5. AN OVOL2-ZEB1 MUTUAL INHIBITORY LOOP RESPONSIBLE FOR GOVERNING A PLASTIC, INTERMEDIATE CELL STATE

Mounting evidence suggests the remarkable lineage plasticity of somatic cells, and EMT represents a phenomenon that describes the innate plasticity of epithelial cells. Moreover, as described, a partial EMT is suggested to influence a SC-state. A hybrid, partial EMT in cells lead to an expression of both epithelial and mesenchymal traits, which can endow these cells with migratory potential and simultaneous maintenance of a degree of epithelial identity and may influence the cell fate plasticity. Mutual repression between EMT-TFs Snail/Zeb1 and EMT-
inhibiting microRNAs miR34 and miR200 are critical for controlling the dynamic EMT process. Mathematic modeling, which incorporates cross-regulation of Zeb1 and Ovol2 into the core EMT-regulatory network, reveals the production of a four-state system, where two intermediate states emerges between terminal epithelial and mesenchymal states (Hong et al., 2015) (FIGURE 1.5). These intermediate states exhibit differential propensity to differentiate toward either an epithelial or mesenchymal identity. Interestingly, modulation of Ovol2 and Zeb1 expression has the capacity to influence the directionality of differentiation (Hong et al., 2015). These findings highlight the possibility that balanced expression of Ovol2 and Zeb1 may be critical for generating a plastic intermediate cell state that is compatible with an SC state with cell fate plasticity to differentiate toward a more epithelial or mesenchymal lineage. Specifically within the MG, a balance between Ovol2 and Zeb1 may be important for a mammary SC fate with the capacity to differentiate towards either a luminal or basal fate.

1.6. THESIS OBJECTIVES

As described in the previous sections, a partial EMT is suggested to expand the lineage plasticity of epithelial cells. Despite a growing number of leads that champion this notion, it has not been vigorously tested in vivo within a physiologically relevant setting. Despite in vitro evidence suggesting ZEB1 to have a role in producing a SC state, its function within adult epithelial tissues is to be determined. Moreover, whether this EMT-TF influences the SC state within the MG is far less clear. Through this study, my goal is to use both in vivo and ex vivo approaches to characterize the role of Zeb1 in determining mammary SC fates and to understand its genetic interaction with Ovol2. In addition, I am interested in understanding the molecular mechanisms underlying such functions. In the following chapters, I will describe 1) the reagents
and methods that I used for my thesis study (Chapter 2); 2) my initial efforts in identifying molecular targets of Ovol2, uncovering the first mechanistic link between Ovol2 and Zeb1 (Chapter 3); and 3) my experiments to characterize the role of Zeb1 in mammary epithelial regeneration and SC differentiation, as well as to elucidate the functional interaction between Zeb1 and Ovol2 in mammary SC fate determination (Chapter 4).
FIGURE 1.1 Schematic representation of a mammary duct and TEB (depictions of mammary epithelium at puberty). The mammary epithelium is composed of a cellular bi-layer that is composed of an inner luminal population and an outer population of basal/myoepithelial cells. Image taken from Visvader et al., 2014.
FIGURE 1.2 Markers of prospectively identified epithelial subsets in the mouse mammary gland. Summary of cell surface markers used for the isolation of epithelial cell subsets from the mouse mammary gland. ER denotes ERα. Image taken from Visvader et al., 2014.
FIGURE 1.3 Schematic diagram of the primary stages of mammary gland ontogeny in the embryo and adult. In the mouse embryo, the placodes (visible at embryonic day 11.5 [E11.5]) evolve into mammary buds that penetrate the underlying mesenchyme around E13.5. These buds sprout by E15.5 and develop a lumen. By E18.5, a small arborized gland that has invaded the developing fat pad is evident. In the postnatal animal, development of the mammary gland remains relatively dormant until puberty at 3 wk, when profound morphogenesis occurs, largely under the control of estrogen (E). In the young adult gland, progesterone (Pg) regulates side branching, while in pregnancy, the steroid hormones estrogen, progesterone, and prolactin (Prl) all play roles in alveolar expansion. In the late stages of pregnancy and during lactation, the peptide hormone prolactin plays a key role in establishing the secretory state. After lactation, the gland involutes and returns to a state that resembles the virgin gland. Image and caption taken from Visvader et al., 2014.
FIGURE 1.4 Schematic diagram describing the biochemical alterations associated with progression of EMT. Intermediate states between epithelial (E) and mesenchymal (M) states represent bi-stable, partial-EMT states that have recently been associated with enhanced stemness. Image taken from Nieto et al., 2016.
FIGURE 1.5 A. Influence diagram of the EMT/MET system. Blue icon: epithelial promoting factor. Yellow icon: mesenchymal promoting factor. B. Summary of Hong et al., 2015 describing a cross regulatory network that controls a four state EMT system. Evidence of the four state system is shown through systems biology approach using human normal mammary epithelial cell line MCF10A. The four-state system is maintained through activity/interaction of Zeb1 and Ovol2. Possible transition among the each state is shown through altering expression of Ovol2, Zeb1, or TGF-b (EMT induction). The multiple intermediate states display distinct capacity to differentiate. Images taken from Hong et al., 2015.
<table>
<thead>
<tr>
<th>Transcription factor</th>
<th>Downregulated expression during EMT</th>
<th>Upregulated expression during EMT</th>
<th>Regulatory signalling pathways</th>
<th>Refs</th>
</tr>
</thead>
<tbody>
<tr>
<td>SNAIL1 and SNAIL2</td>
<td>E-cadherin, caduclin, occludin, Crumbs3, PALS1, PATJ, cytokeratin, desmoplakin and plakoglobin</td>
<td>Fibronectin, N-cadherin, collagen, MMP15, MMP2, MMP9, TWIST, ID1, ID2, ZEB1 and ZEB2</td>
<td>TGFβ–SMAD3, WNT–β-catenin, MAPK</td>
<td>11,45,59,60,61, 66,159,185</td>
</tr>
<tr>
<td>TWIST1</td>
<td>E-cadherin, caduclin, occludin, desmoplakin and plakoglobin</td>
<td>Fibronectin, N-cadherin and α5 integrin</td>
<td>MAPK</td>
<td>11,45,67,68,72</td>
</tr>
<tr>
<td>ZEB1 and ZEB2</td>
<td>E-cadherin, ZO1, Crumbs3 and plakoglobin</td>
<td>N-cadherin and MMPs</td>
<td>TGFβ–SMAD3, WNT–β-catenin and RAS–MAPK</td>
<td>11,45,73–75, 77,142, 143</td>
</tr>
<tr>
<td>FOXD3</td>
<td>Unknown</td>
<td>Unknown</td>
<td>β1 integrin and laminin</td>
<td>266</td>
</tr>
<tr>
<td>FOXC2</td>
<td>E-cadherin</td>
<td>Fibronectin, vimentin, N-cadherin and αSMA</td>
<td>TGFβ–SMAD3</td>
<td>267</td>
</tr>
<tr>
<td>FOXP1</td>
<td>E-cadherin, claudin1, occludin, desmoglein1β, desmoglein2, desmocollin2 and desmoplakin</td>
<td>Fibronectin and N-cadherin</td>
<td>Unknown</td>
<td>268</td>
</tr>
<tr>
<td>FOXQ1</td>
<td>E-cadherin</td>
<td>Fibronectin, N-cadherin and vimentin</td>
<td>Unknown</td>
<td>269</td>
</tr>
<tr>
<td>FOXO3A</td>
<td>E-cadherin</td>
<td>SNAIL1</td>
<td>AKT</td>
<td>270</td>
</tr>
<tr>
<td>FOXA1</td>
<td>E-cadherin, vimentin and SNAIL1</td>
<td>Fibronectin, vimentin and SNAIL1</td>
<td>TGFβ, HGF and AKT</td>
<td>271</td>
</tr>
<tr>
<td>FOXA2</td>
<td>E-cadherin and ZO1</td>
<td>Fibronectin, vimentin, N-cadherin, SNAIL1 and SNAIL2</td>
<td>TGFβ, HGF and AKT</td>
<td>271</td>
</tr>
<tr>
<td>Serpent, GATA4 and GATA6</td>
<td>E-cadherin, Crumbs and claudin</td>
<td>N-cadherin and MMP1</td>
<td>Unknown</td>
<td>81</td>
</tr>
<tr>
<td>HMG2</td>
<td>E-cadherin</td>
<td>SNAIL1, SNAIL2 and TWIST</td>
<td>TGFβ–SMAD3</td>
<td>103</td>
</tr>
<tr>
<td>SOX9</td>
<td>Unknown</td>
<td>SNAIL2</td>
<td>BMPs and PKA</td>
<td>83</td>
</tr>
<tr>
<td>KLF8</td>
<td>E-cadherin</td>
<td>MMP9</td>
<td>Unknown</td>
<td>272</td>
</tr>
<tr>
<td>CBFα–KAP1</td>
<td>Unknown</td>
<td>FSP1</td>
<td>Unkown</td>
<td>273</td>
</tr>
<tr>
<td>ZNF703 (also known as Zippo1 in mice)</td>
<td>E-cadherin</td>
<td>Vimentin, N-cadherin, SNAIL1 and cytokeratin</td>
<td>RHO-GTPase</td>
<td>274</td>
</tr>
<tr>
<td>PRK1</td>
<td>E-cadherin</td>
<td>Vimentin and laminin</td>
<td>BMP2 and TGFβ</td>
<td>275</td>
</tr>
</tbody>
</table>

αSMA, α-smooth muscle actin; BMPs, bone morphogenetic proteins; CBFα, CArG-binding factor A; E-cadherin, epithelial cadherin; EGF, epidermal growth factor; EMT, epithelial–mesenchymal transition; FGF, fibroblast growth factor; FOX, forkhead box; FSP1, fibroblast-specific protein 1; HGF, hepatocyte growth factor; HMG2A, high mobility group A2; ID, inhibitor of differentiation; KAP1, KRAB-associated protein 1 (also known as TIF1β); KLF8, Krueppel-like factor 8; MMP, matrix metalloproteinase; N-cadherin, neural cadherin; NF-κB, nuclear factor-κB; TGFβ, transforming growth factor β; PALS1, protein associated with Lin-7 1; PATJ, PALS1-associated tight-junction protein; PKA, protein kinase A; PRK1, paired-related homeobox 1; SOX, SRY box; ZNF703, zinc-finger 703; ZO1, zonula occludens 1; ZEB, zinc-finger E-box-binding.

**TABLE 1.1** Table describing well-known EMT-TFs and their associated regulatory targets that manifest this trans-differentiation program. Image taken from Nieto et al., 2016.
MATERIALS AND METHODS

Mouse Strains

The conditional knockout Ovol2 mouse line (Ovol2 \(^{\text{flox/flox}}\)) (C57BL/6J background) was previously described by Unezaki et al., 2007. The conditional knockout Zeb1 mouse line (Zeb1 \(^{\text{flox/flox}}\)) (C57BL/6J background) was recently described in Brabletz et al., 2017 and was provided to us by the Thomas Brabletz lab. For our studies on the Zeb1-Ovol2 EMT cross-repression circuit, we crossed Zeb1 \(^{\text{flox/+}}\); Ovol2 \(^{\text{flox/+}}\) males with Zeb1 \(^{\text{flox/flox}}\); Ovol2 \(^{\text{flox/flox}}\) females to generate Zeb1 \(^{\text{flox/flox}}\); Ovol2 \(^{\text{flox/flox}}\) double mutants and controls. Genotyping for these animals was performed using PCR primers that have been previously described (Andl et al., 2004; Dai et al., 1998; Unezaki et al., 2007). Additionally, wild-type C57BL/6J (Jackson Laboratories; Cat. No. 000664) mice were purchased from Jackson Laboratories. All mouse lines were maintained in a pathogen free facility, following mouse procedures that conform and have been approved by the UC Irvine Institutional Animal Care and Use Committee (IACUC) (Protocol No. 1999-2133).

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Genotyping primers</strong></td>
<td></td>
</tr>
</tbody>
</table>
| Zeb1 Flox | F: 5’-CGTGATGGAGCCAGAATCTGACC-3’  
R: 5’-GCCCTGTCTTCTCACAGTGTGG-3’        |
| Zeb1 Null | F: 5’-CGTGATGGAGCCAGAATCTGACC -3’  
R: 5’-GCCATCTCACCAGCCTACTGTC-3’         |
| Ovol2 Flox | F: 5’-CCCAGACCCCTTCTCTGCT-3’  
R: 5’-CGCCCCAGTGACAGGCTAG-3’         |
| Ovol2 Null | F: 5’-GCCATCTCTGCAGCGCATTC-3’  
R: 5’-GCCGGTAAACATCCCAAC-3’            |

Isolation of Mouse Primary Keratinocytes
Preparation of primary keratinocytes was performed as previously indicated (Lee et al., 2014). An updated and improved protocol has since been established by Daniel Haensel (not described in this document). Briefly, skin was isolated from newborn (P0) wild-type C57BL/6J mice (Jackson Laboratories; Cat. No. 000664) and quickly washed with 20 ug/mL gentamycin (Sigma-Aldrich) in 1xPBS and Betadine Microbicide (Betadine), and followed with a rinse with 70% Ethanol. Skin was dissociated into epidermal and dermal components by treatment with 5 U/mL Dispase II solution (Stem Cell Technologies; Cat. No. 7913) for 6-7 hours at 4°C. After, the epidermis was incubated with 0.25% Trypsin-EDTA (Sigma Aldrich) for 10 minutes at 37°C to break intra-cellular adhesion, and subsequently passed through a 70um-pore cell strainer (Fisher Scientific; Cat. No. 08-771-1) to generate a single cell suspension. 10% FBS DMEM was added to inactivate Trypsin-EDTA. Cells are then centrifuged at 1000x rpm to pellet and resuspended with Cnt-02 Mouse Epidermal Keratinocyte Medium (Cell-n-Tec). Cells were counted with hemocytometer using trypan blue exclusion and plated at a high density of 4.00x10^6 cells/25cm^2. Keratinocytes are incubated at 35°C, 5% CO₂ for 1-2 days. Primary keratinocytes were then processed for ChIP-PCR (see below).

Mammary Cell Preparation and FACS

Preparation of single mammary cell suspensions and the procedures for fluorescence-activated cell sorting (FACS) were performed as previously indicated (Gu et al., 2009, Gu et al., 2013). Briefly, all five mammary gland pairs were isolated from 8-12-week-old wild-type C57BL/6J mice (Jackson Laboratories; Cat. No. 000664) and dissociated by incubating with 300 U/mL collagenase (Sigma, Cat. No. C9891) and hyaluronidase (Sigma, Cat. No. H3506) for 1.5 hours at 37°C. The resulting cell mixture was resuspended in red blood cell lysis buffer (Sigma,
Cat No. R7757) and underwent further dissociation by treatment with 0.25% trypsin-EDTA (Gibco, Cat. No. 25200), 10 g/L DNase (Sigma, Cat. No. DN25), and 5 mg/mL Dispase (Stem Cell Technologies, Cat. No. 07913). Cell aggregates were removed from the single cell suspension by filtration using a 40 um-pore mesh filter (Corning, Cat. No. 352098). The total number of single, viable mammary cells resulting from the isolation were calculated by counting with hemocytometer and using trypan blue exclusion.

To sort subpopulations of the mammary gland, the primary mammary cell suspension was treated with fluorophore-conjugated antibodies specific for the following surface markers:
- Anti-CD49f-FITC (1:250, 102205 Bio Legend),
- Anti-EpCAM-PE-Cy7 (1:250, 118215, Bio Legend),
- Anti-Lineage-APC [(1:50; APC-CD45 (559864, BD Biosciences), APC-CD31 (1:50, 551262, BD Biosciences), APC-TER119 (1:50, 557909, BD Biosciences)].

For compensation, single IgG stains that corresponded to each antibody-conjugated fluorophore used in the sort were prepared (same dilution). 7-AAD (BD Biosciences, Cat. No. 559925) was used as a marker to sort live cells for downstream experimentation. Cells were sorted at low pressure (20psi using a 100um nozzle) on a FACSAria (Becton Dickenson UK) equipped with violet (404nm), blue (488nm), green (532nm), yellow (561nm) and red (635nm) lasers. The total cell sample and the collection tube were maintained at 4°C. Single stained samples were used as compensation controls.

