Transcription Factor RFX3 Stabilizes Mammary Basal Cell Identity

By

Erica Marie Tross

Dissertation

Submitted to the Faculty of the

Graduate School of Vanderbilt University

in partial fulfillment of the requirements

for the degree of

DOCTOR OF PHILOSOPHY

in

Cell & Developmental Biology

May 13, 2022

Nashville, Tennessee

Approved:

Christopher Wright, D. Phil.

Ian Macara, Ph.D.

Emily Hodges, Ph.D.

Vito Quaranta, Ph.D.

Mark Magnuson, Ph.D.

Thank you, God, for Your promise in Jeremiah 29:11.

This work is dedicated to the Tross and Walwyn families.
To my ancestors upon whose backs this opportunity was built and the future generations who will carry this work forward.

ACKNOWLEDGEMENTS

This work is dedicated to my grandparents, St. Clair (Grandpa) and Violet (Mom) Walwyn who instilled in me the importance of keeping God first and taking advantage of every learning opportunity. Mom, thank you for teaching me the importance of therapy – retail therapy. My dearest Aunty Ketty (Keturah Trotman), your belief in me kept me going during times that I thought I had nothing else to give. You never missed my calls and you overflowed my cup with encouragement, hot chocolate and a side of freshly baked bread. Thank you for always reminding me to, 'vaya con Dios'.

To my parents, Rodolfo and Julie, you are my backbone, you are my strength. I was so blessed to have you both right by my side during this entire journey. Dad, thank you for being whomever I needed you to be whenever I needed you to be them. My beauty consultant, my therapist, my chef, my mechanic, my everything – all from hundreds of miles away. Mommy. My best friend forever. My twin. Thank you for always burning the midnight oil with me, for always answering the phone and for always taking my side. Even when you didn't agree with me 100%, you still found a way to be in my corner. Loren and Brandon, my brothers, thank you both for being trailblazers. For setting the standard, for jumping every hurdle of adversity and always standing tall in your purpose. Brandon we've been on the graduate school journey together and the vent sessions were epic.

André Kristopher Artís, my partner in crime whom I love, admire and appreciate. Thank you for always celebrating the small steps in this long journey. Thank you for coming to the lab with me and for helping me with my experiments. For being my one man audience as I prepared for presentations and my favorite study buddy. You see my potential and never let me settle, you never enabled me. Thank you for pushing me to the limit, for telling me the things that I didn't want to hear, for giving me tough love. You wanted me to be my best and work hard, but never without also playing hard. Thank you for keeping me balanced.

To my sweet girlfriends that kept me grounded, made sure that I made time for self-care, and taught me to see the silver lining in every situation – Chere Ann Weaver, my love. Thank you for your prayers, love, laughs and encouragement. Even though

you're so far away, it felt like you were much closer. Melanie Victoria Brady, my best friend, my sister, my rib– we started this journey together and stayed the course. The dinner and a movie dates, countless Face Time sessions – you were a critical part of this journey. My lifeline. Jessica Jackson Abner – we made it through all of those classes, exams, experiments, thank you for holding it down with me! Myri, Carcia and Jamisha – thank you for being the fumes in my fuel tank that helped me accelerate through the last few miles of this journey! Ke'Ara and Lana, thank you both for being such sweet mentees and friends who loved me through it all and kept me laughing.

Thank you to Dr. Yeodono Sovyanhadi and Dr. Safawo Gullo who were pillars during my time as a biomedical student at Oakwood University. Thank you to Dr. Brian Nelms who trained and supported me as I took my first steps as a basic scientist in his lab at Fisk University. Thank you to Dr. Hugh Fentress who played a major role as a mentor during my time at Vanderbilt. The love you have for your students is unmatched and I'm glad that you're a part of my tribe.

Dr. Ian Macara, my P.I. THANK YOU- for your endless fountain of patience, support, encouragement and overall passion for research. Thank you to all of my lab mates for your help and encouragement. Thank you to my committee members, for believing in my success, helping me to think critically, and giving me invaluable feedback on my project over the years. Thank you to Vanderbilt's Interdisciplinary Graduate Program, Vanderbilt's Department of Cell and Developmental Biology, the Program in Developmental Biology, the Vanderbilt Program in Molecular Medicine the Initiative to Maximize Student Diversity and the Organization of Black Graduate and Professional Students

Thank you to Dr. Keivan Stassun, Dr. Arnold Burger, Dr. Kelly Holley-Bockelmann, Dr. Dina Stroud and all of the other faculty, staff and students in the Fisk-Vanderbilt Master's-to-PhD Bridge Program. Thank you for the vision that ignited the flame that has guided the path for so many aspiring scientists.

TABLE OF CONTENTS

Page

LIST OF FIGURES

Diagram/Figure **Page**

LIST OF TABLES

Table **Page**

LIST OF TERMS AND ABBREVIATIONS

MG	Mammary gland
E	Embryonic
PTHrP	Parathyroid hormone-related protein
TEB	Terminal end bud
s-SHIP	*stem*-SH2 domain-containing inositol 5-phosphatase
MaSC	Mammary stem cells
FACS	Fluorescence-activated cell sorting
EMT	Epithelial to mesenchymal transition
GH	Growth hormone
PRL	Prolactin
IGF1	Insulin-like growth factor 1
ESR1/ER	Estrogen receptor
PR	Progesterone receptor
MMP3	Metalloproteinase 3
CD45	Protein tyrosine phosphatase receptor type C
TER119	Lymphocyte antigen 76
CD31	Platelet/endothelial cell adhesion molecule 1
CD24	Heat-stable antigen
CD29	β1-Integrin
CD49f	α6-Integrin
α-SMA	Alpha smooth muscle actin
Myh11	Smooth muscle myosin, heavy polypeptide 11
K14	Keratin 14
K5	Keratin 5
K18	Keratin 18
K8	Keratin 8
LRP5/6	LDL receptor-related protein 5/6
Acta2	Actin, alpha 2, smooth muscle, aorta
ProCr	Protein C receptor
LGR5	Leucine-rich repeat-containing G protein-coupled receptor 6
S/L	Small/Large
WAP	Whey acidic protein
PyMT	polyoma middle T

I. INTRODUCTION

1.1 Overview

The mammary gland, a hallmark of mammals, is a highly specialized organ that produces milk to nourish offspring. Murine mammary gland development spans over embryogenesis, puberty, and pregnancy and is regulated by intricate signaling networks between mammary epithelial cells and the stroma. The two main cellular lineages that constitute the mammary gland are the milk producing luminal cells, and the contractile basal/myoepithelial cells. Several epithelial tissues depend on multipotent stem cells for their proper development and maintenance, but the mammary gland is maintained by an alternate mechanism. Transplantation assays and lineage tracing experiments have produced controversial results in the search for the mammary stem cell, but the growing consensus is that postnatal mammary development occurs from unipotent progenitor cells. Interestingly, though, basal mammary epithelial cells retain reprogramming potential that is repressed *in situ*, but observed *in vitro*.

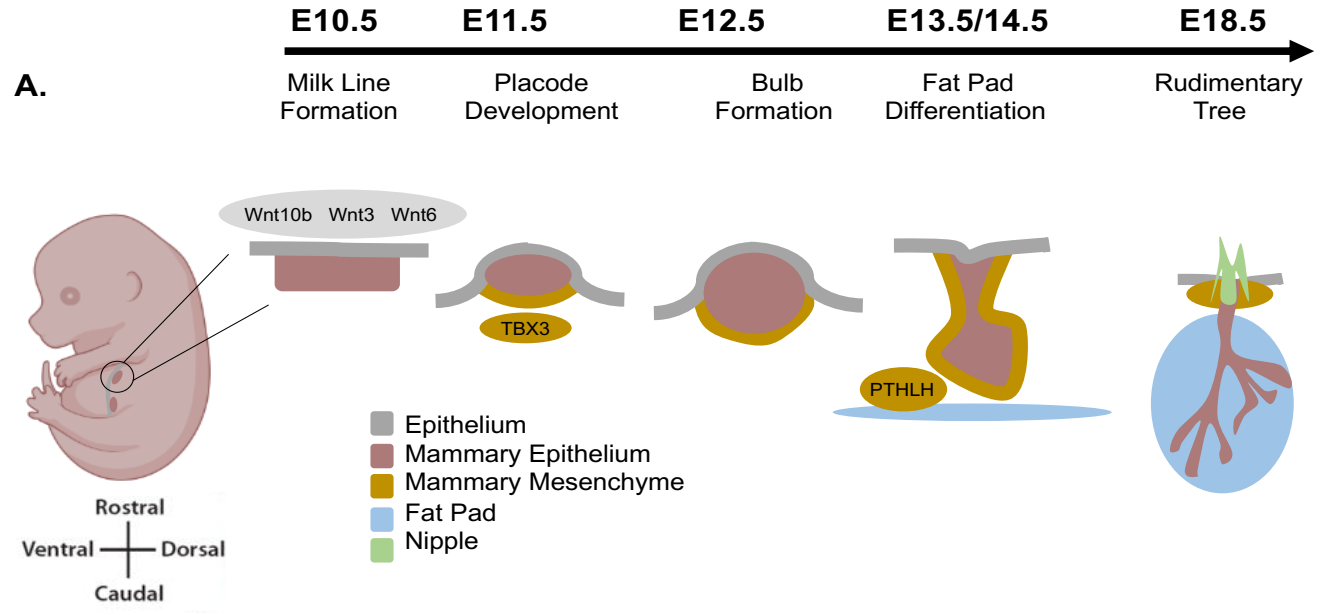
1.2 Murine Mammary Gland Development

The murine mammary gland (MG) is a highly specialized organ and one of the most regenerative in the body. The primary function of the MG is to provide nourishment and passive immunity for offspring. MG development occurs in 3 stages. Embryonic, pubertal, and reproductive development.

1.2.1 Embryonic Development/Primordial

Murine mammary gland (MG) development begins at embryonic day (E) 10 and extends to E18.5. Bilateral stripes of multilayered ectoderm, or milk lines, form on the ventral surface of the embryo and run anterior to posterior from the forelimb bud to the hindlimb bud (Diagram 1). This line of ectodermal cells then migrate and thicken to form symmetrical placodes in the location of the mammary buds (5 pairs) [1-4]. *Wnt10b, Wnt3* and *Wnt6*, are expressed in the mammary line and are important in specification [5, 6]. The T-box transcription factor, *Tbx3*, is expressed very early in restricted areas of mesenchyme and is critical for the proper migration of cells to form mammary placodes[7]. At E11.5, the mammary placodes develop into bulbs of epithelial cells that

are distinct from the surrounding epidermis and appear as elevated knob structures up until E13.5, when they descend into the underlying dermis [8] (Diagram 1, A). Around the sunken epithelial bud, mesenchymal cells become the mammary mesenchyme. The embryonic mammary mesenchyme provides critical inductive signals through parathyroid hormone-related protein (PTHLH) that direct the rudimentary bud formation and specify mammary epithelial cell differentiation. PTHLH is also important for androgen receptor activation in the mesenchyme of male embryos to signal for mammary bud degradation[9]. At E14.5, the fat pad precursor differentiates. Development continues with epithelial cell proliferation and elongation of the bud, forming a stalk connecting the bud to the epidermis like a small tree that will invade the fat pad and begin branching into about 10-15 branches. The nipple arises from epidermal cells that overlay the bud, and lumen formation becomes visible by E16.5. The spaces within the lumen increase in size (via programmed cell death) and number until E18.5[10]. At this point the rudimentary tree remains largely quiescent undergoing a period of allometric growth, keeping up with overall body development.


A.
E10.5
E11.5
E12.5
E13.5/14.5
E18.5
Milk Line Formation
Placode Development
Bulb Formation
Fat Pad Differentiation
Rudimentary Tree
Wnt10b Wnt3 Wnt6
TBX3
PTHLH
Epithelium
Mammary Epithelium
Mammary Mesenchyme
Fat Pad
Nipple
Rostral
Ventral
Dorsal
Caudal