3D-Matrigel Culture (Proliferative conditions and Differentiation conditions)

For 3D-Matrigel culture of mammary epithelial cells (sorted and unsorted), single cells were embedded into 100% chilled-growth factor-reduced Matrigel (BD Biosciences; Cat. No. CB-40230) and plated onto 8-well chamber slides (Nunc LabTek II, Cat. No. C7182) at a density
of 1x10^4 or 2x10^4 cells/50uL Matrigel/well. After the Matrigel set, the gel was covered with 400uL EpiCult-B medium (Stem Cell Technologies; Cat. No. 05610) containing 1x proliferation supplements (included with media), 10 ng/mL EGF (Millipore; Cat. No.01-107), 10 ng/mL FGF-2 (PeproTech; Cat. No. 100-18B), and 4 ug/mL Heparin (Stem Cell Technologies; Cat. No. 07980). The medium was replaced every 3 days. To induce differentiation, EpiCult-B medium was removed after a week and replaced with Differentiation Medium (Shackleton et al., 2006), which consisted of DMEM/F12 Basal Medium (Invitrogen, Cat. No. 12500-062) containing 1% fetal bovine serum (FBS) (Omega Scientific; Cat. No. FB-02), 5 ug/mL Insulin (Sigma; Cat. No. 16634), 500 ng/mL Hydrocortisone (Calbiochem; Cat. No. 386698), and Prolactin (Sigma; Cat. No. L6520). Organoid cultures were cultured up to 21 days, fixed with 4% paraformaldehyde for 30 mins at room temperature, and visualized using a phase contrast light microscope. Organoid colonies were characterized as described in Gu et al., 2013, Jarde et al., 2016, Dontu et al., 2003; Petersen et al., 1992; Shackleton et al., 2006. Morphology of GFP-positive organoids was assessed by visualizing with a Nikon DiaPhot 300 fluorescent microscope (Nikon). NIS Elements microscopy software (Nikon) was used to measure organoid diameter (2 total measurements per organoid; taken at perpendicular regions crossing center of organoid). Statistical analysis: p-value for measurement of the average organoid diameter was determined using two-tailed student t-test between two sample groups.

For expression analysis, RNA was isolated using a SpinSmart Total RNA Mini Purification Kit (Denville, Cat. No. CM-610-50) after digesting the Matrigel with 5 mg/mL Dispase (Stem Cell Technologies, Cat No. 07913) for 30 mins at 37°C. For RT-qPCR analysis, cDNA was prepared using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) per manufacturer’s instructions. Real-time PCR was performed using a CFX96 RT-
qPCR system and SYBR Green Supermix (BioRad). Comparative analysis using delta-delta Ct method was performed between the gene of interest and the housekeeping gene Gapdh/GAPDH.

Adenoviral/Lentiviral Infection

Lentiviral Constructs

shRNAs targeting the mouse Zeb1 transcript were generated by cloning the annealed oligo-nucleotides (Cat. No. TRCN0000070818 and TRCN0000070819, Sigma Aldrich) (TRCN0000070818 Mature Antisense: 5’-TTTGAACTCATAATCCACAGG-3; TRCN0000070819 Mature antisense: 5’-ATGTGGCTCAATCAAATGATGC-3’) into the AgeI/EcoRI site of pLKO.1-ZsGreen (Gu et al., 2013). Scrambled control shRNAs had the mature sequence of 5’-CCTAAGGTTAAGTCGCCCTCGCTC-3’ (Addgene; Cat No.1864). Production of the lentiviral particles was achieved using the calcium phosphate transfection method, as previously described (Gu et al., 2012). Briefly, shRNA-expressing pLKO.1-ZsGreen plasmids (Addgene; Cat No.10878) were co-transfected with psPAX2 (Addgene; Cat No.12260) and pMD2.G (Addgene; Cat No.12259) plasmids into 293T-LentiX cells, where a concentrated viral supernatant was generated and collected 3 and 4 days post transfection, and filtered for purity with a 0.45um pore-filter (Olympus Plastics; Cat No.25-240). The transduction unit of each viral solution was calculated by measuring the percentage of GFP-positive cells through flow cytometry using a BD LSR II (BD Biosciences).

To knockdown expression of the Zeb1, mouse mammary epithelial cells were infected with lentiviral particles in suspension at a multiplicity of infection of 10 (MOI 10), by centrifuging at 350 rcf for 90 minutes at room temperature. The efficiency of gene transfer was elevated with the addition of 4 ug/mL Polybrene (Sigma Aldrich) in the suspension. After
infection, cells were plated at a density of 2-4x10^3 cells/cm^2 and were cultured for 1-2 weeks under proliferative conditions, using EpiCult-B medium (Stem Cell Technologies; Cat. No. 05610) containing 1x proliferation supplements (included with media), 10 ng/mL EGF (Millipore; Cat. No.01-107), 10 ng/mL FGF-2 (PeproTech; Cat. No. 100-18B), and 4 ug/mL Heparin (Stem Cell Technologies; Cat. No. 07980). Successful knockdown of Zeb1 was determined by mRNA and protein expression analysis.

Adenoviral Constructs

Control and Cre-expressing adenoviruses were obtained from Vector Bio Labs (Ad-Cre-GFP #1700 and Ad-GFP #1060) and used to infect an unsorted population of mammary epithelial cells at a multiplicity of infection of 50 (MOI 50). Transduction of mammary epithelial cells with adenoviral particles used the same methodology as transduction with lentiviral particles in suspension. Briefly, mouse mammary epithelial cells were infected with adenoviral particles in suspension at a multiplicity of infection of 50 (MOI 50), by centrifuging at 350 rcf for 90 minutes at room temperature. After infection, cells plated at a density of 2-4x10^3 cells/cm^2 and were cultured for 1 week under proliferative conditions, using EpiCult-B medium (Cat. No. 05610, Stem Cell Technologies) containing 1x proliferation supplements (included with media), 10 ng/mL EGF (Cat. No.01-107, Millipore), 10 ng/mL FGF-2 (Cat. No. 100-18B, PeproTech), and 4 ug/mL Heparin (Cat. No. 07980, Stem Cell Technologies).

Successful recombination of the Zeb1 and Ovol2 floxed loci was determined by genotyping using mouse genotyping primers that identify floxed and recombined (null) loci (see primer sequences below) as well as by mRNA expression analysis.

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genotyping primers</td>
<td></td>
</tr>
<tr>
<td></td>
<td>F: 5'</td>
</tr>
<tr>
<td>--------</td>
<td>--------</td>
</tr>
<tr>
<td>Zeb1 Flox</td>
<td>5'CGTGATGGAGCCAGAATCTGACCCC</td>
</tr>
<tr>
<td>Zeb1 Null</td>
<td>5'CGTGATGGAGCCAGAATCTGACCCC</td>
</tr>
<tr>
<td>Ovol2 Flox</td>
<td>5'CCCAGACCCCTTCCTGTCTTCCTCCT</td>
</tr>
<tr>
<td>Ovol2 Null</td>
<td>5'GCCATCTCTGAGCGCCATTTTC</td>
</tr>
</tbody>
</table>

Cleared Mammary Fat Pad Transplantation Assay

The mammary fat pad transplantation assay was performed as previously described (Morris et al., 2004) with the following modifications. Briefly, FACS-sorted basal mammary epithelial cells were isolated from the mammary glands of an 8-12-week old wild-type female donor C57BL/6J (Jackson Laboratories; Cat. No. 000664) and transduced with lentiviral particles expressing either shZeb1 constructs or a Scrambled control sequence. The transduced cells were gently suspended in 50% chilled-growth factor-reduced Matrigel (BD Biosciences; Cat. No. CB-40230)/50% EpiCult-B medium (Stem Cell Technologies; Cat. No. 05610) containing 1x proliferation supplements (included with media), 10 ng/mL EGF (Millipore; Cat. No.01-107), 10 ng/mL FGF-2 (PeproTech; Cat. No. 100-18B), 4 ug/mL Heparin (Stem Cell Technologies; Cat. No. 07980) at a density of 2x10^3 cells/10uL Matrigel, and injected into the cleared fat pad of mammary gland No. 4 belonging to a 3-week old C57BL/6J (Jackson Laboratories; Cat. No. 000664) mouse host. Mammary outgrowths were analyzed 8 weeks after transplantation. GFP fluorescence was visualized with a Nikon DiaPhot 300 fluorescent microscope (Nikon) or Keyence epi-fluorescence microscope (Keyence USA).
Branching and ductal length measurements of GFP-positive trees were carried out using ImageJ software. Statistical analysis: p-value for measurement of branching and ductal length were determined using two-tailed student t-test between two sample groups.

**Single-cell RNA-seq (scRNAseq)**

To uncover the cellular heterogeneity within the basal compartment of the adult mouse mammary gland, viable mammary epithelial basal cells were FACS-sorted from 8-week old wild-type female C57BL/6J mice (Jackson Laboratories; Cat. No. 000664) using Anti-CD49f-FITC (1:250, 102205 Bio Legend), Anti-EpCAM-PE-Cy7 (1:250, 118215, Bio Legend), Anti-Lineage-APC [(1:50; APC-CD45 (559864, BD Biosciences), APC-CD31 (1:50, 551262, BD Biosciences), APC-TER119 (1:50, 557909, BD Biosciences)] (See above for procedure). For compensation, single IgG stains that corresponded to each antibody-conjugated fluorophore used in the sort were prepared (same dilution). 7-AAD (BD Biosciences, Cat. No. 559925) was used as a marker to sort live cells for downstream experimentation. C1 system operation and cell preparation methodology are described in the Fluidigm reference sheet “Using C1 to Generate Single-Cell cDNA Libraries for mRNA Sequencing.” Briefly, the C1 Harvest Reagent (Fluidigm; Cat. No. 100-6201), C1 Preloading Reagent (Fluidigm; Cat. No. 100-6201), C1 Blocking Reagent (Fluidigm; Cat. No. 100-6201), and Cell Wash Buffer (Fluidigm; Cat. No. 100-6201) were properly loaded onto the IFC and primed in the C1 system by running the mRNA Seq Prime (1771x/1772x/1773x) script.

Promptly after, the viable mammary basal cells in suspension media were mixed with C1 suspension reagent (Fluidigm; Cat No. 100-5315) to reach an optimal buoyant cell suspension and loaded onto the primed IFC (See below: Cell Loading Logistics). The IFC was then placed
in C1 and ran using protocol mRNA Seq: Cell Load (1771x/1772x/1773x) to capture single cells. After completion of the Cell Load script, capture sites on the IFC were imaged using a Keyence epi-fluorescence microscope (Keyence USA) to record positioning of cell singlets and doublets. Reverse transcription and cDNA pre-amplification were performed using the SMARTer Ultra Low RNA Kit (Clontech) and processed in the C1 system by running script mRNA Seq: RT and Amp (1771x/1772x/1773x). A total of 103 cells were captured in two independent IFCs. Position of doublets was recorded upon viewing IFC chip through microscopy and positions were excluded in post-sequencing analysis.

After completion of RT and cDNA amplification, the samples were harvested and quantified through the PicoGreen assay (Thermo-Fisher) using the Agilent Bioanalyzer (Agilent Genomics). Equal amounts of cDNA from each sample were tagmented and prepared for library synthesis using the using Nextera XT DNA Sample preparation kit with 96 indices (Illumina), per the protocol supplied by Fluidigm. Libraries were multiplexed and sequenced as paired-end on a Next-Seq 4000 Illumina sequencing platform.

**Cell Loading Logistics**

Determination of the appropriate size of the microfluidic chip to be used for scRNAseq, the average diameter of FACS-sorted mammary basal cells was determined by visualizing cells using Nikon DiaPhot 300 fluorescent microscope (Nikon) and taking measurements using NIS Elements microscope software (Nikon). Two total measurements per cell; taken at perpendicular regions crossing center). Quantification of cell diameter was validated using a Countess II FL Automated Cell Counter (Thermo Fisher Scientific). The 10-17 um 96-cell IFC (Fluidigm; Cat. No. 100-5763) was used to capture FACS-sorted mammary basal cells.
To prepare for cell loading, an optimized ratio of suspension reagent (Fluidigm; Cat No. 100-5315) to suspension media must be established to create a neutrally buoyant cell suspension that will travel throughout the microfluidics. To determine the optimized ratio, cells were suspended in media at a concentration of 166-250 cells/μL and mixed with suspension reagent (Fluidigm; Cat No. 100-5315) at the following ratios:

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>5:5</td>
<td>5</td>
<td>0</td>
<td>5</td>
</tr>
<tr>
<td>6:4</td>
<td>5</td>
<td>1</td>
<td>4</td>
</tr>
<tr>
<td>7:3</td>
<td>5</td>
<td>2</td>
<td>3</td>
</tr>
<tr>
<td>8:2</td>
<td>5</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>9:1</td>
<td>5</td>
<td>4</td>
<td>1</td>
</tr>
<tr>
<td>10:0</td>
<td>5</td>
<td>5</td>
<td>0</td>
</tr>
<tr>
<td>Total</td>
<td>30</td>
<td>15</td>
<td>15</td>
</tr>
</tbody>
</table>

Cell mixtures were resuspended gently by pipetting up and down, transferred to unused inlets of a C1 IFC, and monitored by microscopy over a period of ten minutes to assess buoyancy of cells. The optimal ratio of media to suspension reagent will not change buoyant distribution of cells throughout solution.

Chromatin Immunoprecipitation (ChIP)-PCR and ChIP-seq

Chromatin Immunoprecipitation was performed as previously described (Gu et al., 2009) with several modifications. HC11 cells, MDA-MB-231 cells, and primary keratinocytes were grown in 150-mm dishes and cross-linked with 1% formaldehyde for 10 minutes at room temperature, followed by quenching with 0.125M glycine for 5 minutes at room temperature. After washing, chromatin was sheared to produce ~100-500 bp fragments using the Bioruptor Sonicator (Diagenode Inc.) at “high” setting and cycled 30s ON /30s OFF for 30m. A small
aliquot of the recovered supernatant underwent subsequent reverse crosslinking and DNA purification, which was used to assess concentration and shearing efficiency (input sample). For HC11 ChIP-seq, immunoprecipitation was performed by incubating 25 ug of crosslinked chromatin overnight at 4C with control IgG (Santa Cruz Biotechnology, sc-2027) or anti-Ovol2 antibody (Mackay et al., 2006). For MDA-MB-231 ChIPseq, immunoprecipitation was performed by incubating 25 ug of crosslinked chromatin overnight at 4C with control IgG (Santa Cruz Biotechnology, sc-2027) or anti-Zeb1 (Novus Biologicals, NBPI-88845) or anti-Zeb1 (Santa Cruz Biotechnology, h-102). Note: Both antibodies work well for ChIP and IF. Conversely, anti-Zeb1 (Santa Cruz Biotechnology, E-20) requires optimization for IF.

Antibody-chromatin immuno-complexes were purified with Dynabeads protein A (Thermo-Fisher, Cat. No. 10001D) that underwent washes with Low Salt Buffer, High Salt Buffer, Lithium Chloride Buffer, and TE buffer, for 5 minutes each at room temperature to remove impurities. Immuno-precipitated DNA was eluted and reverse-crosslinked by incubating overnight with freshly prepared elution buffer with Proteinase K at 65C. ChIP DNA was recovered and purified using phenol: chloroform: isoamyl alcohol purification.

Prior to library preparation and sequencing, both ChIP DNA quality and fragment length were determined using Qubit Sensitivity DNA assay (Thermo-Fisher, Cat. No. Q32851). ChIP DNA libraries were prepared following instructions for NEXTflex ChIPseq Library Systems (BioScientific). Briefly, ChIP DNA fragment ends were repaired and phosphorylated using Klenow, T4 DNA polymerase and T4 polynucleotide kinase. After adaptors were ligated using NEXTflex ChIPseq barcode kit, DNA was size-selected by gel purification and amplified through PCR using NEXTflex ChIP Primer Mix (NEXTflex ChIPseq Barcode Kit) under the following conditions: 2 minutes at 98C, 30 seconds at 98C, 30 seconds at 65C (16 cycles), 1
minute at 72C, and 4 minutes at 72C. Library quality was determined using the Qubit DNA Sensitivity Assay. Sequencing was performed at the University of California, Irvine, High Throughput Genomics Facility using an Illumina Genome Analyzer IIx, using 1 lane for all samples, 50bp singleton sequencing. For HC11 ChIPseq, sequencing was performed using the HiSeq2500, using 1 lane per sample, 50bp singleton sequencing. For MDA-MB-231 ChIPseq, sequencing was performed using the HiSeq4000, using 1 lane per sample, 50bp singleton sequencing.

ChIPseq Analysis

After sequencing, reads were mapped to the appropriate species-specific reference genome using Bowtie (Langmead et al., 2009). Short sequence reads of low complexity that mapped or complex repeats that did not map to the reference genome by chance were removed from the downstream analysis. BED files were created and used as the input for downstream data processing, as well as for visualization in the UCSC Genome Browser (http://genome.ucsc.edu/index.html). Peak identification was accomplished using Model-based Analysis of ChIPseq (MACS) (Zhang et al., 2008) using the following cutoff parameters: \( mfold \) 10, \( bandwidth \) 300bp, \( p-value < 10^{-5} \). To associate peaks to specific genes (peak-to-gene analysis), a RefSeq (NCBI37/mm9) file containing transcriptional start site coordinate information was used. To identify motifs present at binding peaks, DNA peak sequences were retrieved using Galaxy Web Browser (http://main.g2.bx.psu.edu) (Liu et al., 2011) and analyzed using MEME (Bailey et al., 2006).

ChIP-PCR primers were designed based on the Ovol2 locations identified through ChIPseq analysis. Primer sequences used for the study are listed below:
Ovol2 ChIPseq

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChIP-PCR primers</td>
<td></td>
</tr>
</tbody>
</table>
| Cdh2 | F: 5’-AGCTCCTTGATCTCCCGTCT-3’  
R: 5’-AAACAAATAGCGGGCTCGG-3’ |
| Gapdh | F: 5’-TGTGCCAAGCATTGGTATAAC-3’  
R: 5’-TATGTCTGACCAGAGGAGG-3’ |
| Snai1 | F: 5’-GCGCAAGGTCGATGCTCA-3’  
R: 5’-TGCTCCGCTCTCAGTGCTCC-3’ |
| Snai2 | F: 5’-GGAGCTCAGTCGGGAGGAG-3’  
R: 5’-ACCAGCGTTTTACTCGGCAC-3’ |
| Twist1 | F: 5’-ACTAGAGGTTCAGACTAGAGGT-3’  
R: 5’-GGCGCGAGGCAAGAAAGT-3’ |
| Vim | F: 5’-GGCTCTCCCCGGTCTGCAAG-3’  
R: 5’-AGAGCCACTGCTCTGAGTCTCC-3’ |
| Zeb1 | F: 5’-CGTTATGGCTCCTGCTG-3’  
R: 5’-ATGCGGCTCGTCTCTTTG-3’ |
| Zeb2 | F: 5’-GTTTCTTGCTGAGTCGCTG-3’  
R: 5’-CCCGCTCTCGCTCCTCTT-3’ |

ZEB1 ChIPseq

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>ChIP-PCR primers</td>
<td></td>
</tr>
</tbody>
</table>
| OVOL2 (proximal) | F:5’-GTGATAGGGGTATGAAGG-3’  
R: 5’-CACAGAGGAACATTTGGGAGT-3’ |
| OVOL2 (distal) | F: 5’-AGCCAGAAACCTCCTACCA-3’  
R: 5’-CCTACTCAGGAGGATCTGCT-3’ |
| CCND1 | F: 5’-GAAAACCTCGGAGCCTGAG-3’  
R: 5’-GGAACCTCCTCCAGGGG-3’ |
| AXIN2 (site 1) | F: 5’-TCGTATTTGCTCAGGACTACA-3’  
R: 5’-ACCTAACCACACGGATTCTCC-3’ |
| AXIN2 (site 2) | F: 5’-ATCGTATTGGCGCAACTACA-3’  
R: 5’-CCTACTCACCACAGGTCTCC-3’ |
| GAPDH | F: 5’-TCCTGTTCATCACAGGCTG-3’  
R: 5’-GAATGTCAACAGGGAGG-3’ |

** CCND1 and GAPDH ChIP primers were also used for YAP1 ChIP-PCR study.**

RNAseq
MCF10A cells were infected with either pHIV-ZsGreen-Ovol2, pHIV-Z-Zeb1, or pHIV-ZsGreen-Empty Vector (EV) at a multiplicity of infection (MOI) of 10. Six days after infection, the GFP+ population was isolated by FACS and total RNA was extracted with a SpinSmart RNA Mini Purification kit (Cat. No. CM-610-50, Denville). Optimal-quality RNAs were considered for cDNA library preparation (RIN > 8.8). Full-length cDNA library amplification was performed as previously described (Picelli et al., 2013; Picelli et al., 2014). Briefly, 1ng of total RNA was reverse-transcribed and the resulting cDNA was pre-amplified for 17 cycles. Tagmentation of cDNA was carried out using the Nextera DNA Sample Preparation Kit (Illumina). The Tn6 tagmentation reaction was carried out at 55°C for 5mins and purified using a PCR Purification Kit (Qiagen). Adapter-ligated fragments were amplified using limited cycle enrichment PCR with Nextera barcodes for 7 continuous cycles. The resulting libraries were further purified using AMPure XP beads (Beckman Coulter). Libraries were multiplexed and sequenced as paired-end on a Next-Seq 2500 Illumina sequencing platform.

**RNAseq Analysis**

After sequencing, reads were mapped to the appropriate species-specific reference (Hg19) genome using BowTie2 (Langmead et al., 2012) and splice junctions between exons were mapped using Tophat2 (Kim et al., 2013). Differential analysis of gene and transcript expression was accomplished using CuffDiff (Trapnell et al., 2013). Gene FPKM values were used for downstream analysis that included Gene Set Enrichment Analysis (GSEA) (Subramanian et al., 2005) and Gene Ontology (GO) Analysis (Ashburner et al., 2000).

**Cell Lines**
MDA-MB-231 was purchased from ATCC. MDA-MB-231 cells were cultured and maintained using Dulbecco’s Modified Eagle’s Medium (DMEM) that was supplemented with 10% fetal bovine serum (FBS) and 1% Pen/Strep.

**RT-PCR Analysis**

For RT-qPCR analysis, cDNA was prepared using High-Capacity cDNA Reverse Transcription Kit (Thermo Fisher Scientific) according to manufacturer’s instructions. Real-time PCR was performed using a CFX96 RT-qPCR system and SYBR Green Supermix (BioRad). Comparative analysis using delta-delta Ct method was performed between the gene of interest and the housekeeping gene *Gapdh/GAPDH*. Primers used to analyze gene expression are listed below:

<table>
<thead>
<tr>
<th>Name</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mouse RT-PCR/qPCR primers</strong></td>
<td></td>
</tr>
<tr>
<td><em>Axin2</em></td>
<td>F: 5’-TGACTCTCCTTCCAGATCCCA-3’&lt;br&gt;R: 5’-TGCCCCACACTAGGCTGACA-3’</td>
</tr>
<tr>
<td><em>Ccnd1</em></td>
<td>F: 5’-CGCACAACGCACTTTTCTTC-3’&lt;br&gt;R: 5’-AGACCAGCCTCTTCCACCCAC-3’</td>
</tr>
<tr>
<td><em>Gapdh</em></td>
<td>F: 5’-CCTGCACAAGTATAGATGAC-3’&lt;br&gt;R: 5’-GGAGTGTGCTGTGAAGTCT-3’</td>
</tr>
<tr>
<td><em>Hey1</em></td>
<td>F: 5’-CAGGAGGGAAAGTTATTTGACG-3’&lt;br&gt;R: 5’-TAGTTGTTGAGATGGGAGACCAGC-3’</td>
</tr>
<tr>
<td><em>Krt8</em></td>
<td>F: <em>Sequence by collaborator</em>&lt;br&gt;R: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td><em>Krt14</em></td>
<td>F: <em>Sequence by collaborator</em>&lt;br&gt;R: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td><em>Krt19</em></td>
<td>F: 5’-GCTGCAGATGACTTCAGAACC-3’&lt;br&gt;R: 5’-CTCCTGAGGCAGTAATTTC-3’</td>
</tr>
<tr>
<td><em>Notch1</em></td>
<td>F: 5’-AGGACCCTCATACTCAGACGC-3’&lt;br&gt;R: 5’-TCTTTGTTAGCCCCGTCTTTCAG-3’</td>
</tr>
<tr>
<td><em>Ovol2</em></td>
<td>F: 5’-AATCTCCAGAGCTTACGACG-3’&lt;br&gt;R: 5’-CTGGGAGGCGCCTTTGTTAC-3’</td>
</tr>
<tr>
<td><em>Zeb1</em></td>
<td>F: 5’-ACCGGCCTATTTATCCTGAG-3’&lt;br&gt;R: 5’-CATCTGGGTGTCCCGTTTTCCATCA-3’</td>
</tr>
<tr>
<td>Name</td>
<td>Sequence</td>
</tr>
<tr>
<td>------------------</td>
<td>--------------------------------------------------------------------------</td>
</tr>
<tr>
<td>Human RT-PCR/qPCR primers</td>
<td></td>
</tr>
<tr>
<td><strong>CCND1</strong></td>
<td>F: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td></td>
<td>R: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td><strong>GAPDH</strong></td>
<td>F: <code>5'-GGACCTGACCTGCCGTCTAGAA-3’</code></td>
</tr>
<tr>
<td></td>
<td>R: <code>5'-GGTGTCGCTGTTGAAGTCAGAG-3’</code></td>
</tr>
<tr>
<td><strong>KRT8</strong></td>
<td>F: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td></td>
<td>R: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td><strong>KRT14</strong></td>
<td>F: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td></td>
<td>R: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td><strong>KRT19</strong></td>
<td>F: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td></td>
<td>R: <em>Sequence by collaborator</em></td>
</tr>
<tr>
<td><strong>OVOL2</strong></td>
<td>F: <code>5'-AGCTGTGACCTGTGGCAAG-3’</code></td>
</tr>
<tr>
<td></td>
<td>R: <code>5'-ACGAATGCCTGTGGTCG-3’</code></td>
</tr>
<tr>
<td><strong>ZEB1</strong></td>
<td>F: <code>5'-TTGCTCCCTGTGCAAGTTACA-3’</code></td>
</tr>
<tr>
<td></td>
<td>R: <code>5'-CGTTTCTTGTGAGTTTGGGCA-3’</code></td>
</tr>
</tbody>
</table>

**Western Blot Analysis**

To detect Zeb1 expression after shRNA-mediated knockdown, and ZEB1 and YAP1 expression in human breast cancer cell line MDA-MB-231, cytoplasmic and nuclear protein extracts were isolated as previously described (Wells et al., 2009). Protein concentrations were quantified using BioRad Protein Assay Reagent (BioRad Inc.) and equal amounts were loaded onto 10% polyacrylamide gels to be analyzed by SDS-PAGE. The gels were transferred by electro-blotting onto nitrocellulose membranes (Millipore) and probed with the following antibodies:

- anti-Zeb1 (Rabbit; 1:250; Novus Biologicals, NBPI-88845);
- anti-Zeb1 (Rabbit; 1:250; Santa Cruz Biotechnology, h-102);
- anti-YAP1 (Rabbit; 1:500; Source).

**Immunocytochemistry**
Immunofluorescence imaging was performed on freshly frozen mammary glands that were embedded in Optimum Cutting Temperature (OCT) Compound (Tissue-Tek) and stained following the previously described immunofluorescence protocol (Lee et al., 2014). Briefly, either sectioned tissue or cells grown in 8-well chambered slides (Nunc LabTek II) to optimal cell density were fixed with 4% paraformaldehyde and made permeable with 0.2% Triton-X-100. After washing and blocking with 5% milk in phosphate-buffered saline, cells were stained overnight at 4°C with antibodies specific for Zeb1 (Rabbit; 1:250), Acta2 (Rabbit; 1:500), or Krt14 (1:500). Cells were subsequently labeled with secondary antibodies conjugated with Alexa Fluor-488 (1:500) or Alexa Fluor-568 (1:500). Counter staining was carried out using DAPI.

**Statistical Analysis**

p-values were determined using two-tailed student t-test between two sample groups unless otherwise indicated. Statistical significance of gene expression overlap using GSEA gene signatures was determined using http://nemates.org/MA/progs/overlap_stats.html. Heat-map statistics were generated using Broad Institute Morpheus software.
CHAPTER 3:

GENOME-WIDE ASSESSMENT OF OVOL2 BINDING TARGETS REVEALS ITS FUNCTION AS A GATEKEEPER OF EPITHELIAL IDENTITY

3.1 ABSTRACT OF CHAPTER

The concept describing that epithelial cells exhibit remarkable plasticity is demonstrated by these cells’ capacity to undergo EMT to undergo biochemical changes that influence the production of a mesenchymal-like cellular identity. While milder variations of the EMT phenotype are observed in both developmental and regenerative stages of epithelial tissues, such as the mammary gland and the epidermis, questions pertaining as to how epithelial identity is maintained and governed throughout these critical processes in these tissues remain largely elusive. My data contributes to the novel finding showing transcriptional repressor Ovol2 to function as an important guardian of epithelial identity in stem/progenitor cells of both the MG and basal epidermis. Specifically, using ChIP-seq and ChIP-PCR, I found Ovol2 to significantly bind directly to the gene regulatory regions of EMT transcriptional drivers and its downstream effectors that affect cellular junctions and cytoskeletal organization. Among these targets, Ovol2 shows strongest binding to Zeb1, a gene encoding a potent EMT-TF that is also highly implicated in many aggressive forms of cancer. Importantly, my findings provide a molecular mechanism for the function of Ovol2 as an important mediator of proper ductal morphogenesis and regeneration in the mammary gland, and epidermal differentiation in the epidermis.
3.2 INTRODUCTION

Epithelial cells undergoing developmental and regenerative events can reversibly change their phenotype to support formation of new tissues (Nieto, 2013). EMT is a naturally occurring manifestation of cellular plasticity that characterizes the conversion of epithelial cells to a mesenchymal identity (Kalluri et al., 2013; Nieto et al., 2016). Complete EMT is fundamental to many processes, including the formation of the mesoderm and neural crest during embryogenesis (Alberga et al., 1991; Thiery et al., 2006), while a partial, reversible onset of this program is critically implicated in proper wound healing and other regenerative events (Lamouille et al., 2014). Interestingly, recent findings, many of which have been carried out in vitro, also uncover a striking parallel link between the partial onset of EMT and acquisition of SC-like traits, such as multi-lineage differentiation potential and the capacity to self-renew (Chaffer et al., 2013; Mani et al. 2008; Morel et al. 2000). Expectedly, tight control of EMT is important for proper organismal development, as the invasive properties engendered from aberrant onset of this program are believed to be major contributors to progression and metastasis of various cancers (Thiery et al., 2009; Chakrabarti et al., 2012; Li et al., 2016). Thus, EMT underscores the spectacular capacity of epithelial cells to enhance their lineage plasticity, which prompts questions as to what mechanisms regulate and maintain epithelial identity.

Epithelial cells that pass through a partial or complete EMT undergo dramatic morphogenic cellular changes that are induced by the activity of a collection of EMT-TFs. Master transcriptional drivers of EMT include members from the ZEB (Zeb1 and Zeb2), SNAIL (Snail, Slug), and TWIST (Twist1 and Twist2) families, which directly regulate expression of genes that govern epithelial and mesenchymal states. Hallmark epithelial features affected by progression of EMT are intercellular junctions and apico-basal polarity, which are progressively
shed to produce a motile and invasive mesenchymal-like cellular entity. In addition to up-regulated expression of EMT-TFs, progression of EMT is traditionally characterized by down-regulation of calcium-dependent cell adhesion protein and tumor suppressor E-cadherin (Cdh1), and up-regulated expression of mesenchymal intermediate filament vimentin (Vim) and N-cadherin (Cdh2). While several EMT regulatory networks specific aspects of this cellular process have been identified (Bracken et al., 2008; Siemens et al., 2011; Lu et al., 2013), it is yet to be determined whether a master EMT regulatory program exists to regulate expression of both transcriptional drivers and downstream EMT effectors.

Our recent studies provide evidence that highlights the importance in maintaining epithelial identity of epithelial tissues undergoing morphogenic events, and together with the investigations of others, uncover transcriptional regulators belonging to the Ovol family as critical guardians of this cellular state during normal development as well as in the context of cancer (Lee et al., 2014; Watanabe et al., 2014; Roca et al., 2013). Ovol proteins belong to a family of evolutionarily conserved zinc finger-containing TFs that act as downstream of Wnt and TGF-beta/BMP developmental signaling pathways in embryonic and adult epithelial tissues (Lee et al., 2014; Watanabe et al., 2014; Zhang et al., 2013; Chizaki et al., 2011; Unezaki et al., 2007). Within mammalian systems, Ovol1 and Ovol2 are the most thoroughly studied of three homologs (Ovol1, Ovol2, Ovol3), functioning predominantly as transcriptional repressors through the presence of a SNAG domain, a component described to be present in several positive EMT regulators (Dai et al., 1998; Wells et al., 2009; Teng et al., 2007).

Function of Ovol1 and Ovol2 has begun to be elucidated within epithelial tissues that include the MG and skin epidermis (Lee et al., 2014; Watanabe et al., 2014; Zhang et al., 2013; Chizaki et al., 2011; Unezaki et al., 2007). Within the developing epidermis and during stages of
re-epithelialization in wound repair, tight control of a partial EMT is necessary to restrict plasticity and stemness of the stem/progenitor basal cell population and to allow these cells to undergo proper terminal differentiation (Lee et al., 2014; Turner et al., 2006; Savagner et al., 2005). Activity of EMT driver Slug (Snai2) has been shown to be necessary in development and maintenance of the continuously-renewing tissue, specifically by regulating the proliferative propensity of keratinocytes (Savagner et al., 2005). In the developing epidermis, Ovol1 and Ovol2 function in a compensatory manner to enable proper formation of the impermeable epidermal barrier. Studies utilizing an SSKO mouse model of Ovol1 and Ovol2 reveal that double knockout of these transcriptional regulators results in an expanded number of stem/progenitor basal keratinocytes, resulting in part from up-regulated expression of EMT-TFs that include Snai2 and Zeb1, as well as disruption to cell-cell adhesion in a manner that is accompanied by a defective capacity of basal keratinocytes to terminally differentiate (Lee et al., 2014). Within this developing tissue, failure to regulate expression of EMT-associated genes leads to aberrations in the SC state.