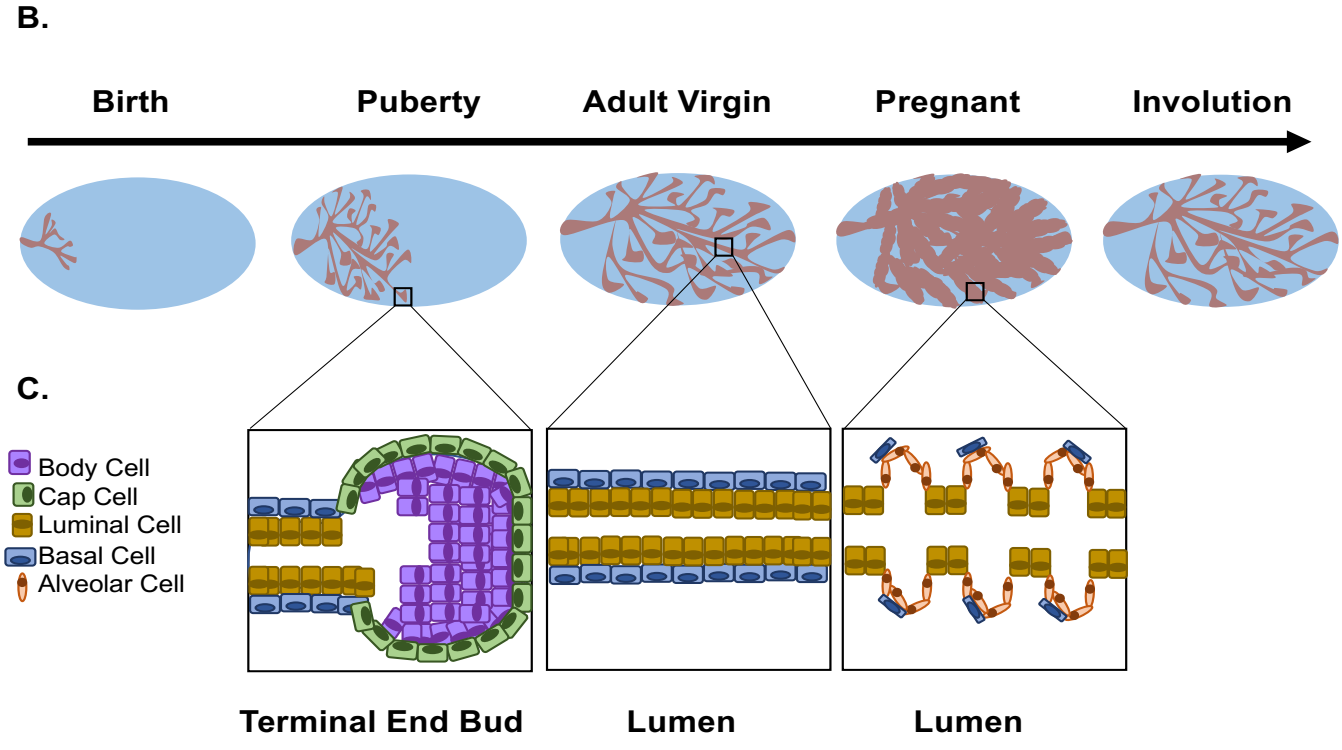

B.
Birth
Puberty
Adult Virgin
Pregnant
Involution
C.
Body Cell
Cap Cell
Luminal Cell
Basal Cell
Alveolar Cell
Terminal End Bud
Lumen
Lumen

Diagram 1. Murine Mammary Gland Development. A. Mammary gland development begins around embryonic day 10 (E10) with formation of a milk line (gray) above the mammary epithelium (pink). At E11.5, placodes form symmetrically along the milk lines and begin to invaginate to form buds (E12.5-E14.5). At E15.5 the mammary epithelium begins to proliferate and elongate to form the primary sprout that drives migration through the mammary mesenchyme towards the fat pad (blue). Lumen formation gives rise to the nipple (green). On E18.5, the mammary epithelium forms a rudimentary branched structure connected to the nipple. B. The embryonic mammary anlage is present at birth and remains quiescent until puberty. Terminal end buds (TEBs) drive the mammary epithelium to proliferate and fill the fat pad in response to pubertal hormones. TEBs, specific to rodents, have an outer layer of cap cells surrounding an inner layer of body cells. Adult mice have an epithelial tree-filled fat pad absent of TEBs. The ducts of the epithelial tree have an inner layer of luminal cells and an outer layer of contractile basal cells. Hormonal changes that occur during pregnancy drive an expansion of milk-producing alveolar cells upon lactation. The milk-filled alveoli fill the majority of the fat pad. Upon weaning, involution occurs by cell death returning the gland to a resting state.

1.2.2 Pubertal Development (Mammary Gland Components, The Terminal End Bud)

During puberty expansive proliferation of the mammary epithelium occurs driving mammary anlage invasion into the mammary fat pad. In mouse, this ductal elongation and branching morphogenesis is driven by highly proliferative terminal end buds (TEBs) [11, 12]. TEBs are club shaped structures at the ends of the growing ducts that contain a single proliferative outer layer of epithelial cap cells, surrounding a multilayered body of epithelial cells located at the invading front of the branch[13, 14]. The primary ductal system is generated by TEB bifurcation regulated by the stroma. Secondary branches grow laterally from the primary ducts. The epithelial cells that make up the cap of the TEB differentiate into myoepithelial cells [15] (Diagram 1, B). Branching morphogenesis continues until the TEBs reach the distal end of the fat pad and then the structure disappears. The production and activation of TGFβ and other mechanical and local cues regulate the termination of branching[16]. Under cyclical ovarian stimulation, or the estrus cycle, short tertiary branches will grow in response to progesterone, only differentiating into milk producing alveolar buds under the influence of pregnancy hormones. In the absence of pregnancy hormones, tertiary branches apoptos, leaving the mammary gland with primarily primary and secondary outgrowths. Several hormones and growth factors are important for normal branching morphogenesis during puberty, including growth hormone (GH), prolactin (PRL), insulin-like growth factor 1 (IGF1), and estrogen receptor (ESR1)[17-19].

Mammary Gland Components- The mammary gland (MG) is composed of multiple cell types that work together to maintain a functional organ. These include epithelial, fibroblast, vascular, immune, and lymphatic cells with adipose making up the majority of the organ. ***Epithelial cells*** form a bilayered ductal system of apically oriented luminal cells surrounded by a layer of basally oriented contractile myoepithelial or basal cells that contact the basement membrane (Diagram 1, B). The luminal cell population expresses keratins 8 and 18 and can be subdivided into estrogen receptor $(ER)^+$/progesterone receptor $(PR)^+$ and ER^-/PR^- cells. Lineage-tracing of the luminal cell population has revealed that during pregnancy, lactation and involution, ER^+ progeny are restricted to the ER^+ lineage and ER^- progeny are restricted to the ER^-

lineage [20-22]. Myoepithelial cells express keratins 5 and 14 and smooth muscle actin, which mediates their contractile function. Several putative stem cells were thought to exist among the basal population [23-25]. As will be discussed below, caps cells and body cells are specialized epithelial cells within the TEB that are less differentiated than basal cells or luminal cells and are highly proliferative [14, 26]. Lastly, during pregnancy, luminal cells differentiate into milk producing cells that form alveoli. ***Fibroblasts****,* found on the basal side of the epithelium embedded in the fat pad serve many functions. Fibroblasts provide instructions to the epithelium during branching morphogenesis via growth factors and proteases [27, 28]. Fibroblasts also support epithelial cell survival and morphogenesis and synthesize several extracellular membrane components [29, 30]. ***Vascular and immune cells*** are intercalated throughout the mammary gland. Throughout pubertal MG development the lymphatic system extends in close association with epithelial outgrowth and blood vasculature [31]. During branching morphogenesis, macrophages and eosinophils are recruited to the TEBs to mediate invasion through the fat pad [32]. Macrophages are also required during involution for the clearance of dead cells and repopulation of adipocytes[33]. ***Adipocytes*** compose the majority of the adult, non-lactating murine MG. Adipocytes serve an endocrine function in the MG, regulating epithelial growth and function and mediating communication with other cell types. During pregnancy and lactation, adipocytes provide energy to the epithelial cells for the metabolically demanding milk production and cell contraction[34, 35].

The Terminal End Bud- The terminal end bud (TEB) is unique to the pubertal MG and is made up of two compartments. The outer compartment is a single layer of highly proliferative ‘cap cells’, differentiating into myoepithelial cells as the duct elongates [14] (Diagram 1, B). The inner compartment is the ‘body cell’ layer – a multi-cellular layer approximately 4-6 cells thick. This layer is made up mainly of luminal and alveolar progenitors that differentiate into mature luminal cells as the duct elongates[14]. Cap cells express several markers including keratin 5, smooth muscle actin and p63. Cap cells also express a stem cell specific isoform of SH2-containing inositol 5’-phosphatase (s-SHIP), though its role in the MG is unknown. The least differentiated and most proliferative cells are located in the bulbous region and the more differentiated and less proliferative are in the neck and subtending duct. TEBs invade through the fat

pad, undergoing regular bifurcation events until they reach the edge of the fat pad, when they regress completely to form blunt-ended ductal termini or small rounded buds[36]. Proper formation of the TEB requires estrogen, FGF10 and growth hormone. Wnt, SLIT and TGFB are also important for TEB morphology and function [37, 38].

1.2.3 Reproductive Development/Alveolar

During pregnancy, in preparation for lactation, the alveolar epithelium (luminal cells that have further differentiated) proliferates rapidly in response to circulating hormones. These include progesterone and prolactin, which are important to the developing secretory alveoli that are capable of producing milk [37, 39]. During lactation, the apically oriented luminal epithelial cells synthesize and secrete milk proteins into the lumen of the alveoli. Oxytocin release, caused by the suckling infant initiates contraction of the surrounding myoepithelial cells and progression of the milk through the ductal tree and to the nipple. During weaning, the stimuli for milk production are lost and the expanded epithelial compartment apoptos in an event referred to as 'involution'. This process is regulated by signal transducer and activator of transcription factor 3 (STAT3) and several other factors including STAT5a, STAT5b, sulfated glycoprotein-2 (SGP-2), interleukin-1b converting enzyme (ICE) and TGFβ3 [40]. 80 % of the mouse mammary epithelium is lost within 72 hours in a tightly regulated series of events [41]. The gland is remodeled by a number of proteases, of which metalloproteinase 3 (Mmp3) is the most prominent to return to the MG to 'resting' state. The epithelial tree returns to a state similar, but not identical to the resting virgin mammary gland (MG). Thus, the epithelial component and the surrounding tissue architecture go through a significant amount of remodeling during each pregnancy [42, 43].

1.3 Differences Between Human and Mouse Mammary Gland Development

Human and mouse mammary gland (MG) development are largely very similar, with a few key differences. *In utero*, there are two major differences between species. Firstly, in humans, the mammary milk line condenses into only a single pair of placodes that develop into MGs, in comparison to the 5 pairs in mice. Also, while male mice have irreversible condensation of the mammary mesenchyme between E13.5 and E15.5,

human MGs develop similarly in both males and females [44]. At birth, the main difference is that the human MG has several minor ductal networks joined at the nipple, and mice just have a single network [45]. During mouse development the mammary epithelium is composed of terminal end buds at the leading edge of the tissue driving invasion, humans do not have terminal end buds. The human adult mammary fat pad is composed of less adipocytes than that of a mouse and a much more complex and fibrous branching system [45, 46]. Finally, at the onset of pregnancy, the mouse MG develops lobules, whereas in human, the lobules are always present [47]

1.4 The Existence of Mammary Stem Cells (MaSCs)

Stem cells are characterized as having the ability to self-renew and give rise to more specialized differentiated cells, depending on the tissue. Several studies have been done with a goal to identify the mammary stem cell (MaSC), giving rise to both basal (myoepithelial) and luminal mammary epithelial cells. The classical methods used to identify MaSCs are transplantation assays and lineage tracing.

1.4.1 Transplantation Assays

In the 1950's serial transplantation assays, the original 'gold standard', hinted that MaSCs may exist in the mammay gland (MG). DeOme *et al.* revealed that portions of the normal mammary epithelia, when transplanted from donor mice to recipient fat pads cleared of endogenous tissue, could reproduce functional mammary epithelia [48] (Diagram 2, A). The cleared mammary fat pad allowed transplantation and growth of normal, pre-neoplastic and malignant mammary tissues, giving rise to either normal or tumorous tissue, respectively. In 1998, Kordon and Smith showed that a single cell could recapitulate the entire mammary epithelium, and this was further verified by Shackleton *et al.* in 2006 [49]. Not only could these cells reconstitute the MG, but progeny cells from the primary transplant tissue displayed serial transplantability at a clonal level [24, 50-54]. It was following the success of these studies that serial transplantation was regarded as the 'gold standard' assay for detection and validation of MaSCs [55, 56]. Several studies over time have used transplantation assays to determine the regenerative capacity of several cell types ([23, 48, 53, 57-62]). However, a major caveat to

using transplantation assays is whether removal of donor cells from their endogenous environment may drastically alter their normal behavior.

1.4.2 Markers used to isolate Mammary Stem Cells

Classically, fluorescent-activated cell sorting (FACS) is one of the key techniques used to isolate cells to determine reconstitution capacity of potential stem cells using transplantation assays [23, 24, 63-69]. The first step is usually to remove hematopoietic and endothelial cells with the markers CD45, Ter119, and CD31 (also named as Lin for lineage negative). Then, several studies use CD24 (heat-stable antigen), CD29 (β1-integrin), and CD49f (α6-integrin) as MaSC specific markers [23, 24, 64, 65, 68] . Other studies have reported additional markers used to identify MaSCs, including Lrp5/6 [63]Axin2 [65] CD1d [68] Lgr5 [66] and Procr [67]. It is notable that these factors are part of the Wnt signaling pathway, which has proved instrumental for MaSC renewal and expansion [70]. Also, ALDH1 [25, 71] *α*-SMA$^+$ and Myh11$^+$ myoepithelial cells have mammary repopulating unit capacity [60].

1.4.3 Lineage Tracing

Genetic lineage tracing is a powerful tool used for mapping cell fate. Lineage tracing allows direct observation of the progeny of a single cell under physiological conditions in the mouse [72]. In this technique, cell- or tissue- specific recombinase enzyme is expressed to specifically activate the expression of a conditional reporter gene. The reporter will permanently genetically label all the progeny of the marked cells [73] (Diagram 2, B). Because this technique does not perturb cells in their endogenous environment, it has taken over as the 'gold standard' to determine cell fate, and identify MaSCs. The *Cre-loxP* system is the preferred approach of genetic lineage tracing in mice, owing to its high recombination efficiency [74]. The *Cre-loxP* system functions by Cre recombinase expression under a cell-specific promoter, specifically activating a reporter in cells that express the promoter by excision of the STOP cassette in a *loxP*-STOP-*loxP* sequence upstream of the promoter. The *Cre-loxP* system can also be altered to make the Cre activity temporal and spatial. CreER is a mutated CRE whose Cre activity is inducible via the ER ligand, tamoxifen (Diagram 2, B). Several important

lineage tracing studies of mammary gland (MG) have been done. The keratin family of markers are commonly used to label stem cells in lineage tracing studies.