In contrast, during post-natal development of the pubescent MG, a transient, partial EMT occurs during ductal morphogenesis of the mammary epithelium (Ewald et al., 2008; Simian et al., 2001). Ductal elongation and branching depend on a population of basal/myoepithelial cap cells that reside on the bulbous TEBs of the emerging ducts that proliferate and direct a tightly coordinated partial EMT phenotype. A partial EMT permits collective migration of the epithelium by influencing expression of genes that determine epithelial adhesion and cellular polarity (Nanba et al., 2001; Hinck and Silberstein, 2005; Kouros-Mehr and Werb, 2006; Ewald et al., 2012). Ovol2 uniquely plays a significant role within the developing mouse breast, as ablation of this TF results in a severe ductal elongation defect during post-pubertal mammary
gland development (Watanabe et al., 2014). Investigations on the TEBs show that loss of Ovol2 results in abnormal onset of EMT that leads to trans-differentiation of mammary epithelial cells to fibroblast-like cells in vivo (Watanabe et al., 2014). Simultaneously, in vivo transplantation studies performed at limiting dilution reveal that ablation of Ovol2 results in an inability to form the ductal network compared to wild-type controls, suggesting that aberrant EMT compromises SC potential in development and regeneration.

The findings prompt interesting questions as to how Ovol proteins maintain epithelial identity of these epithelial tissues. Epithelial cell culture systems offer an opportunity to dissect molecular mechanisms involved in regulating various processes pertinent to skin and mammary epithelial biology. Primary keratinocyte cultures provide a useful model to decipher gene regulatory mechanisms in skin epidermis. Moreover, specific to the MG, normal mammary epithelial cell line HC11, derived from BALB/c mice at mid-pregnancy (COMMA-1D), has been widely used to study various aspects of mammary epithelial biology, including cell proliferation, signal transduction, and differentiation, primarily for these cells’ display of characteristics associated with both luminal and basal/myoepithelial fates (Humphreys et al., 1997; Williams et al., 2009). Through this chapter, I first perform experiments coupling chromatin immunoprecipitation with high-throughput sequencing to identify the potential regulatory targets of Ovol2 within mammary epithelial model HC11. I next performed ChIP-PCR using an immortalized mouse mammary epithelial cell line (Watanabe et al., 2014) and cultured keratinocytes to validate Ovol2 binding to select targets. Together, my results identify Ovol2 as a direct transcriptional repressor of EMT with a role in protecting epithelial identity in mammary epithelial cells, and highlight the conservation of molecular Ovol function in both mammary and skin epithelial tissues.
3.3 RESULTS

3.3.1 COMPARATIVE ANALYSIS OF MOLECULAR WEIGHT SUGGESTS HC11 MAMMARY EPITHELIAL CELLS EXPRESS THE TRANSCRIPTIONAL REPRESSOR OF OVOL2

Our previous work has identified Ovol2 as a critical transcriptional regulator responsible for maintaining epithelial identity in cells of the mammary epithelium (Watanabe et al., 2014). To gain insight into the molecular mechanisms involved in this process, I employed well-established mammary epithelial cell model HC11, for its display of properties associated with both basal and luminal epithelial lineages (Humphreys et al., 1997). I began assessing the functional role of Ovol2 in the mammary epithelium by first identifying whether the Ovol2 isoform expressed in HC11 cells recapitulated the isoform expressed in mammary tissues (Watanabe et al., 2014). In both human and mouse, the genetic locus of Ovol2 can produce a total of three transcripts, which differ on the presence or absence of a sequence encoding the SNAG (Snail1/GFI) repressor domain on the amino (N)-terminus of the nascent polypeptide chain (Lin et al., 2002). The longest isoform of Ovol2, encoding a transcriptional repressor, is the predominant isoform detected in epithelial tissues that include the epidermis and mammary epithelium. Indeed, my Western blot analysis shows that only the longest isoform, of approximately 37kDa, is detectable in HC11 nuclear protein extracts at varied amounts (FIGURE 3.1).

3.3.2 GENOME-WIDE ANALYSIS IDENTIFIES A MYRIAD OF EMT-RELATED GENES TO BE BINDING TARGETS OF OVOL2
To understand the molecular basis of Ovol2 function, we carried out a genome-wide assessment of Ovol2-binding targets in HC11 cells by coupling chromatin immunoprecipitation with high-throughput sequencing (ChIP-seq) (FIGURE 3.2). Briefly, chromatin from HC11 cells was harvested and cross-linked after culturing cells under proliferative conditions to near confluency, and sheared to appropriate size (for details, refer to CHAPTER 2: MATERIALS AND METHODS). We applied a computational algorithm created by Model-Based Analysis for ChIP-seq (MACS) (Zhang et al., 2008) to identify genomic regions enriched by Ovol2 over control (IgG) experiments. Utilizing this method, we identified 3,092 total Ovol2-binding sites, of which 1,328 sites exhibited less than a 5% false discovery rate (FDR). The distribution of Ovol2 binding peaks was highly specific to gene promoters, as 29.0% of peaks were restricted within 3000bp of the transcriptional start site (TSS) (25.8% of peaks fell within 1kb of the TSS) and 6.4% of peaks were associated with the 5′ untranslated region (UTR) (FIGURE 3.2A, B).

Previous molecular studies carried out using keratinocytes used Cyclic Amplification of Selected Targets (CAST) as an unbiased approach to identify the consensus binding sequence of Ovol1 and Ovol2 (Nair et al., 2007; Wells et al., 2009). We analyzed the ChIP-seq peak regions identified by MACS with Multiple EM for Motif Elicitation (MEME-ChIP) to discover motifs. Our de novo motif analysis of ChIP-seq binding peaks corroborated previous findings, as we identified 5′-CCGTTA-3′ as the primary sequence recognized by Ovol2 (Z-score 31.791; P-value 1x10^{-30}) (FIGURE 3.2C). Remarkably, 1,277 of the 1,328 (96.2%) identified peaks with <5%FDR possessed this consensus sequence, suggesting the high specificity of this transcription factor.

We further associated each Ovol2 binding peak with the nearest gene using the Peak-to-Gene script from Galaxy/Cistrome (for details, refer to CHAPTER 2: MATERIALS AND
and validated binding to previously identified Ovol2 gene targets, including *Myc*, *Id1*, *Id2*, and *Ovol1* (FIGURE 3.3A). To our surprise, amongst the top gene targets bound by Ovol2 were those associated with EMT and those related to metastasis, most prominently including genes encoding potent EMT-TFs *Zeb1*, *Zeb2*, and *Twist1*, EMT-promoting cytokine *Tgfb3*, and EMT-associated intermediate filament *vimentin* (FIGURE 3.3B and FIGURE 3.3C). EMT-related Ovol2 gene targets were further validated through ChIP-PCR using an immortalized mouse mammary epithelial cell line (iMMEC) that was manipulated genetically to delete *Ovol2* after induction of Cre recombinase (Watanabe et al., 2014). (FIGURE 3.3C). Moreover, Gene Ontology (GO) analysis of the Ovol2-bound gene list (<5% FDR) identified greatest significant enrichment of genes associated with EMT/metastasis, as well as cardiac differentiation/muscle contraction, neuronal differentiation, and glycoprotein synthesis/intracellular trafficking (TABLE 3.2).

To determine whether genes containing Ovol2-bound regions (binding peak located within 2kb of TSS; 770) are indeed regulated by Ovol2, we followed our analysis by associating expression of this ChIP-seq-generated gene list with microarray data from control and Ovol2 SSKO TEBs. Our analysis implies remarkable fidelity of this transcriptional regulator to function as a transcriptional repressor, as Gene Set Enrichment Analysis (GSEA) shows significant up-regulation of this set of genes (P-value < 3x10⁻⁶) after *Ovol2* deletion (FIGURE 3.3D). Together, these findings suggest that Ovol2 functions in mammary epithelial development at least in part by directly repressing expression of a collection of EMT genes.

### 3.3.3 OVOL2 BINDS TO EMT-RELATED GENES IN EPIDERMAL KERATINOCYTES

In the developing skin, loss of Ovol proteins leads to up-regulation of genes associated with EMT and stemness, which lead to a defective and permeable protective barrier (Lee et al.,
To determine whether Ovol2 also directly regulates EMT genes in skin epidermis, I performed chromatin immunoprecipitation coupled with PCR (ChIP-PCR) on primary wild-type keratinocytes taken from newborn C57BL/6J mice to determine whether Ovol proteins directly binds to promoters of genes involved in EMT. Because of the functional redundancy of Ovol1 and Ovol2, and our knowledge that Ovol2 is normally expressed within basal stem/progenitor keratinocytes, we sought to identify specific binding targets of Ovol2 in keratinocytes cultured under a low calcium condition. Utilizing the identified and validated Ovol2 binding sites that were generated by HC11 ChIP-seq as a reference (FIGURE 3.4A), we reveal Ovol2 to bind to the promoter sequences of Zeb1, Vim, Snai2, and Myc (FIGURE 3.4B). Interestingly, skin-specific binding of Ovol2 was also detected, as binding to the Trp63 (Delta-N p63) promoter was detected at consensus motifs only in keratinocytes but not in mammary epithelial cells. Taken together, these findings associate a molecular mechanism to the pleiotropic cellular effects of Ovol transcriptional regulators in governing epithelial differentiation by inhibiting cellular plasticity.

3.4 DISCUSSION

The findings described in this chapter highlight a novel molecular mechanism through which transcriptional regulator Ovol2 functions to maintain epithelial identity and to regulate plasticity of cell populations within mammary and skin epithelial tissues. Primarily, this is carried out through tight regulation of EMT, a program which is greatly implicated in development and regeneration of epithelial tissues. Within the developing pubertal mouse MG, a partial-EMT in the TEBs enables cohesive migration of the elongating mammary ducts (Ewald et al., 2012). Expression of transcriptional drivers of EMT govern up-regulation of genes that
encode ECM-modifying enzymes, such as matrix metalloproteases (MMPs) (Nieto et al., 2016; Kalluri et al., 2009), which are associated with enabling the emergent epithelial buds to penetrate and migrate through the surrounding fatty stroma. Simultaneously, recent findings have also identified an interesting link between a partial EMT and cancer SC-like characteristics (Guo et al., 2012; Mani et al., 2008; Morel et al., 2008; Chaffer et al., 2013; Nieto et al., 2016), suggesting not only the pleiotropic function of EMT-TFs, but also the intriguing possibility that progression of EMT may also regulate cellular plasticity that is yet to be elucidated. Importantly, the critical importance in tightly regulating the expression of these EMT factors is shown to be necessary to preventing aberrant progression of this program that may lead to the progression of cancer. Indeed, clinical studies show expression of Ovol2 and of genes associated with maintaining epithelial identity display greater favorable outcomes in various carcinomas (Fu et al., 2017; Roca et al., 2014; Roca et al., 2013), supporting the idea that regulation of EMT-mediated invasiveness contributes to cancer progression.

My genome-wide Ovol2 binding data provides mechanistic insight into our in vivo analysis of Ovol2 function in the pubertal MG (Watanabe et al., 2014), and identifies transcriptional repressor Ovol2 as a master negative regulator of various genes involved in progression of EMT, such as $Snai1$, $Snai2$, $Vim$, and $Zeb1$. Amongst these targets, Snail ($Snai1$) and Slug ($Snai2$) are EMT-TFs which have been shown to promote onset of EMT through distinct mechanisms that regulate unique gene targets responsible for controlling mammary SC and circulating tumor cell (CTC) activity (Ye et al., 2016). Moreover, vimentin ($Vim$) is an mesenchymal intermediate filament that is not only a structural protein, but also functional within the nucleus to regulate EMT (Mergui et al., 2010), although, its role within the MG has not been studied. Interestingly, recent studies identify Zeb1 as a potent mediator of EMT with a
critical function in the conversion of non-SCs to that of a SC-like state in breast cancer (BrCa) models (Chaffer et al., 2013), which highlight the importance of Ovol2 repression of this EMT-TF. Interestingly, Zeb1 was the most strongly bound target by Ovol2, and its upregulation had been shown to be partly responsible for the defective ductal morphogenesis phenotype associated with Ovol2 deletion (Watanabe et al. 2014). Using ChIP-PCR, I also found Zeb1 to bind directly to the promoter of Ovol2, raising the possibility that, if EMT governs cellular plasticity in a parallel manner to epithelial plasticity, a tight balance between EMT-TF Zeb1 and EMT-inhibitor Ovol2 may engender a partial EMT that is amenable to a SC-like state. Intriguingly, using a systems biology approach, we recently uncovered that a mutual inhibitory circuit between Ovol2 and Zeb1 governs bi-directional differentiation capacity of normal mammary epithelial cells (MCF10A) toward both epithelial and mesenchymal terminal states (Hong et al., 2015).

While several cross-regulatory gene networks that are responsible for regulating SC activity have been described in other settings, one wonders whether involvement of an EMT-regulatory network that is responsible for maintaining an intermediate cellular identity between epithelial and mesenchymal states can also be compatible with a stem cell fate. This poses questions as to what function EMT regulators, such as Zeb1, have in normal mammary gland development and regeneration. Moreover, what is the physiological function of an Ovol2-Zeb1 mutual inhibitory loop in mammary epithelial cells in vivo? The analysis described in the following chapter introduces findings that begin to decipher these questions.
FIGURE 3.1 The Ovol2 isoform, with identified transcriptional repressor function, is the exclusive isoform produced in HC11 mammary epithelial cells. (A) Detection of Ovol2 through Western Blot using 15 and 20 ug of nuclear protein extract from HC11 cells.
FIGURE 3.2 Genome-wide analysis identifies Ovol2 to function as a transcriptional repressor that binds to promoter-proximal gene regions with high specificity. (A) Distribution of Ovol2-binding regions on the genome relative to RefSeq gene loci annotation. Intergenic region refers to all locations that are not categorized as promoter regions, 5’UTR, exons, introns, 3’UTR, or downstream regions of each RefSeq gene. (B) Enrichment of Ovol2 peaks reside in gene promoter region that lie <1kb from TSS and on 5’UTR. P-value represents enrichment over random genomic distribution of peaks. (C) Presence of Ovol2-binding motif is present in detected Ovol2 ChIPseq peaks.
FIGURE 3.3 Genome-wide analysis identifies a myriad of EMT-related genes to be binding targets of Ovol2 (A) Visualization of ChIPseq peaks for the indicated genes. Ovol2 binding peaks (red boxes) were visualized on the University of California, Santa Cruz (UCSC) genome browser for (i) known targets and (ii) EMT-related genes. Gapdh is shown as a random example of a gene that does not exhibit an Ovol2 binding peak. (B) ChIP-PCR using immortalized mouse mammary epithelial cells (iMMEC) reveals Ovol2 directly binds to the promoter of a myriad of EMT-transcriptional regulators as well as down-stream effectors by binding to genomic regions containing 5’-CCGTTA-3’ binding motif. The Gapdh promoter is used as a negative binding control. (D) Gene Set Enrichment Analysis (GSEA) performed on the TEB microarray data with Ovol2 binding gene set showing significant correlation between Ovol2 binding gene targets and up-regulated genes resulting from Ovol2 deletion.
Figure 4. ChIP-Seq and ChIP-PCR Analyses Identify EMT Genes as Ovol2 Targets

(A) Distribution of Ovol2-binding regions on the genome relative to RefSeq genes. Intergenic region refers to all locations other than promoter, 5′ UTR, exon, intron, 3′ UTR, or downstream.

(B) Enrichment of Ovol2 binding at promoter (<1 kb) and 5′ UTR compared with random genomic distribution.

(C) Presence of Ovol-binding motif within the ChIP-seq peaks.

(D) GSEA performed on the TEB microarray data with Ovol2-bound gene set.

(E) Western blot analysis of Ovol2 with or without Cre induction in iMMEC(N). Lysates from 293T cells transiently expressing the indicated Ovol2 isoforms were run as size references. EV, empty vector. Asterisk indicates a previously unrecognized Ovol2 isoform the structure and function of which are currently unknown. LaminA/C was used as a loading control.

(F) ChIP-PCR. Note absence of ChIP signals in Ovol2-deficient iMMEC(N) cells. Gapdh promoter was used as a negative control. Values on the right of the gels represent percentage of input of Ovol2 occupancy obtained for iMMEC(N)+EV.