Van Keymeulen *et al*. showed that $K14^+$ (keratin14) cells from embryonic mice were multipotent, while $K14^+$ postnatal cells were unipotent, only contributing to basal-oriented myoepithelial cells throughout life. Van Keymeulen also found that two additional putative stem cell markers, $K5^+$ (keratin5) and $Lgr5^+$, preferentially labelled the myoepithelial stem cells. Lineage tracing of luminal cells showed that $K8^+$ (keratin8) cells were unipotent, giving rise only to luminal cells, which went on to differentiate into luminal and milk-producing cells during pregnancy. $K18^+$ (keratin18) cells also only labelled the luminal cells [75]. Overall, Van Keymeulen *et al*. showed that while multipotent stem cells exist within the basal layer during embryonic development, only unipotent progenitors exist within the luminal and myoepithelial cell lineages after birth.

However, Rios *et al.* reported that adult multipotent stem cells exist in the mammary gland. Lineage tracing depicted that K5, K14 and Lgr5 targeted long-lived multipotent stem cells that contributed to the expansion of both the luminal and myoepithelial lineages in the pubertal and adult mammary gland [76]. The discrepancies between these studies may be attributed to the studies using different lineage tracing mouse models and differences in labeling efficiency and specificity, or the administration of different concentrations of tamoxifen [25]. Leakiness has been a common problem with inducible systems. Wuidart *et al*. showed that the inducible CreER is not lineage specific, complicating the interpretation of data that suggest identification of multipotent stem cells in the adult mammary gland [77].

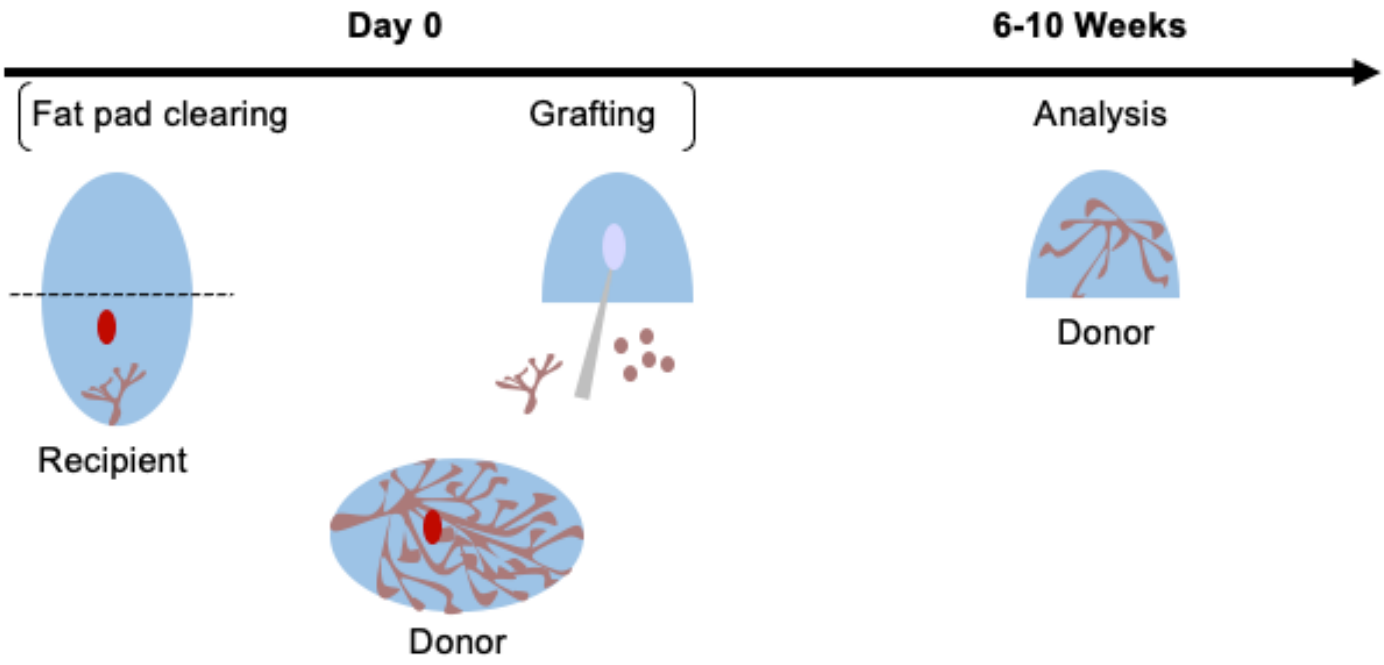
Members of the Notch family were also thought to mark MaSCs, including Notch1+, Notch2+, and Notch3+. Rodilla *et al*. found that Notch1 targeted multipotent stem cells in the embryonic mammary rudiment, but restricted their lineage potential to the ER- luminal lineage postnatally [21]. Sale *et al*. uncovered the existence of distinct Notch2+ progenitors that represented two novel mammary epithelial cell lineages, which they termed S (small) and L (large) cells. The S and L cells are morphologically, topologically, genetically, developmentally, and functionally distinct from classical luminal and myoepithelial cells [78]. Lafkas *et al*. revealed that $Notch3^+$ cells were a

luminal progenitor population that was highly clonogenic and transiently quiescent that gives rise to a ductal lineage [79].

Several additional lineage tracing studies have been done (stem cells or progenitors expressing Axin2 [80], Acta2 [60], WAP [81], Procr [67], Lgr5 [75, 76], Lgr6[77], Sox9[20, 77] prominin1[20], p63[77, 82] ,ER[22], and Blimp1[83]) that support both the existence of multi- and unipotent stem cells in the mammary gland and unveil the fate of mammary stem cells. Though studies have published confounding results, there is now a growing consensus that postnatal mammary development occurs from unipotent progenitor cells that separately generate the luminal and basal cell lineages forming the ductal tree.

Amongst the luminal cells are subpopulations of luminal progenitors that have been defined using lineage tracing [25]. Mature luminal and luminal progenitor cells can be differentiated by different cell marker combinations, one of them being CD49b-CD14-Sca1+ which corresponds to the mature luminal cell population [61]. The luminal progenitor population is defined by CD49b+CD14+Sca-1−Aldh1+ which marks the largest luminal progenitor subset [84]. This subset can then be further divided into committed hormone responsive estrogen receptor $(ER)^-$ or ER^+ luminal progenitor cells. These cells will then give rise to functionally distinct mature ER^- and ER^+ luminal cells. The ER^- luminal population is alveolar based and the ER^+ luminal cell population is ductal [84].

A. Transplantation Assay



B. CRE-Lox Lineage Tracing

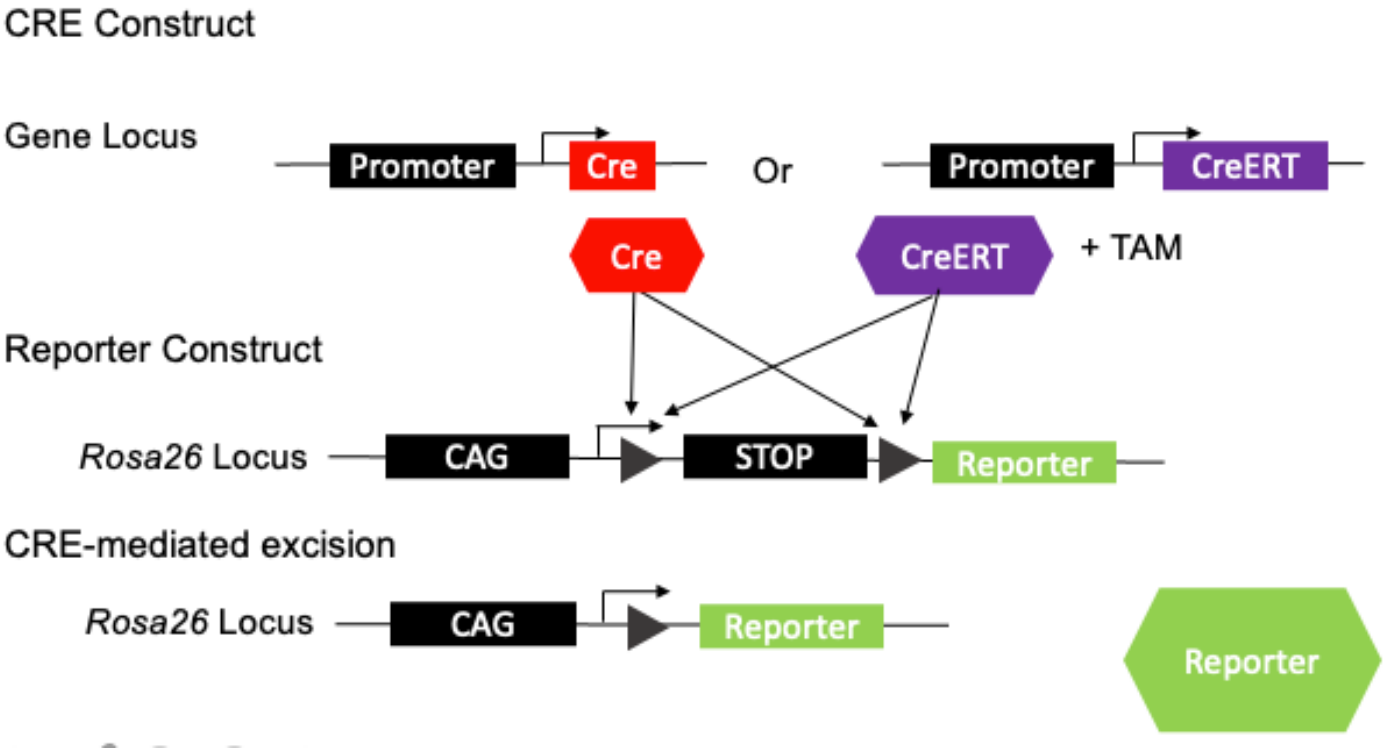


Diagram 2. Key Methods used to Identify Mammary Stem Cells. A. Transplantation Assay. In the recipient mouse, the portion of the fat pad containing rudimentary mammary gland is excised, as indicated by the dotted lines. The lymph node (red) is often used as an excision landmark. Epithelial fragments or single cells are isolated from the donor mouse and injected into the remaining (cleared) fat pad. After 6-20 weeks, the recipient mouse fat pads are isolated for outgrowth analysis. B. CRE-Lox Genetic lineage tracing. Schematic of the CRE-Lox system. The CRE-Lox system consists of two main elements. The CRE construct in which a gene for CRE or a modified tamoxifen-dependent form (CRE-ERT) is placed under control of a lineage-specific promoter. The reporter construct from which the expression of a marker protein occurs after CRE-mediated excision of a STOP cassette flanked by two LoxP sites. In the absence of CRE, the STOP cassette inhibits expression of the marker.

1.5 Cell Identity/State Switch/Reprogramming

The most common reasons for cell fate changes outside of development are for regeneration and repair. One of the mechanisms associated with regeneration is dedifferentiation which involves a terminally differentiated cell reverting back to a less differentiated state from within its own lineage[85]. In many studies mammary epithelial cell reprogramming has been observed. Early transplantation studies revealed that mammary epithelial cells retain multipotent potential postnatally [23, 60, 86]. Differentiated myoepithelial cells, which are unipotent *in vivo*, are also capable of reverting to multipotent stem cells, regenerating a mammary gland [24, 75, 80]. Human and mouse mammary luminal progenitors were shown to display “plasticity” under non-physiological conditions, suggesting that cell fate decisions among this lineage are not irreversible. It was shown that transplantation of both ER+ and ER− luminal progenitors into cleared mammary fat pads could not regenerate a mammary gland [61, 84, 87], but can be induced to display reversibility and reprogramming capacity. Transient co-expression of Slug and Sox9 or transient YAP/TAZ activation efficiently convert differentiated luminal cells into MaSCs *in vitro* [88, 89]. In addition, another study showed that polyoma middle T (PyMT) antigen or ErbB2 signaling activation in differentiated luminal cells can reprogram them into bipotency, giving rise to basal cells[90]. One of the transcriptional repressors that controls gene expression programs of cell state transitions between stem and progenitor cells and their differentiated progeny is SLUG (snai2) [88, 91-94]. A lingering question in the field is why can differentiated basal mammary epithelial cells exhibit stem cell behavior in culture or under other conditions when this behavior is not observed in the endogenous tissue under normal conditions? This behavior is only observed endogenously in response to DNA or tissue damage. Why do these cells retain this capability? One possible explanation is that normal reprogramming potential is repressed *in situ* under normal conditions but revealed in response to damage, or *in vitro,* and might contribute to breast cancer when this repression is perturbed.

1.6 The s-SHIP Marker

SHIP (SH2 domain-containing inositol 5′-phosphatase) was first identified as a tyrosine phosphorylated signaling protein that forms a complex with the adaptor proteins Shc and Grb2 during blood cell signaling [95-97]. The full-length SHIP structure consists of an N- terminal SH2 domain, a central inositol-polyphosphate with 5′ phosphatase enzymatic activity and an ~350 amino acid C-terminal region with numerous protein interaction motifs, which bind PTB, SH2, and SH3 domains. SHIP functions as a negative regulator by converting the plasma membrane PI3K-produced product, phosphatidylinositol(3,4,5)P_3, to phosphatidylinositol(3,4)P_2 [98]. SHIP was also later found to be expressed testes [99].