See also Figures S4 and S5 and Table S2.
TABLE 3.1 GO analysis reveals the top candidate processes associated with Ovol2-bound genes.
FIGURE 3.4 Ovol2 directly represses multiple downstream targets genes involved with EMT and stemness. (A) Positioning of Ovol2-binding region on the promoter loci of various stemness/proliferation genes. (B). ChIP-PCR using primary mouse keratinocytes reveals Ovol2 binds to the promoter of Myc, Trp63, Zeb1, Vimentin, and Snai2, at specific genomic regions containing 5’-CCGTTA-3’ binding motif, but not on control regions on the Trp63 promoter or random region on Gapdh.
A

-1139

-1812 -31

+103 +286 +338

-749 -41

+83

Myc

Trp63

Zeb1

Vim

Snai2

B

Fold over IgG

Gapdh  Myc  ΔNp63 +31

Trp63  Zeb1  Vim  Snai2  Trp63-c

Input  Ovol2  IgG
CHAPTER 4:

AN OVOL2-ZEB1 EMT REGULATORY CIRCUIT REGULATES MAMMARY BASAL-LUMINAL BINARY DIFFERENTIATION

4.1 ABSTRACT OF CHAPTER

EMT is a recognized form of cell fate plasticity that converts epithelial cells (i.e., epiblast) into mesenchymal cell types (i.e., primitive mesenchyme) during embryogenesis. EMT-like processes are also strongly implicated in cancer metastasis and wound healing. Recent studies, mostly utilizing cultured cell lines, uncover an interesting parallel between EMT and stemness: differentiated epithelial cells that undergo a transient, partial-EMT, acquire SC-like traits that include multipotency and capacity to self-renew, suggesting transcriptional regulators of EMT may govern stemness through mechanisms that are parallel to EMT. Despite emerging data supporting the physiological relevance of this finding, questions on whether the dynamic interplay of EMT-promoting and -inhibiting factors influence SC fate during epithelial morphogenesis and regeneration is yet to be elucidated. We recently identified Ovol2 as a master gatekeeper of epithelial identity that carries out its function by regulating expression of a myriad of EMT-related genes, most notably that of potent EMT-inducing EMT-TF Zeb1. Utilizing a systems biology approach, we uncovered the importance of a mutual inhibitory loop between Ovol2 and Zeb1 for the robust production of a plastic, partial-EMT state, that exhibits propensity to differentiate into cells of both epithelial and mesenchymal lineages. In this study, we report the SC-containing basal compartment of the mammary epithelium, but not the luminal compartment, to exist in a partial EMT state that is regulated by the activity of Zeb1 and Ovol2.
Utilizing both genetic knockout models and a lentiviral knockdown system, I provide evidence suggesting Ovol2 promotes mammary epithelial cells to adopt a luminal fate, whereas Zeb1 promotes determination toward a basal/myoepithelial lineage. In addition to protecting precocious differentiation of basal epithelial cells toward a luminal fate, I ascribe Zeb1 with a novel function in supporting basal SC self-renewal/proliferation; a role that might be important in regulating ductal branching during mammary regenerative/morphogenic processes. Together, my results support a model where positive and negative regulation of EMT controls the binary basal-luminal decision within the differentiating mammary stem/progenitor cell populations, and provide evidence of EMT-TF Zeb1 to have a functional role in epithelial tissues.

4.2 INTRODUCTION

Emerging evidence suggests partial onset of EMT is relevant for producing a plastic cellular state that confers differentiated epithelial cells with SC-like characteristics (Nieto et al., 2016). In both normal and neoplastic settings, transient expression of EMT-TFs from the SNAIL, ZEB, and TWIST families engender programs that heighten proliferative potential and expand the capacity to differentiate in vitro (Mani et al., 2008; Guo et al., 2012; Chaffer et al., 2013). This observation raises the possibility that the SC state is not a fixed cellular entity but rather a transient state that exists through the dynamic interplay of tightly guarded EMT-regulatory networks responding to various forms of stimuli (Nieto et al., 2012). Many EMT-regulatory networks have been identified, which also integrate the regulatory function of microRNAs (miRNAs) to modulate expression of specific mediators of this cellular process (Nieto et al., 2013; Zhang et al., 2012; Bracken et al., 2008; Gregory et al., 2008; Park et al., 2008). Interestingly, recent investigations uncover proteins from the Ovol family to function as
transcriptional gatekeepers of epithelial identity and plasticity within developing skin and mammary epithelial tissues, specifically by negatively regulating the expression of a multitude of EMT-TFs and downstream effectors (Lee et al., 2014; Watanabe et al., 2014; Roca et al., 2014).

Among this gene family, Ovol2 was found to have a critical role in MG development and regeneration, as abrogation of this transcriptional repressor within the mammary epithelium led to severe defects in ductal elongation that likely result from trans-differentiation of mammary epithelial stem/progenitor cells to a mesenchymal-like phenotype (Watanabe et al., 2014). Detailed molecular analysis of Ovol2-null mutants revealed that Ovol2 functions as a direct master regulator of a myriad EMT-TFs or other EMT-related genes, that include Snai1, Snai2, Twist1, Vim, and most prominently Zeb1 (Watanabe et al., 2014). EMT-TF Zeb1 is a notable positive regulator of EMT whose effects on producing SC-like traits in a pathological context have been well-documented (Lehmann et al., 2016; Chaffer et al., 2013; Vandewalle et al., 2008). The regulation of Zeb1 by Ovol2 is functionally relevant, as knockdown of the Zeb1 protein in Ovol2-null glands partially rescued the ductal elongation defects (Watanabe et al., 2014). Additionally, the relevance in maintaining balanced control of Zeb1 and Ovol2 was shown by taking a systems biology approach using mammary epithelial cells, which described a cross-regulatory circuit between Zeb1 and Ovol2 that is necessary for producing a plastic, partial-EMT state with propensity to differentiate into cells of both epithelial and mesenchymal lineages (Hong et al., 2015). These findings prompt important questions pertaining to what role EMT-promoting factors have on governing the epithelial SC state, and whether an EMT cross-regulatory circuit between Zeb1 and Ovol2 has relevance in mammary epithelial SCs in vivo.

The MG serves as an outstanding system to study regulation of epithelial SC fate plasticity for its simple bi-lineage composition. Specifically, the mammary epithelium is
composed of an inner population of luminal epithelial cells that are capable of terminally differentiating into milk-producing cells, and an outer, epithelioid basal population that possesses both epithelial and mesenchymal characteristics (Prater et al., 2016; Visvader et al., 2014). Substantial progress has been made to identify the mammary epithelial SC populations that support the dynamic, morphological changes associated with post-natal MG development (Shackleton et al., 2006; Stingl et al., 2006; Visvader et al., 2014). Of the two lineages, the basal population contains mammary SCs, which are functionally defined by their ability to generate a complete mammary epithelial network composed of both basal and luminal lineages upon transplantation (Visvader et al., 2014; Rios et al., 2014; Van Keymeulen et al., 2011). While there is some controversy on whether bipotent SCs exist under physiological conditions, existing evidence suggests that both bipotent and unipotent SC populations are likely to exist, with environmental conditions determining which population drives morphogenesis (Prater et al., 2014).

During pubertal and pregnancy-associated stages of post-natal MG morphogenesis, partial EMT occurs during the cohesive migration of the elongating ductal network and linked to SC activity (Oakes et al., 2008; Chakrabarti et al., 2012a,b; Ye et al., 2015). Whether and how an EMT-regulatory network between Zeb1 and Ovol2 plays any role in governing stemness and plasticity of adult basal SCs is the focus of the experiments described in this chapter. In this study, I use a regeneration model to investigate how regulation of EMT, specifically through cross repression of Zeb1 and Ovol2, affects the mammary SC state. Implementing an ex vivo differentiation model, I show how loss of Zeb1 likely promotes basal SC fate determination toward a luminal lineage, while loss of Ovol2 likely promotes these cells to differentiate toward a basal/myoepithelial fate. Specifically, within the basal epithelial cells, I find Zeb1 to be
required for basal SC self-renewal and/or proliferation. Transplantation of Zeb1-deficient basal cells results in defective regenerative capacity, as regenerated mammary trees fail to form a properly branched epithelial ductal network. The in vitro and in vivo phenotypes together are consistent with previous findings that show how compromised basal to luminal epithelial cell ratio affects the extent of mammary ductal branching. Interestingly, my findings show Zeb1 promotes SC-activity in basal cells possibly by activating the expression of two Wnt downstream targets, Ccnd1 and Axin2. Positive regulation of CCND1 may be mediated by the interaction of ZEB1 with YAP1, as both proteins bind to the same promoter region.

This body of work describes provides strong evidence for the importance of regulating EMT in mammary basal cells for maintenance of epithelial plasticity and a SC-state during stages of regeneration, and suggests that regulation of EMT may also play a role in governing adult epithelial cellular plasticity. Specifically, my findings suggest a model that regulation of EMT is mediated through cross inhibition of antagonistic transcriptional regulators Zeb1 and Ovol2, which are necessary in specifying mammary basal-luminal binary differentiation.

4.3 RESULTS

4.3.1 ZEB1 EXPRESSION IS ACTIVATED IN THE BASAL COMPARTMENT OF MAMMARY EPITHELIUM AT TIMES OF ACTIVE MORPHOGENESIS

Previous work suggests mammary epithelial basal cells display enriched expression of genes that are characteristic of having undergone a partial EMT (Prater et al., 2014, Ye et al., 2015). Our comprehensive transcriptome-wide analysis corroborates this notion, as Gene Set Enrichment Analysis (GSEA) of microarray data from both FACS-sorted adult mammary basal and luminal epithelial populations, revealed significantly enriched expression of hallmark EMT
genes within the sorted basal population but not in the luminal population (FIGURE 1A, B; FIGURE S1A- S1C). Of interest, mRNA levels of potent EMT-TF Zeb1 were dramatically higher in basal cells when compared to luminal cells (FIGURE 1C), albeit at a dramatically lower level than in the stroma (data not shown). Conversely, expression levels of Ovol2, a transcriptional repressor of Zeb1 (Lee et al., 2014; Watanabe et al., 2014; Roca et al., 2014), showed an inverse mRNA expression pattern to that of Zeb1 (FIGURE 1C). Together, these data are consistent with the current understanding that adult mammary basal cells exhibit traits characteristic of epithelial cells bearing a partial-EMT state, whereas luminal cells represent a differentiated/differentiating epithelial cell state.

While recent studies uncouple a previously suggested direct, intimate link between onset of EMT and enhanced stemness (Ocaña et al., 2012, Chang et al., 2013, Stankic et al., 2013), new findings suggest this phenomenon likely occurs from pleiotropic functional roles of EMT-TFs that act in parallel to that of EMT (Nieto et al., 2016). These findings engender new interest to defining the unique role specific EMT-TFs have in governing the SC state. Recent in vitro findings reveal a SC-like state could be induced from non-SCs through the epigenetic conversion of the ZEB1 promoter from a “poised”/bivalent configuration to an active state (Chaffer et al., 2013), which is functionally important in generating a more proliferative phenotype. Comprised of both repressive H3K27 tri-methylation (H3K27me3) and permissive H3K4 tri-methylation (H3K4me3) histone modifications, the “poised”/bivalent epigenetic state allows transcriptional activation to be highly responsive to environmental stimuli and is a modification found on the promoter of many genes critical for embryonic development (Bernstein et al., 2006; Voigt et al., 2013). We wondered whether Zeb1 expression may be activated in a similar manner to that of neoplastic conditions during stages of active stages of normal MG development.
We began addressing this question by inquiring how Zeb1 expression changed from virgin levels in response to pregnancy, a developmental phase that is characterized by elevated SC activity that stimulates dramatic expansion and remodeling of the mammary epithelium (Visvader et al., 2014). qRT-PCR analysis of FACS-sorted basal cells taken from age-matched adult virgin and pregnant mice (12 days post coitus, DPC) revealed a substantial 40-fold increase in Zeb1 mRNA expression in basal cells belonging to the pregnant gland (FIGURE 1D). We followed this finding with an assessment of the chromatin landscape at the Zeb1 locus to determine whether epigenetic activation of the Zeb1 promoter underlined the observed transcriptional up-regulation of this gene in the pregnant gland. We isolated chromatin from freshly isolated basal cells taken from adult virgin and pregnant glands at 12 DPC, and performed micro ChIP-qPCR (Dahl et al., 2008, Gu et al., 2010) to describe the distribution of H3K4me3 and H3K27me3 histone modifications at the Zeb1 locus. Both activating and repressive histone marks were present on the Zeb1 promoter in basal cells of the adult virgin (FIGURE 1E). In the pregnant state, however, the abundance of activating H3K4me3 mark showed a 3.5-fold increase, while H3K27me3 remained largely unchanged (FIGURE 1E). Together, these findings suggest the possibility that pregnancy induces chromatin activation at the Zeb1 locus in at least a subset of the basal cells.

Next I employed immunocytochemistry to determine how Zeb1 protein levels were affected at different stages of post-natal MG development and to identify whether Zeb1 expression may be subject to post-transcriptional regulation. We specified stages characterizing various SC states of post-natal MG morphogenesis: active morphogenesis (pubertal and pregnancy-induced), homeostasis (adult virgin), and lactation (Visvader et al., 2014). In all settings observed, Zeb1 protein was detected in the nuclei of cells in the surrounding mammary
stroma, supporting the fidelity of our assay (FIGURE 1F-H; data not shown). Utilizing a-SMA to stain cells of the basal lineage, we readily detected co-localization a-SMA with several scattered Zeb1-positive nuclei in puberty and pregnancy stages, where the mammary epithelium is undergoing active morphogenesis (FIGURE 1F, H). A three-dimensional view of Z-stack images belonging to these pregnant glands further supports the co-existence of Zeb1 in nuclei of a-SMA-positive basal cells (FIGURE S1D). Conversely, while we cannot definitively conclude the absence of Zeb1-positive nuclei in basal cells in the adult virgin or during lactation, we were unable to detect expression of Zeb1 in the a-SMA-positive (FIGURE 1F, H) or cyto-keratin14-positive basal compartment (data not shown) when multiple sections and samples were analyzed. Together, these data suggest Zeb1 exhibits a dynamic expression pattern, where its expression is most abundant and readily detected during stages of active morphogenesis.

4.3.2 ALTERED ZEB1 EXPRESSION RESULTS IN MISREGULATION OF BASAL AND LUMINAL BINARY DIFFERENTIATION

Our previous findings identify a cross regulatory circuit between Zeb1 and Ovol2 in mammary epithelial cells that is necessary for governing a stable, partial-EMT state that exhibits bidirectional propensity to differentiate into epithelial and mesenchymal lineages (Hong et al., 2015). We questioned whether regulation of EMT, specifically through regulation of Zeb1 expression, is necessary for controlling cell fate plasticity (i.e., precocious differentiation) of basal SCs. To begin evaluating the functional relevance of this regulator of EMT within the MG, we utilized a standardized 3D Matrigel differentiation assay (Gu et al., 2013) and infected mammary epithelial cells with lentiviral particles containing a bi-cistronic cassette, co-expressing green fluorescent protein (GFP) and either an shRNA targeting the Zeb1 transcript or
a scrambled control sequence (FIGURE S2 A- C) (Watanabe et al., 2014) to ask whether shRNA-mediated knockdown of Zeb1 influenced cell fate plasticity of unsorted mammary epithelial cells that were cultured ex vivo. As previously reported, differentiation of mammary epithelial cells produces two major, morphologically distinct types of organoids: branched/solid and acinar-like organoids, which are derived from basal stem and luminal/alveolar-restricted progenitor cells, respectively (Dontu et al., 2003; Petersen et al., 1992; Shackleton et al., 2006; Gu et al., 2013). While the total number of GFP+ organoids was not affected by Zeb1 knockdown, depletion of Zeb1 resulted in a significant reduction in the ratio of branched/solid to acinar-like organoids (FIGURE 2A, B), suggesting that basal SCs precociously adopt a luminal fate under differentiating conditions. Expression analysis by qRT-PCR using mRNA harvested from FACS-sorted basal cells that were cultured under differentiation conditions further validated our observation, as in addition to reaffirming transcriptional knockdown of Zeb1 (FIGURE 2C), we observed elevated levels of Krt19 and Notch4 (FIGURE 2C), two genes known to be expressed specifically within luminal cells (Bartek et al., 1985, Bouras et al., 2008, Raafat et al., 2011, Sun et al., 2010). This said, the expression of basal marker Krt14 and luminal marker Krt8 was unchanged (FIGURE S2D, E). These data suggest that Zeb1 plays a cell-intrinsic role in preventing mammary basal cells from acquiring at least some luminal features.

To validate and further examine our findings on the Zeb1-Ovol2 cross-regulatory circuit in regulating basal and luminal fate determination, we turned to mammary epithelial cell line MCF10A, a reliable normal human basal-like cell model that expresses markers of both mammary basal and luminal epithelial lineages (Qu et al., 2015). We ablated ZEB1 expression in MCF10A cells and examined the effect on cytokeratin marker expression, namely basal and transitional/luminal cytokeratin markers K14 and K19, respectively, to determine whether ZEB1
over-expression recapitulated our experimentally observed expression and morphological changes. Upon ZEB1 deletion, MCF10A cells showed significant up-regulation of KRT19 expression, while expression levels of KRT14 decreased (FIGURE 2D, courtesy of Kazuhide Watanabe). We found expression of these cytokeratin markers to be specifically regulated by ZEB1 and not because of EMT progression, as TGF-β, a well-known positive regulator of EMT (Nishimura et al., 2006), was only able to induce expression changes in KRT14 and KRT19 in control MCF10A cells, but not after ZEB1 was deleted (FIGURE 2D, courtesy of Kazuhide Watanabe).

Additionally, we introduced a doxycycline-regulated lentiviral vector system to induce ZEB1 expression in MCF10A (FIGURE S3A, C, courtesy of Kazuhide Watanabe) and stained for cytokeratin markers K14 and K19 to determine if ZEB1 over-expression produced the opposite effects. Immunostaining of untreated MCF10A with these cytokeratin markers revealed a heterogeneous mixture of K14 and K19 single- and double-positive cells (FIGURE 2E, F, courtesy of Kazuhide Watanabe). Remarkably, after as early as 24h post induction, over-expression of ZEB1 led to a significant increase in the number of K14⁺; K19⁻ cells while accompanied with a decrease in number of K14⁺; K19⁺, K14⁺; K19⁻, and K14⁻; K19⁻ cells (FIGURE 2E, F, courtesy of Kazuhide Watanabe). Consistent with our finding, qRT-PCR of these samples revealed increased KRT14 expression and decreased KRT19 expression, while KRT8 expression remained unchanged (FIGURE S3D, courtesy of Kazuhide Watanabe). To obtain a transcriptome-wide assessment of stemness-related changes associated with ZEB1 over-expression, we cultured these MCF10A cells (6 days) and isolated total RNA from these samples for RNAseq. As expected, GSEA found that induced expression of ZEB1 resulted in positive enrichment of EMT hallmark genes (p-value 0.00; FDR q-value 0.00) (FIGURE 2G).
Interestingly, our analysis detected significant enrichment of genes that are normally up-regulated in mammary SCs (p-value 0.00; FDR q-value 1.0x10^{-4}) (FIGURE 2G) and down-regulation of genes that are normally down-regulated in mammary SCs (p-value 0.00; FDR q-value 0.00) (Lim et al., 2010) (FIGURE S3E). Moreover, over-expression of ZEB1 results in down-regulation of gene signatures describing epithelial differentiation modules (p-value 0.03; FDR q-value 0.06) (Bosco et al., 2010) (FIGURE S3E). Collectively, our data support a facilitative role of Zeb1/ZEB1 in basal fate and suppressive role in luminal fate.

4.3.3 ZEB1 DEPLETION IN MAMMARY BASAL CELLS COMPROMISES THEIR ABILITY TO UNDERGO OPTIMAL BRANCHING MORPHOGENESIS UPON TRANSPLANTATION

Recent studies identify mammary basal cell number, or more specifically the ratio of basal to luminal mammary epithelial cells, to be directly correlated to the extent of ductal branching (Macias et al., 2011). Having identified the relevance of the Zeb1-Ovol2 cross regulatory circuit in regulating bidirectional differentiation propensity of mammary epithelial cells, we questioned whether transcriptional knockdown of Zeb1 would compromise the extent of branching in regenerative mammary transplants. To begin evaluating the functional relevance of Zeb1 in mammary basal cells, I employed the mammary fat pad transplantation assay to investigate whether depletion of Zeb1 in the basal population affected the regenerative properties of these cells (FIGURE 3A-D). We infected FACS-sorted basal cells with lentiviral particles containing a bi-cistronic cassette, co-expressing green fluorescent protein (GFP) and either an shRNA targeting the Zeb1 transcript or a scrambled control sequence (FIGURE S3 A, C) and introduced cells into the cleared fat pad of 3-week old C57BL/6J WT recipient mice. While both
control and Zeb1-depleted basal cells produced GFP-positive mammary outgrowths with no detectable difference in length (FIGURE 3B), Zeb1-depleted basal cells generated trees that exhibited significantly less branching than Zeb1-intact cells (FIGURE 3C). A ratio of the number of branching points to total ductal length of GFP-positive outgrowths revealed a near 2.5-fold reduction that resulted from the loss of Zeb1 (FIGURE 3D). With our expression analysis describing dramatic up-regulation of Zeb1/Zeb1 during pregnancy, we also sought to determine whether Zeb1 knockdown disrupted a specific function role of Zeb1 that may initiate at pregnancy. To elucidate this potential role, we mated transplanted host and visualized each gland after 12 DPC to assess morphological defects resulting from depletion of Zeb1 (FIGURE 3E-H). Expectedly, the gland of each host mouse exhibited an apparent increase in the overall extent of branching in response to pregnancy-associated hormonal changes, but, consistent with the findings on the adult virgin host mice, the reduced branching phenotype in GFP-positive Zeb1-depleted transplants persisted (FIGURE 3E-H). Results from these regenerative assays suggest Zeb1 possesses a role that influences ductal branching in regeneration.

4.3.4 OVOL2 SUPPRESSES EMT/BASAL GENES IN MAMMARY BASAL CELLS WHILE PROMOTING THEIR LUMINAL FEATURES

Regulation of epithelial plasticity by cross repression of Zeb1 and Ovol2 prompted us to wonder whether Ovol2 plays a role opposite to that of Zeb1 in mammary basal cells, namely by promoting differentiation of these cells toward a luminal fate. We first determined whether knockout of Ovol2 resulted in specific up-regulation of Zeb1 in the basal population, the presumptive population containing bipotent mammary SCs. qPCR using total RNA taken from FACS-sorted basal, luminal, and stromal populations belonging to WT and Ovol2 SSKO glands
revealed Zeb1 to be significantly up-regulated exclusively within the Ovol2-deficient basal population, while the expression differences observed between WT and SSKO luminal or stromal cells were not statistically significant (FIGURE 4A). During post-natal mammary morphogenesis, basal cap cells of the TEBs are enriched with basal stem/progenitor cells that drive ductal morphogenesis (Ewald et al., 2008). We isolated and immuno-stained TEBs from WT and Ovol2-SSKO glands with basal lineage-specific marker K14 and primitive progenitor marker K6 (Grimm et al., 2006, Stingl et al., 2005). TEBs from Ovol2-SSKO glands showed striking elevated levels of K14 and a modest decrease in K6 expression (FIGURE 4B, courtesy of Kazuhide Watanabe), suggesting possible terminal differentiation of mammary basal SCs toward a basal/myoepithelial cell fate.

To gain further insight into how regulation of EMT, specifically through the balance of Zeb1 and Ovol2, regulates mammary basal fate decisions, we generated Zeb1<sup>loxflox</sup>; Ovol2<sup>floxflox</sup> and Zeb1<sup>loxflox</sup>; Ovol2<sup>floxflox</sup> single mutant and Zeb1<sup>loxflox</sup>; Ovol2<sup>floxflox</sup> double mutant mouse lines. We efficiently deleted Zeb1 and Ovol2 in unsorted mammary epithelial cells by introducing Cre-recombinase through infection with Adeno-viral particles (FIGURE 4C, D), and again induced the organoid colonies to differentiate ex vivo. In support of our findings using the Zeb1-knockdown model, deletion of Zeb1 in these cells resulted in a significant reduction to the number of branched/solid, basal-derived organoids as compared to control (FIGURE 4E), suggesting Zeb1 has a role in inhibiting precocious differentiation toward a luminal fate. Simultaneously, deletion of Ovol2 resulted in the inverse phenotype, showing a significant increase to the ratio of branched/solid to acinar-like organoids (FIGURE 4E). Surprisingly, restoring the equilibrium of the cross regulatory circuit through deletion of both Zeb1 and Ovol2 rescued the imbalance between the number of branched/solid and acinar-like organoid colonies.
back to control levels (FIGURE 4E). As an interesting note, several organoid colonies derived from \(Zeb1^{\text{floxed/floxed}}; Ovol2^{\text{floxed/floxed}}\) exhibited an indiscernible morphology that could not easily be distinguished as either basal- or luminal/alveolar cell derived (FIGURE S4A), suggesting the possibility that deletion of both Zeb1 and Ovol2 may disrupt the initially stages of mammary epithelial differentiation. These findings describe a novel role of EMT regulators in governing the SCs by regulating mammary SC fate determination and plasticity.

We returned to the MCF10A model to ask whether inducible expression of \(OVOL2\) could enhance luminal features but compromise basal competence. We again examined expression lineage specific cytokeratin markers, K14 and K19. Remarkably, \(OVOL2\) over-expression significantly increased the number of K14\(^+\)K19\(^+\) cells at the expense of K14\(^-\)K19\(^-\) and K14\(^-\)K19\(^+\) cells (FIGURE 4F-G, courtesy of Kazuhide Watanabe) – a phenotype that is exact opposite of the changes induced by ZEB1 over-expression. We also followed-up our observation by asking whether inducible expression of \(OVOL2\) in MCF10A would engender global transcriptional changes that contrast the expression changes manifested by ZEB1 (ZEB1-MCF10A). We performed RNA-seq after induction of \(OVOL2\) in MCF10A (OVOL2-MCF10A) and used GSEA to assess expression changes to specific gene signatures. GSEA confirmed \(OVOL2\) as a transcriptional inhibitor of EMT, as we witnessed significant down-regulation of EMT hallmark genes (p-value 0.00; FDR q-value 0.00) (FIGURE 4H). Interestingly, 166 of the 200 EMT hallmark genes exhibited an inverse expression pattern after over-expression of \(OVOL2\) or \(ZEB1\) (p-value 0; representation value 101.7) (FIGURE S4B), suggesting the possibility that this gene subset may be co-regulated by ZEB1 and OVOL2. Importantly, over-expression of \(OVOL2\) resulted in enriched expression of luminal mammary epithelial cell signatures and decreased expression of gene expression signatures associated with basal epithelial cells (FIGURE 4H).
Interrogation of this data found $OVOL2$ over-expression to be inversely related to expression of genes found to be up-regulated in mammary SCs (p-value 0.00; FDR q-value 0.00) and those found to be down-regulated in mammary SCs (p-value 0.00; FDR q-value 0.00) (Lim et al., 2010) (FIGURE 4H). Simultaneously, there is a direct correlation between expression of $OVOL2$ and epithelial differentiation gene signatures (p-value 0; FDR q-value 0) (FIGURE S4C) (Bosco et al., 2010) (FIGURE S4C). Together, these data indicate that Ovol2/OVOL2 is important in basal cells to prevent them from undergoing complete EMT and to preserve their epithelial features.

4.3.5 **ZEB1 IS REQUIRED FOR BASAL CELL PROLIFERATION AND DIRECTLY ACTIVATES CCND1 EXPRESSION**

Our findings showing the involvement of Zeb1 in governing cell fate plasticity prompted us to question if this EMT-TF may also regulate self-renewal/proliferation of basal cells. Utilizing our established Zeb1 knockdown model, we examined the effects of Zeb1 depletion on basal cell clonogenicity after culturing these cells under proliferative conditions (FIGURE 5A). Analysis of GFP-positive colonies derived from control and Zeb1-depleted cells revealed that knockdown of Zeb1 resulted in a significant reduction to the number and size of organoid colonies (FIGURE 5B, C; S5A). mRNA expression analysis of these cultures showed knockdown of Zeb1 led to decreased levels of Axin2, a gene encoding a direct transcriptional target of Wnt/b-Catenin signaling (Zeng et al., 2010, Van Amerongen et al., 2012), which is implicated in mammary SC self-renewal, as well as cell cycle gene, $Ccnd1$. (FIGURE 5D).

To identify whether down-regulation of Axin2 and $Ccnd1$ is a consequence of direct transcriptional regulation by Zeb1, we interrogated a ChIP dataset generated using human
HepG2 liver cells (UCSC Genome Browser). Excavation of this dataset found ZEB1 binding to bind on promoter proximal regions on both \textit{CCND1} and \textit{AXIN2} gene loci (FIGURE S5B). We followed these leads asking whether Zeb1/ZEB1 binding to these genes also extended to mammary epithelial cells. We first attempted to perform Zeb1 ChIP-seq utilizing FACS-sorted basal mammary epithelial cells from the pregnant gland, but were unable to get successful signal due to the low expression level of this transcriptional regulator. To address this possibility and gain insight into the molecular function of ZEB1 in governing mammary basal proliferative potential, we performed ChIP-PCR using human basal-like breast cancer cell line MDA-MB-231 for its high expression levels of \textit{ZEB1} (FIGURE 6A). In agreement with the genome-wide assessment in HepG2, ZEB1 showed specific binding to the \textit{CCND1} promoter and on two sites within the promoter of \textit{AXIN2} (FIGURE 6B). A random region on the \textit{GAPDH} promoter was used as a negative control for ZEB1 binding (FIGURE 6B). While numerous investigations reveal Zeb1/ZEB1 to function primarily as a transcription repressor, a small number of studies identify specific conditions where this TF may function as a transcriptional activator (Postigo et al., 1999; Postigo et al., 2003; Vanderwalle et al., 2009). Specifically, recent studies uncover a novel interaction of ZEB1 with YAP1 that converts ZEB1 into an activator upon complexing with YAP1, a regulatory protein which itself does not possess DNA-binding capacity (Lehmann et al., 2016). YAP1 is an important mediator of the Hippo signaling pathway (Ramalho-Santos et al., 2002; Huang et al., 2005; Camargo et al., 2007), which regulates cell proliferation and apoptosis of SC populations.

We wondered whether ZEB1 interacted with YAP1 to activate transcription of \textit{CCND1} in mammary epithelial cells. Utilizing MDA-MB-231 cells, we detected protein expression of YAP1 (FIGURE 6C) and performed ChIP-PCR experiments targeting either ZEB1 or YAP1 to
determine whether both ZEB1 and YAP1 bound to the promoter of \textit{CCND1}. Indeed, we found both ZEB1 and YAP1 to exhibit near equal enrichment at the CCND1 promoter, suggesting a scenario where ZEB1 and YAP1 may activate transcriptional expression of Cyclin D1 (FIGURE 6D). We returned to MCF10A to ask whether regulation mediated by either ZEB1 or OVOL2 activated expression of YAP1 target genes, which encode genes involved in cell proliferation, regulation of apoptosis, and SC self-renewal (Zhang et al., 2008). Interestingly, over-expression of \textit{ZEB1} in MCF10A resulted in enriched positive correlation of a YAP1 target genes, while the same gene signature was inversely correlated with over-expression of \textit{OVOL2} (FIGURE 6D). Conversely, basal mammary epithelial cells of \textit{Ovol2} SSKO mice exhibit elevated levels of \textit{Yap1} which may work together with the elevated Zeb1 levels to support transcriptional activation of a subset of genes that Zeb1 may functionally activate (FIGURE 6E). Together, the data provides evidence for a novel role of EMT in regulating cell fate plasticity and proliferative potential of mammary basal cells, specifically through a tight regulatory circuit maintained by EMT-TF Zeb1 and master EMT inhibitor, Ovol2, that includes regulation of downstream Wnt signaling and YAP1 targets.

4.4 DISCUSSION

My work demonstrates the importance of a Zeb1-Ovol2 cross repression circuit in governing mammary epithelial cell fates. While loss of Zeb1 and Ovol2 compromises basal and luminal fates, respectively, their simultaneous ablation is able to normalize these defects, highlighting the importance of their balanced expression/activity. The finding that I have made using knockdown/knockout models in mouse cells is well-supported by similar findings generated by Dr. Kazuhide Watanabe using a human mammary epithelial cell model. To date,
my study is the first to reveal the influence of balanced expression of EMT-regulators on maintenance of proper epithelial lineages within the MG.

What is the molecular mechanism by which Zeb1 functions? While the number of Zeb1-binding targets continues to enumerate, the mechanism by which this EMT-TF functions remains largely elusive. Most associations of Zeb1 function rely on our understanding of its capacity to complex with other regulatory proteins, which confer this Zeb1 with either an activating or inhibiting activity. Yet, despite this knowledge, much of the understanding we have on Zeb1 has been deduced from its function in early embryogenesis and pathological settings. A primary reason for this stems from embryonic studies that show germline deletion of Zeb1 (Zeb1<sup>−/−</sup>) results in perinatal lethality due to respiratory failure and multiple skeletal defects (Takagi et al., 1998). Thus, investigations addressing Zeb1 function in adult tissues necessitate alternative approaches, such as the implementation of conditional knockout alleles or knockdown studies.

Function of Zeb1 in normal epithelial tissues, particularly within the MG, has not been previously reported. Utilizing a lentiviral Zeb1 knockdown approach coupled with transplantation, I have provided evidence describing the critical importance of Zeb1 during mammary epithelial regeneration in vivo, possibly by maintaining a basal/SC fate. To date, Zeb1 has been most closely associated to functioning as a potent EMT-TF for its ability to down-regulate expression of E-cadherin (Cdh1) and other epithelial-specific genes that dissolve epithelial identity and produce a mesenchymal cellular phenotype (Schmalhofer et al., 2009; Vanderwalle et al., 2008; Grootenlaes et al., 2000). Zeb1 has also been shown to induce trans-differentiation of epithelial cells by initiating vast transcriptome-wide changes through regulation of various histone modifiers, such as Hdac1 and Hdac2 (Aghdassi et al., 2012). From the findings presented in this report, I provide no definitive evidence describing Zeb1 to function as
a transcriptional driver of EMT in adult mammary epithelium. Instead, the intermediate levels of Zeb1 that are observed in the native basal population could potentially highlight the pleiotropic functional roles Zeb1 may possess.