Along with full length SHIP and its spliced isoforms, also expressed from this locus was a 104 kDa protein lacking the N-terminal SH2 domain [100]. This product also differed from full length SHIP as its mRNA begins with a 44-nucleotide region not previously found in any murine SHIP cDNA. This novel 44-nucleotide region was designated the stem-SHIP region (SSR). The 104 kDa protein arises from an internal promoter within the *ship* locus and is not generated by splicing[101]. Tu *et al.* proposed that there was a distinct promoter within intron 5 of the ship 1 gene. This was based on the observation that the SSR included 44 nucleotides from intron 5 at its 5' end just before the *ship* sequence similarity from introns 6 to 27 [98]. The protein generated from this promoter was named s-SHIP, with the "s" signifying stem cell as this product was found to be specifically expressed in embryonic stem cells and hematopoietic stem cells [100]. Rohrschneider made a transgenic mouse (named Tg11.5kb-GFP) expressing GFP behind this s-SHIP promoter to facilitate analysis of its expression in other mouse tissues [98]. Bai and Rohrschneider used the mouse in an attempt to identify stem cells in the mammary gland [38].

They found the expression of GFP in cap cells of mammary buds during embryonic development, in the cap cells of TEBS during puberty, and within basal alveolar bud cells in pregnancy, but not in mature basal cells. GFP+ cells exhibited self-renewal and regenerative capabilities, giving rise to a functional MG upon transplantation. While these studies suggest that cap cells are multipotent stem cells,

endogenous studies using lineage tracing confirmed that they are unipotent stem cells, giving rise only to the basal cell population and not luminal cells [38].

1.7 Overview of work presented in this dissertation

In vivo, mature mammary myoepithelial cells are unipotent, showing no evidence of multipotency; but when isolated from their endogenous environment, cultured, or transplanted into a recipient mouse, they spontaneously convert to a stem-like state and can convert into luminal-like cells [102,103]. The mechanism enabling transdifferentiation of mature myoeptithelial cells to the luminal lineage has not been resolved. We addressed this problem by asking whether myoepithelial cells switch lineages directly into the luminal lineage or whether they are required to pass through an intermediate, stem-like state. What factors are involved in stabilizing cell state transitions? I found that myoepithelial cells preferentially de-differentiate towards a progenitor state before changing lineage and that the transcription factor RFX3 plays a critical role in the stabilization of the mammary basal cell lineages.

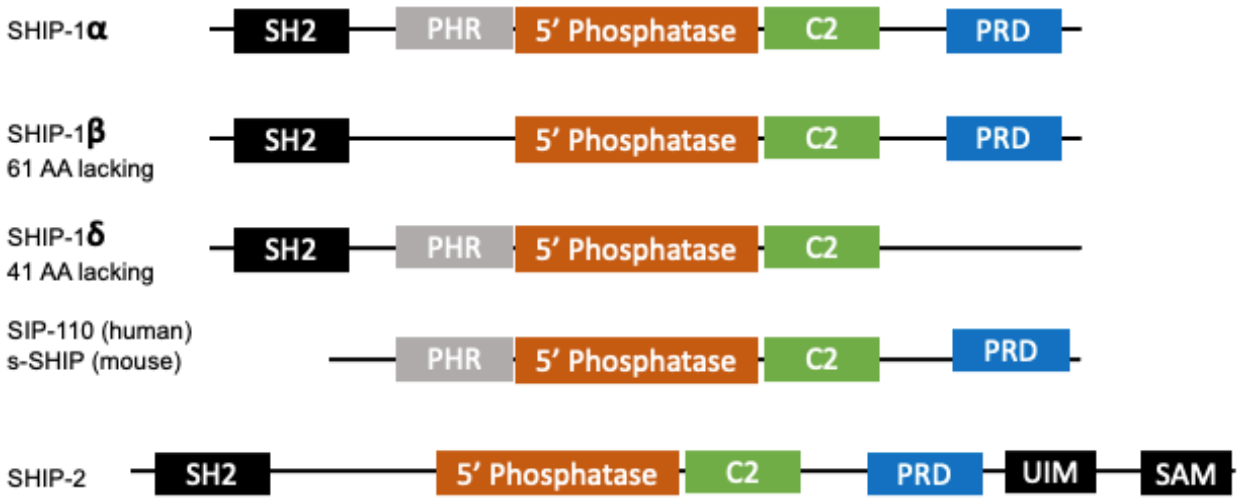

SHIP-1α
SH2
PHR
5' Phosphatase
C2
PRD
SHIP-1β
61 AA lacking
SH2
5' Phosphatase
C2
PRD
SHIP-1δ
41 AA lacking
SH2
PHR
5' Phosphatase
C2
SIP-110 (human)
s-SHIP (mouse)
PHR
5' Phosphatase
C2
PRD
SHIP-2
SH2
5' Phosphatase
C2
PRD
UIM
SAM

Diagram 3. s-SHIP. Diagram described in text.

II. TRANSCRIPTION FACTOR RFX3 STABILIZES MAMMARY BASAL CELL IDENTITY

Adapted from Tross *et al.* Transcription Factor RFX3 Stabilizes Mammary Basal Cell Identity (Tross *et al*. 2021)

2.1 Abstract

The myoepithelial cell compartment of the murine postnatal mammary gland is generated from basal cap cells in the terminal end bud and maintained by self-renewal. Transdifferentiation to the luminal lineage does not normally occur, but can be induced by DNA damage, luminal cell death or transplantation into a recipient mammary fat pad. Myoepithelial cells cultivated in vitro can also transdifferentiate towards the luminal lineage. Little is known about the molecular mechanisms and gene regulatory networks underlying this plasticity. Using a transgenic mouse (Tg11.5kb-GFP) that marks cap cells with GFP, we discovered that mature myoepithelial cells placed in culture begin to express GFP within ~24 hrs and later express the Keratin 8 (K8) luminal marker. Cell tracking showed that most K8+ cells arose from GFP+ cells, suggesting that myoepithelial cells de-differentiate towards a progenitor state before changing lineage. Differential gene expression analysis, comparing pure GFP+ cap cells with mature myoepithelial cells, identified multiple transcription factors that iRegulon predicted might regulate the myoepithelial to cap cell transition. Knockout of one of these genes, Regulatory Factor 3 (Rfx3), significantly reduced the population of GFP+ cells and increased differentiation to the K8+ luminal lineage. Rfx3 knockout also reduced mammosphere growth and mammary gland regeneration efficiency in a transplantation assay but had no effect on proliferation in vitro. Together, these data support a key role for Rfx3 in the stabilization of the mammary basal cell lineages.