Specifically, my finding showing Zeb1 to directly bind to and regulate expression of Axin2 and Ccnd1 provide an additional mechanism by which basal SC activity is likely to be regulated by Zeb1. Indeed, positive regulation of proliferation-associated genes by Zeb1 has been documented (Sanchez-Tillo et al., 2013; Liu et al., 2010). Studies on mantle cell lymphoma show how reduced levels of Zeb1 lead to a reduction in cell viability and proliferation capacity that results from decreased expression of various proliferative markers, among them CCND1 (Sanchez-Tillo et al., 2013). Additionally, both Ccnd1 and Axin2 serve as downstream targets of Wnt and YAP signaling pathways that are involved in cell proliferation and capacity to self-renew (Teo et al., 2010; Zeng et al., 2010; Chen et al., 2014). Regulation of these Ccnd1 and Axin2 signaling targets are thus likely to underlie the reduced capacity of basal cells to proliferate and self-renew ex vivo. To further investigate the function of Zeb1 as a transcriptional activator of genes important for proliferation/self-renewal, I further associate the functional role of Zeb1 in positively regulating basal proliferation by implementing the luciferase reporter assay to study regulation of gene expression by ZEB1 and YAP at the CCND1 locus. Moreover, I can profile Zeb1+ basal cells through single-cell RNA-seq (scRNA-seq) to see if there is a significant enrichment of proliferation/cell-cycle genes during stages of active morphogenesis. Functional relevance of this correlation can be further demonstrated by introducing Ccnd1 or Axin2 into Zeb1-deficient basal cells to see if the proliferation defect is rescued.

My implementation of a regeneration model reveals that depletion of Zeb1 compromises the capacity of basal SCs to form a proper epithelial ductal network. Results from this assay
reveal that depletion of Zeb1 compromises the extent of mammary branching within these mammary transplants. Interestingly, recent studies from the Hinck laboratory at UCSC uncovered that the extent of mammary branching is, at least partly, a consequence of the ratio between basal to luminal cells (Macias et al., 2011). Integrating these findings with our data, it appears likely that Zeb1 by maintains a basal/SC fate that is necessary to maintain a balance of basal to luminal epithelial cells, which is relevant to regenerative mammary branching.

These tantalizing findings prompt us to next determine what role Zeb1 may have under physiological conditions, and to further excavate the role of Zeb1 in regeneration. Potential off-target effects are associated with use of shRNAs (Jackson et al., 2010). To reduce the possibility of erroneously introducing false-positive effects, I have utilized a total of two shRNAs that target unique regions of the Zeb1 transcript, and took a knockout approach using Zeb1 and Ovol2 floxed alleles. Moreover, within the lab, we have now generated mammary-epithelial specific Zeb1 knockout mice to examine the function of this EMT-TF has during active stages of post-natal development (i.e. pubertal and pregnancy-associated morphogenesis). These tools will allow us to further examine the molecular mechanisms affecting basal cell fate determination upon deletion of Zeb1.
FIGURE 4.1 Zeb1 is expressed in the basal compartment of the mammary epithelium during active morphogenesis, and chromatin activation underlies its pregnancy-induced upregulation.

(A) GSEA of microarray (left) and RNA-seq (right) data collected from sorted mammary basal and luminal cells reveals enrichment of a hallmark EMT signature in basal cells. (B) Heat map of microarray data depicting elevated levels of EMT/mesenchymal and epithelial genes in basal and luminal cells, respectively. (C) qPCR analysis comparing expression of Zeb1 and Ovol2 within the epithelial populations of the adult virgin mammary gland (N=2). (D) qPCR analysis showing dramatic transcriptional up-regulation of Zeb1 in pregnant basal cells compared to virgin basal cells (N=2). (E) ChIP-qPCR analysis of H3K4me3 and H3K27m3 histone modifications at the Zeb1 promoter in luminal and basal cells of virgin and pregnant glands. (F-H) Immunofluorescence detecting nuclear Zeb1 protein in basal cells from mammary glands isolated from pubertal (F) and pregnant (G) mice, but not from adult virgin. Basal cells are marked by Acta2 (a-SMA) expression.
Figure A: Hallmark epithelial mesenchymal transition

Figure B: Basal vs. Luminal

Figure C: Ovol2 vs. Zeb1

Figure D: Zeb1

Figure E: H3K4me3

Figure F: H3K27me3
FIGURE 4.2 Zeb1 deficiency in unsorted mammary epithelial cells results in an imbalance of basal and luminal/alveolar cell-derived primary organoid colonies.

(A) Representative images showing acinar-type organoids derived from luminal/alveolar cells or solid/branched-type organoids derived from basal cells. (B) Quantification of the two types of organoids in control and Zeb1 knockdown cultures. Note the significant increase of acinar/branched organoid ratio in the latter. Only GFP-positive organoids are included in the quantitative analysis. Data presented in main figure uses shRNA 818. Similar data were obtained using shRNA 819. See Supplemental Figure. (C) qPCR analysis describing differential expression of the indicated genes, using total RNA isolated from colonies derived from control and Zeb1-depleted organoids. (N = 3 independent experiments). Note efficient knockdown down of Zeb1 correlates with increased expression of luminal markers. (D) Knockout of ZEB1 in MCF10A results in increased expression of basal-luminal transitory keratin KRT19, but decreased expression of basal marker KRT14 (N = 4). Note that EMT-inducing factor TGF-b induces KRT14 expression while suppressing KRT19 expression, but these effects do not seem to require ZEB1. Courtesy of Kazuhide Watanabe. (E) Dox-inducible expression of Zeb1 in MCF10A cells expands the number of K14+ cells at the expense of K19+ and/or K14- cells. Courtesy of Kazuhide Watanabe. (F) Quantification of % populations from (G). (I) GSEA of microarray data on MCF10A reveals enriched expression of both EMT and stem cell gene signatures upon Zeb1 induction.
A. **shScramble** vs **shZeb1**

B. **Luminal / Basal Organoid**

C. **Zeb1**, **Krt19**, **Notch1**

D. **KRT14**, **KRT19**

E. **DOX(h)** vs **0h**, **24h**, **48h** vs **72h**

F. **ZEB1 induction**

G. **MCF10A-ZEB1**

H. **NES score** 4.16

I. **NES score** 2.62

J. **NES score** 1.11E-04

K. **OE** vs **EV**
FIGURE 4.3 Zeb1 deficiency in sorted basal cells compromises ductal branching upon transplantation.

(A) Representative images of regenerated mammary trees derived from FACS-sorted basal cells that were infected with GFP-expressing lentiviral particles containing shRNAs targeting the Zeb1 transcript or scramble control. Data presented in main figure uses shRNA 818 (N = 5). GFP-positive trees were visualized under fluorescence microscope and analyzed for (B) total ductal length, (C) total number of branching points, and (D) a ratio between branching points and ductal length. (E-H) Representative images and quantification of transplanted glands from hosts that were subsequently mated to WT males (N = 2). Analysis was performed at 8 weeks after date of transplantation. Pregnancy-associated changes were observed at 13 DPC. shRNA 819 was used for transplantation experiments assessing the pregnant gland.
FIGURE 4.4 Deletion of Ovol2 in mammary epithelial cells compromises luminal differentiation, and this defect is rescued by simultaneous deletion of Zeb1.

(A) qPCR analysis showing Zeb1 expression in MG sub populations belonging to control and Ovol2 SSKO mice. Note Zeb1 is specifically expressed in basal cells. (B) Expression of mammary epithelial subpopulation-specific cyto-keratins in cultured TEBs derived from control and Ovol2 SSKO knockout mice. Note the expansion of K14+ cells in SSKO TEBs. (C) Semi-qPCR analysis reveals truncated Ovol2 transcripts in cultured Ade-Cre-infected mammary epithelial cells from Ovol2 (heterozygous and homozygous) floxed, and Ovol2/Zeb1 double floxed mice. Gene expression was normalized against GAPDH. (N = 4). No P-value (gel). (D) qPCR analysis reveals reduced levels of Zeb1 transcripts in cultured, Ade-Cre-infected mammary epithelial cells from Zeb1 (heterozygous and homozygous) floxed mice. Gene expression was normalized against Gapdh (N = 4) – P value = 0.01 (E) Quantification of the two types of organoids generated in control, Ovol2 single, Zeb1 single, and Ovol2/Zeb1 KO cultures. Briefly, unsorted MEC were cultured first under proliferative conditions then induced to differentiate. Quantification after 2 weeks of culturing cells (N = 4 independent experiments). P-value between control and Ovol2 KO = 0.0081; P-value between control and DKO = 0.3096; P value between control and Zeb1 KO = 0.0074. (F) Dox-inducible expression of Ovol2 in MCF10A cells expands the number of K19+ cells at the expense of K14+ cells. Courtesy of Kazuhide Watanabe. (G) Quantification of % populations from (F). Courtesy of Kazuhide Watanabe. (H) Gene set enrichment analysis of MCF10A reveals de-enrichment of both EMT and stem cell gene signatures upon Ovol2 induction – effects just opposite of that of Zeb1 induction.
### Table:

<table>
<thead>
<tr>
<th></th>
<th>Avg. Organoid Total</th>
<th>Avg. Abnormal Organoids</th>
</tr>
</thead>
<tbody>
<tr>
<td>Control</td>
<td>23.2 ± 6.3</td>
<td>4.23% ± 0.1</td>
</tr>
<tr>
<td>Zeb1 KO</td>
<td>28.3 ± 15.9</td>
<td>5.18% ± 1.9</td>
</tr>
<tr>
<td>Ovol2 KO</td>
<td>16.1 ± 4.4</td>
<td>4.54% ± 0.5</td>
</tr>
<tr>
<td>Double KO</td>
<td>35.5 ± 19.9</td>
<td>14.27% ± 0.5</td>
</tr>
</tbody>
</table>

### Figures:

#### A. Zeb1

- Relative Expression
- WT Luminal, WT Basal, Zeb1 KO, Ovol2 KO, Double KO

#### B. Krt14 and Krt6

- Staining
- Control, SSKO

#### C. Zeb1

- Relative Expression
- Control, Zeb1 KO, Ovol2 KO, Double KO

#### D. Ovol2

- Western Blot
- Control, Zeb1 KO, Ovol2 KO, Double KO

#### E. Graphs

- Bar Graphs
- Control, Zeb1 KO, Ovol2 KO, Double KO

#### F. DOX (h)

- Images
- 0h, 24h, 48h, 72h

#### G. OVOL2 induction

- Graph
- Percent Population
- DOX (h): 0, 24, 48, 72

#### H. KRT19 / KRT14 / Nuclei

- Images
- MCF10A-OVOL2

---

**Legends:**
- NES score: -3.56, NOMp-value: 0, FDRq-value: 0
- NES score: -3.44, NOMp-value: 0, FDRq-value: 0
FIGURE 4.5 Zeb1 deficiency in sorted basal cells results in reduced clonogenicity and altered cell cycle/EMT gene expression.

(A) Representative images of organoids colonies derived from FACS-sorted basal cells that were infected with lentiviral particles expressing GFP and either shZeb1 or scramble control. (B) Quantification of the total number of organoid colonies from control and Zeb1 knockdown cultures. Note that only GFP-positive organoids are included in the quantitative analysis. (C) Quantification of colony diameter of organoids from control and Zeb1 knockdown cultures. (D) qPCR analysis describing differential expression of the indicated genes, using total RNA isolated from colonies derived from control and Zeb1-depleted organoids. (N = 3 independent experiments).
FIGURE 4.6 Zeb1 directly regulates the expression of EMT/ cell cycle genes and activates transcription in conjunction with YAP.

(A) Western Blot showing expression of ZEB1 and YAP1 in MDA-MB-231 cells. (B) ChIP-PCR results validate specific ZEB1 binding to the CCND1 and AXIN2 promoter in MDA-MB-231 cells. IgG was used as IP control. GAPDH promoter used as control genomic region. (C) ChIP-PCR shows co-occupancy of ZEB1 and YAP on CCND1 promoter in MDA-MB-231 cells. (D) Immunofluorescence analysis reveals elevated nuclear YAP1/TAZ signals in the basal mammary epithelium of Ovol2 skin/mammary epithelial-specific knockout glands. (E) GSEA analysis of the expression of YAP/TAZ and well-documented downstream target genes in MCF10A cells expressing either ZEB1 or OVOL2.
**Figure S1.** (A). FACS profile of lineage-negative mammary cell populations, as identified with EpCAM and CD49f surface markers. (B). Gene set enrichment analysis (GSEA) of microarray data collected from sorted mammary basal and luminal cells reveals enrichment of a EMT gene signature associated with metastasis. (C). Validation of Q-RT-PCR analysis Figure 1D, showing expression of basal and luminal markers (K14 and K8, respectively) are restricted within basal and luminal compartments of the mammary gland. (D). Confocal IF representation of a section taken from the pregnant mammary gland, revealing Zeb1-positive nuclei are within a-SMA cell boundary.
Figure S2. (A). Illustration depicting location of where shRNAs complementarily bind to and target Zeb1 mRNA transcript degradation. (B). Dot Plot showing infection efficiency of lentiviral particles in an unsorted population of MECs derived from the adult mouse. (C). Quantitation of Western Blot using ImageJ to measure relative protein knockdown efficiency of two shRNAs targeting Zeb1 and scramble control in 3T3 mouse embryonic fibroblasts. (D). Representative images of GFP-positive colonies produced by control and Zeb1-depleted basal cells, that were used to analyze differentiation and clonogenic capacity. (E). qRT-PCR analysis showing Krt8 expression change as a result of Zeb1 knockdown in primary mammary organoids derived from basal cells.
Figure S4. (A). Venn Diagram describing inverse relationship of EMT gene expression after Zeb1 or Ovol2 expression in MCF10A. (B). GSEA of mammary SC gene enrichment and differentiation modules after Ovol2 over-expression in MCF10A.
Figure S5. (B). Effect on organoid size after Zeb1 knockdown in FACS-sorted mammary basal cells. (C). Publicly available ZEB1 ChIP-seq dataset in HEPG2 cells depicting ZEB1 binding to AXIN2 and CCND1 promoter.
CONCLUSIONS AND FUTURE PERSPECTIVES

The findings described in this thesis provide evidence supporting a novel, previously unidentified, role of Zeb1 in maintaining a basal/SC fate within the mammary epithelium. While the study delineates the function of Zeb1 under regenerative conditions, this work prompts us to implement methodology that validates these data and enables us to further elucidate the functional role of this transcriptional regulator within a more physiologically relevant setting. Germline deletion of Zeb1 is not a viable model to investigate the role of this EMT-TF in postnatal MG development due to perinatal lethality (Takagi et al., 1998). Recently, the Dai laboratory obtained Zeb1<sup>fl<sub>ox</sub>/fl<sub>ox</sub></sup> mice (described in Brabletz et al., 2017), which offers a model to investigate the physiological role of Zeb1 in MG development. Specifically, Zeb1<sup>fl<sub>ox</sub>/fl<sub>ox</sub></sup> line will be crossed with K14-Cre driver (to generate Zeb1<sup>fl<sub>ox</sub>/fl<sub>ox</sub></sup>; K14-Cre) to generate mice with a mammary epithelial-specific deletion of Zeb1 (Zeb1 SSKO) which will be used to investigate functional role of Zeb1 within the basal population during pubertal morphogenesis as well as upon pregnancy.

In the scenario where Zeb1 SSKO mice fail to produce a viable ductal network, we can investigate the role of Zeb1 in the pregnant state by utilizing an inducible model (Zeb1<sup>fl<sub>ox</sub>/fl<sub>ox</sub></sup>; K5-rtTA; TetO-Cre; mTmG mice), to induce recombination at the appropriate developmental time point. Utilizing this method would allow us to determine whether successful recombination has occurred within the mammary epithelium (mTmG) and to further assess the contribution of these cells to formation of the mammary tree. Like the compensation exhibited by Ovol proteins in skin development (Lee et al., 2014), Zeb proteins, namely, Zeb1 and Zeb2, may have similar
compensative/redundant function within the MG. This said, the results presented through my work show this possibility is unlikely, as Zeb1 knockdown manifests a detectable phenotype.

My preliminary ex vivo experiments provide strong evidence for a functional interaction between Zeb1 and Ovol2. While Ade-Cre-mediated deletion of Zeb1 or Ovol2 alone led to reduced basal or luminal cells, simultaneous deletion of both normalized these deleterious effects. This finding implicates the importance of the Zeb1-Ovol2 cross-repression circuit in governing mammary basal cell fate plasticity. To address the genetic interaction between Zeb1 and Ovol2 in vivo, basal specific double knockout of Zeb1 and Ovol2 (Zeb1-Ovol2 DKO) will be generated using a K14-Cre driver.

Mechanistically, my data suggests a model that Zeb1 transcriptionally activates the expression of Ccnd1 and Axin2 to positively regulate basal SC activity. Intriguingly, a recent study shows that Zeb1 can be converted from a transcriptional repressor to an activator by complexing with YAP (Lehmann et al., 2016). Supporting this notion, my preliminary ChIP-PCR analysis detected YAP binding to CCND1 promoter in MDA-MB-231 cells at the same site as ZEB1 does. Moreover, Luciferase reporter assays need to be performed to determine whether introducing Zeb1 alone represses, but together with YAP activates, Ccnd1 and Axin2 promoter activity. Finally, to develop a more comprehensive understanding of the regulatory functions of Zeb1, I have conducted a genome-wide analysis of the Zeb1 binding targets utilizing MDA-MB-231 human basal breast cancer cell line. Analysis of the ChIP-seq data is currently ongoing but to date, has not produced data that gives clear indication of the genome-wide binding targets of this transcriptional regulatory protein. The number of ZEB1 binding peaks are low when referenced against both IgG or input DNA, and the distribution of the ZEB1 peaks do not conform to any specific chromosomal regions, suggesting that the sample is noisy. Surprisingly,
a small subset of the detected ZEB1 peaks exhibit enriched binding to SMAD and ZEB1 binding motifs, suggesting that a small number of identified peaks may show true binding. To gain a better understanding of the genes regulated by ZEB1, one possible alternative approach to take would be to attempt single cell ChIP-seq (seChIP-seq) or by Flag-tagging ZEB1 and immunoprecipitating protein-DNA complexes.
APPENDIX A:

UNCOVERING THE CELLULAR HETEROGENITY WITHIN THE BASAL COMPARTMENT OF THE ADULT MOUSE MAMMARY GLAND

5.1 INTRODUCTION

Recent advances in Next-Generation Sequencing (NGS) technologies provide an unparalleled opportunity to generate new, deeper insight into many biological systems. Single-cell RNA-seq (scRNA-seq) is a novel, powerful tool used to identify cellular heterogeneity existing within tissues in an unbiased manner (Picelli et al, 2014), and has been implemented to characterize sub-populations and rare cell types through expression of unique molecular identifiers. Additionally, time-series expression analysis from scRNA-seq data can be utilized to order the temporal transcriptional dynamics of individual cells in an unsupervised manner to reveal cell fate decisions and lineage histories (Trapnell et al., 2014). Since its inauguration in 2009 (Tang et al., 2009), scRNA-seq has been successfully applied to describe the existing heterogeneity in various tissues such as lung (Treutlein et al., 2014), brain (Pollen et al., 2014), and in myriad cancer types (Tirosh et al., 2014; Patel et al., 2014).

Isolation and characterization of tissue-specific SCs is fundamental to understanding organ development and homeostasis. The MG is a glandular organ present in mammalian organisms with a function in providing newborn offspring with a source of nutrients and antimicrobial protection (Peaker, 2002; Watson and Khaled, 2008). This complex secretory organ will constitute an extensive, functional network of epithelial ducts and milk-producing alveolar structures that will develop through distinct developmental stages that take place largely after birth. Structurally, the epithelial component of this organ is established from a cellular bi-layer that is maintained by two major mammary lineages derived from a single embryonic cellular
origin (Watson et al., 2008). The lumen of the mammary epithelium is lined with an inner luminal population of epithelial cells that can onset terminal differentiation programs to become milk-producing cells (Visvader et al., 2014). Conversely, the outer layer of the mammary epithelium is composed of a unique population of basal/myoepithelial cells possessing an epithelioid identity between epithelial and mesenchymal states (Prater et al., 2014), which are responsible for producing the contractile forces that eject the nutrient-rich milk (Raymond et al., 2011).

As described, the extensive morphogenic alterations of this dynamic organ are manifested by a complex hierarchy of mammary SCs (Prater et al., 2014). Specifically, both in vivo lineage tracing and regenerative transplantation studies identify the basal compartment of the MG to be the resident population of multi/bi-potent SCs that have the capacity to give rise to a mature epithelial ductal network consisting of the two cellular (basal and luminal) lineages (Rios et al., 2014; Wang et al., 2014). However, there is also evidence of long-lived unipotent SC/progenitors within this population (Van Keymeulen et al., 2011), as well as the terminally differentiated myoepithelial cells (Prater et al., 2014) within the basal compartment. Although the molecular mechanisms that regulate epithelial plasticity and the mammary epithelial SC state remain poorly understood, significant advancements have been made toward identifying the properties of mammary SCs. Various markers have been identified that depict unique basal sub-populations that are associated with capturing the mammary SC phenotype (Visvader et al., 2014). In addition, partial EMT and its link with epithelial stemness continues to engender great interest, and is a concept that is prominent within cells of the basal compartment of the MG.

Through this study, I utilize scRNA-seq to begin resolving the cellular heterogeneity existing within the basal population of the adult virgin MG, a stage characterized by a mature
ductal network and homeostatic growth. Mammary SCs within this compartment reside in a quiescent state but are responsive to external stimuli presented by onset of pregnancy (Joshi et al., 2010). Through this preliminary assessment, our findings describe 3-4 unique cellular clusters within the basal compartment that exhibit unique expression of EMT-related genes that can be sub-clustered into five populations ranging in the extent of expression. The findings prompt interesting questions as to the role EMT may have in delineating basal to luminal fate decisions. Moreover, my data provides support of two possible quiescent mammary SC populations.

5.3 MATERIALS AND METHODS

For mammary cell preparation and FACS C1 loading, and single cell analysis, please refer to CHAPTER 2 MATERIALS AND METHODS.

5.3 RESULTS

To characterize the cellular heterogeneity in the SC-containing compartment of MG during homeostasis, I performed scRNA-seq using FACS-sorted basal cells from C57BL/6J WT adult virgin mice (8 weeks of age). From two independent experiments derived from two separate mice, I captured FACS-sorted basal cells at the cell capture site of the Fluidigm C1 10-17um 96 IFC and automated cell lysis and generated cDNA libraries. I captured and sequenced a total of 103 single cells. After sequencing, I aligned each sample’s reads using RSEM and implemented SEURAT version 1.2, a widely-used R-based toolkit for cell genomics, used commonly for its integration of analytical techniques that identify highly variable genes, reduce sample dimensionality (PCA, ICA, tSNE), and discover differentially expressed genes and
markers (Satija et al., 2015). From the 103 basal cells sequenced, a total of 93 cells expressed more than 1000 genes at a detectable level, which were used for downstream analysis.

tSNE (t-Distributed Stochastic Neighbor Embedding) analysis, using the eleven top principal components as input, identified the presence of 3-4 unique clusters, which exhibited expression of unique molecular identifiers (FIGURE 5.1A). Expectedly, basal markers, Krt14 and Acta2 displayed high expression in almost every cell sequenced. Interestingly, from this analysis, I observe that there is a small subset population that is not K14\textsuperscript{high};Acta2\textsuperscript{high}, but rather only expresses one or the other cytoskeletal filament marker (FIGURE 5.1B, C). Moreover, expression of Krt5, another well-established basal marker within the MG, was expressed in only a fraction of the total cells sequenced (FIGURE 5.1D).

EMT-like cellular changes accompany ductal elongation of the mammary epithelium, and a partial EMT has been associated with producing a SC-like state. The findings presented in this thesis (REFER TO CHAPTER 4) demonstrate that the basal population is the mammary epithelial compartment enriched with expression of EMT-related genes, including various EMT-TFs that drive progression of the program. I sought to determine the relationship between EMT gene expression and cellular subsets observed above. Utilizing a modified Gene Ontology hallmark term for EMT that integrated expression of well-known EMT drivers, I uncovered 5 unique population clusters that, except for one, correlated with the clusters generated by t-SNE analysis (FIGURE 5.2). The finding showing a possible gradient of expression of EMT-related genes between the identified basal sub-populations will enable us to further seek how the expression pattern of these genes correlates with genes that have previously been attributed to the mammary SC population.
While the adult virgin MG is characterized by homeostatic maintenance, it is highly responsive to hormonal stimuli presented upon pregnancy (Joshi et al., 2010). Previously, several distinct basal populations have been shown to present a SC phenotype (Visvader et al., 2014) including those which define a quiescent state. Wnt target Lgr5 has received significant attention as potential SC markers through its ability to denote SC populations in other epithelial tissues (Barker et al., 2013). Indeed, a small subset of cells within the adult basal population are recognized as Lgr5\(^+\). Through work published by Fu et al., (2016), expression of Lgr5 in tandem with Tspan8 marks mammary SC subsets that differ from others in state of activity. Lgr5\(^+\); Tspan8\(^{hi}\) denotes quiescent mammary SCs, while the Lgr5\(^+\) and Tspan8\(^{lo}\) population demonstrate strong regenerative activity (Fu et al., 2016). Within the virgin adult scRNA-seq dataset, I have identified both Lgr5\(^+\), Tspan8\(^{lo}\), and Lgr5\(^+\); Tspan8\(^{hi}\) subsets (FIGURE 5.3). Interestingly, recent findings demonstrate that expression of protein receptor C (Procr) identifies a bipotent mammary SC population that exhibits heightened regenerative capability (Wang et al., 2015). Again, Procr\(^+\) cells were identified within this dataset, and surprisingly, showed high correlation to Lgr5\(^+\) cells, as cells identified as Procr\(^+\) are Lgr5\(^+\), although not all Lgr5\(^+\) cells express Procr (FIGURE 5.3). A small number of Zeb1\(^+\) cells was also detected, which shows some overlap with Lgr5\(^+\) basal cells (FIGURE 5.3). These findings provide new insight into the heterogeneity existing within the MG and the unique subpopulations expressing SC-traits.

### 5.3 SEURAT SCRIPT USED FOR ANALYSIS

SEURAT v1.2 for Single Cell RNAseq Analysis on RStudio Version 0.99.903

```r
# Command used to initiate Seurat and the associated tools used for analysis and visualization
library(devtools)
library(Seurat)
library(ggplot2)
library(Matrix)
```

107
# Command used to upload scRNAseq FPKM matrix in
nbt.data=read.table("/Users/alvarovillarreal/Documents/Results/scRNAseq/scRNAseq_1/wtbasal
1matrixgrcm38_with_genename.txt",sep="\t",header=TRUE,row.names=1)

# Conversion of FPKM (Fragments per Kilo-base Million) values to TPM (Transcripts per
# Kilo-base Million) values
nbt.data<-sweep(nbt.data,2,colSums(nbt.data),'/')*(10^6)

# Transformation of TPM values to log scale
nbt.data=log(nbt.data+1)

# Visualization of the top left corner of the matrix file (of a rectangular data set)
corner(nbt.data)

# Command asking for the dimensions of the data file – DIM = Dimension
dim(nbt.data)

# Command calling Seurat to consider the transformed data as raw data
nbt=new("seurat",raw.data=nbt.data)

# Command for Outlier Analysis: For genes: genes must be present >3 cells. For cells: cells
# expressing >1K genes
nbt=setup(nbt,project="NBT",min.cells = 3,names.field = 2,names.delim = "_",min.genes =
1000,is.expr=1,)

# Command calling for dimensions of raw data after outlier removal
nbt

# Command to identify variable genes across the single cells and for a visualization of the
# variable genes; identifies genes that are outliers on a mean variability plot. Average
# expression of gene (x-axis) vs. variability of expression (y-axis) followed by dividing genes
# into num.bin (default = 20). Overall purpose to identify variable genes while controlling for
# strong relationship between variability and average expression. – the cutoff is 2 fold
difference in expression.
nbt=mean.var.plot(nbt,y.cutoff = 2,x.low.cutoff = 2,fxn.x = expMean,fxn.y = logVarDivMean)

# Command to determine number of variable genes
length(nbt@var.genes)

# Produce PCA analysis using PC1 and PC2 of the variable genes as input. These PC values
depict the greatest dimensionality between samples.
nbt= pca(nbt,do.print=FALSE)
pca.plot(nbt,1,2,pt.size=2)
print.pc(nbt,1)
pca.plot(nbt,dim.1=1,dim.2=2,pt.size = 2)

# Plot PCA of total data
autoplot(prcomp(t(nbt@data)))

# The previous analysis was performed on <400 variable genes. To identify a larger gene
# set that may drive biological differences, but did not pass our mean/variability thresholds,
# we first calculate PCA scores for all genes (PCA projection)—Project Principal Component
# Analysis onto full dataset- takes a pre-computed PCA and projects this onto the entire
dataset.- there are now gene scores for all genes
nbt=project.pca(nbt,do.print=FALSE)
pcHeatmap(nbt,pc.use = 1,use.full = TRUE,do.balanced = TRUE,remove.key = FALSE, cexRow = .5, cexCol = .5)

# Run tSNE using our 11 significant PCs as input (in our case 1:5)
nbt=run_tsne(nbt,dims.use = 1:5,max_iter=2000,dim_embed=3)
plot(nbt@tsne.rot[,1],nbt@tsne.rot[,2])

# Density cluster the tSNE map - note that the G.use parameter is the density parameter for the clustering - lower G.use to get finer settings
# Cells which are 'unassigned' are put in cluster 1 - though in this case there are none
# Assigned cluster will be placed in the 'DBClust.ident' field of nbt@data.info. Putting set.ident=TRUE means that the assigned clusters will also be stored in nbt@ident
nbt=DBclust_dimension(nbt,1,2,reduction.use = "tsne",G.use = 5,set.ident = TRUE)
tsne.plot(nbt,pt.size = 1)

# ips.markers is file showing the top differentially expressed genes in pop 3 in comparison to 1 and 2, threshold is +/- 2 fold
ips.markers=find.markers(nbt,3,thresh.use = 2)
# Print top five lines of the file ips.markers
print(head(ips.markers,5))

# Write a txt file containing the differentially expressed genes for better visualization
write.table(ips.markers, file="Cluster3_Markers.txt", sep="\t",col.names = TRUE)

# Visualization of 3D PCA and 3D T-Sne plot
library("rgl", lib.loc="/Library/Frameworks/R.framework/Versions/3.3/Resources/library")
plot3d(nbt@pca.rot[,1],nbt@pca.rot[,2],nbt@pca.rot[,3])
nbt=run_tsne(nbt,dims.use = 1:5,max_iter=2000,dim_embed=3)
plot3d(nbt@tsne.rot[,1],nbt@tsne.rot[,2],nbt@tsne.rot[,3])

# Visualization in 2D PCA and 2D T-Sne plot
library(ggfortify)
autoplot(prcomp(nbt@data))
autoplot(prcomp(t(nbt@data)))
nbt@tsne.rot[,1]
plot(nbt@tsne.rot[,1],nbt@tsne.rot[,2])
plot(nbt@pca.rot[,1],nbt@pca.rot[,2])
nbt=DBclust_dimension(nbt,1,2,reduction.use = "tsne",G.use = 8,set.ident = TRUE)
tsne.plot(nbt,pt.size = 1)

# Determine which category each cell falls under
nbt@ident

# Identify the differentially expressed markers that classify each category
ips.markers=find.markers(nbt,3,thresh.use = 2)
print(head(ips.markers,5))

# Visualize specific genes
par(mfrow=c(1,2))
feature.plot(nbt, c("ENSMUSG00000026728_Vim", "ENSMUSG00000026872_Zeb2"),cols.use = c("blue", "grey"), pt.size = 1)

# In violin plot
vlnPlot(nbt, genes.viz)
# In t-SNE plot
feature.plot(nbt, genes.viz, pt.size = 1)

matrix_genes <- rownames(nbt@data) %in% EMT_genes[, 1]
genesis_ind <- grep("TRUE", matrix_genes)
do.HeatMap(nbt@data[genes_ind, ], genes.use = NULL, order.by.ident = TRUE, slim.col.label = TRUE, remove.key = TRUE, cexRow = 0.3)
FIGURE 5.1 Single-cell RNA-seq analysis of the basal mammary epithelium at adult stage of MG development. (A) t-SNE plot generated from 11-top most significant principal components reveals the presence of 3-4 clusters, which express unique molecular identifiers (not shown). (B-D). Expression of Krt14, Acta2, and Krt5 visualized on t-SNE plot shows cell-specific expression of these markers. Intensity of expression is determined by the shade of blue in each cell.
FIGURE 5.2 Expression of EMT-related genes in basal subsets of the adult virgin. (A)

Associating a hallmark EMT signature to tSNE-generated basal population sub-clusters
FIGURE 5.3 Expression and co-localization of proposed SC markers within the MG. (A)

Expression of *Lgr5*, *Zeb1*, and *Procr* visualized on t-SNE plot shows cell-specific expression of these markers. Intensity of expression is determined by the shade of blue in each cell.
REFERENCES


Teo, J.-L., & Kahn, M. (2010). The Wnt signaling pathway in cellular proliferation and
differentiation: A tale of two coactivators. *Advanced Drug Delivery Reviews, 62*(12), 1149–

journey of morphogenesis and commitment. *Development (Cambridge, England), 135*(6),


Zeng, Y. A., & Nusse, R. (2010). Wnt proteins are self-renewal factors for mammary stem cells
and promote their long-term expansion in culture. *Cell Stem Cell, 6*(6), 568–577.
http://doi.org/10.1016/j.stem.2010.03.020

van Amerongen, R., Bowman, A. N., & Nusse, R. (2012). Developmental Stage and Time
Dictate the Fate of Wnt/β-Catenin-Responsive Stem Cells in the Mammary Gland. *Cell

(2014). Mammary stem cells have myoepithelial cell properties. *Nature Cell Biology,
16*(10), 942–950. http://doi.org/10.1038/ncb3025

Identification of multipotent mammary stem cells by protein C receptor expression. *Nature,
517*(7532), 81–84. http://doi.org/10.1038/nature13851

Visvader, J. E., & Stingl, J. (2014). Mammary stem cells and the differentiation hierarchy:
http://doi.org/10.1101/gad.242511.114

evidence and unresolved questions. *Nat Rev Cancer, 8*. http://doi.org/10.1038/nrc2499


Van Keymeulen, A., Rocha, A. S., Ousset, M., Beck, B., Bouvencourt, G., Rock, J., … Blanpain,

Rios, A. C., Fu, N. Y., Lindeman, G. J., & Visvader, J. E. (2014). In situ identification of
http://doi.org/10.1038/nature12948