2.2 Introduction

The mammary gland undergoes dynamic changes that contribute to the development, structure and function of the organ. In utero, development is initiated by bipotent mammary stem cells that give rise to the mammary rudiment, which remains quiescent until puberty [64, 75, 104]. At the onset of puberty, the estrus cycle drives expansion of the rudiment into a branching network of ducts that fill the mammary fat

pad [47]. Mature ducts consist of an inner layer of polarized luminal cells (which express the keratin K8), surrounded by spindle-shaped basal, myoepithelial cells that are marked by K14 expression. Outgrowth occurs primarily from terminal end buds at the tip of each duct, filled with immature, proliferating luminal body cells enveloped by a single basal layer of cap cells, which are a metastable population of myoepithelial progenitors [82, 105, 106]. These large end buds, and the cap cells, disappear when the ducts reach the edges of the fat pad [36].

At homeostasis the basal and luminal lineages are lineage-restricted and maintained separately by self-renewal [60, 75, 77]. Each population appears to be stable; however, we discovered that DNA damage to the mammary gland triggers transdifferentiation of myoepithelial to luminal cells. Ablation of luminal cells in situ by diphtheria toxin also activates transdifferentiation, possibly through multiple signaling pathways [107, 108].

Interestingly, this same lineage switch appears to be induced by cultivating myoepithelial cells in vitro or by their transplantation into cleared mammary fat pads in a recipient mouse. However, almost nothing is known about the underlying mechanisms and gene regulatory networks that determine the stability of the lineages or their reversion to a stem-like state. To begin to untangle the processes driving myoepithelial cell transdifferentiation, we employed a transgenic mouse strain that expresses GFP from the promoter of the s-SHIP gene (11.5kb-GFP), which is expressed in embryonic and hematopoietic stem cells, and several other progenitor/stem cell lineages [98]. It also marks cap cells but is silent in mature myoepithelial cells [38].

After plating purified myoepithelial cells in culture, we detected both GFP+ cells and cells expressing the luminal marker K8 arising within 24 – 48 hrs. Single cell RNAseq also identified a population of luminal cells after 96 hrs culture. Cell tracking revealed that most K8+ cells arise from the GFP+ population, suggesting that myoepithelial cells de-differentiate before undergoing transdifferentiation towards the luminal lineage. Bulk RNAseq of purified cap cells (GFP+) and mature myoepithelial cells identified several transcription factors that were preferentially expressed in the cap cell population, which by iRegulon analysis are predicted to control multiple other genes induced in the cap cells. For one such factor, Rfx3 (regulatory factor 3), its ablation by

CRISPR-Cas9 editing in mature myoepithelial cells caused a significant reduction in the frequency of GFP+ cell formation, with increased formation of K8+ cells. Rfx3 knockout also reduced mammosphere formation and reduced ductal outgrowth after transplantation. Rfx3 ablation in pure cap cells also increased the formation of K8+ cells.

We propose that Rfx3 normally stabilizes basal cell identities, so that in its absence the myoepithelial and cap cell populations more rapidly convert to the luminal lineage, thereby losing the ability to regenerate mammary glands in transplantation assays.

2.3 Results

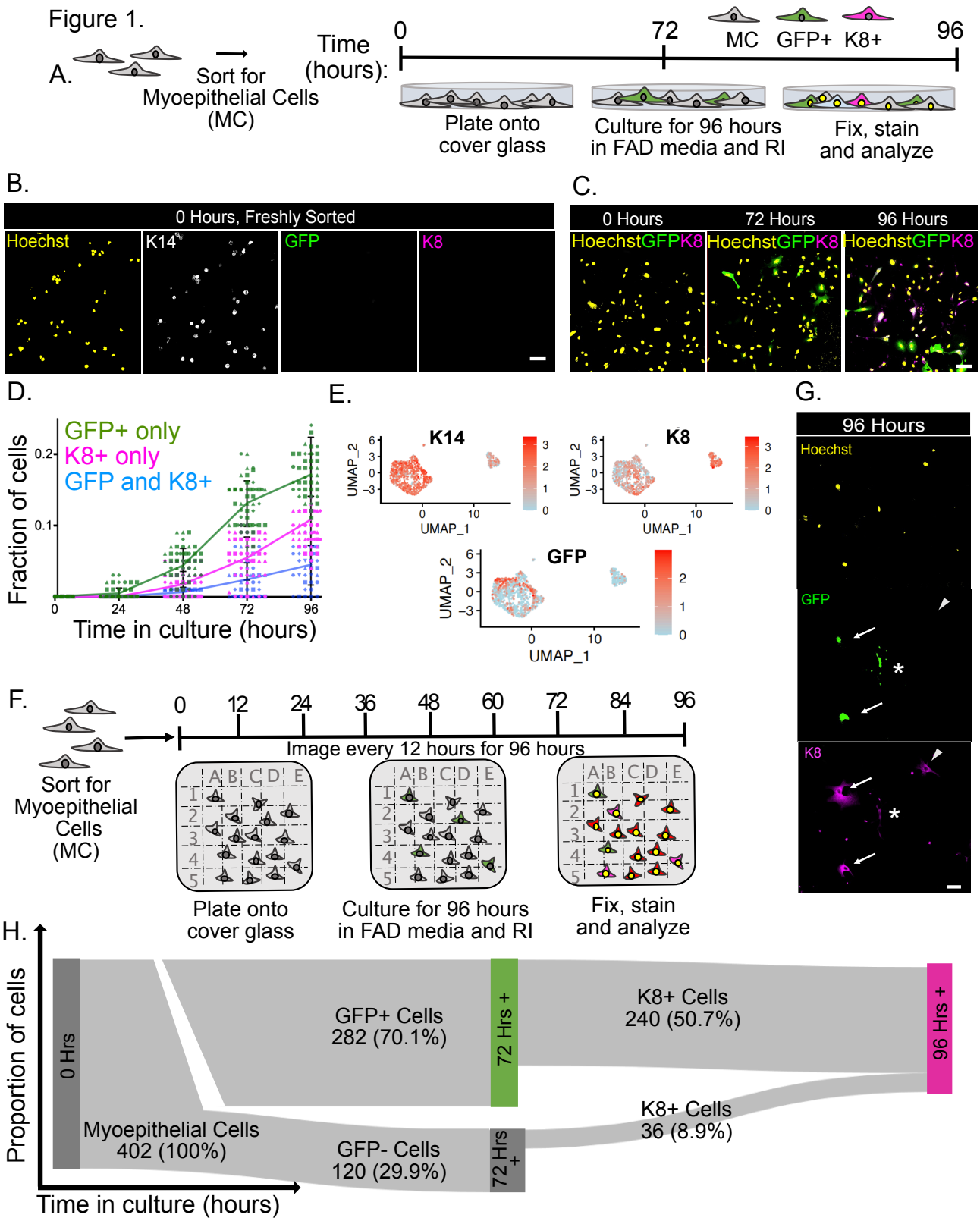
2.3.1 Myoepithelial cells in culture can transdifferentiate towards cap cell and luminal cell lineages

Lineage-restricted mammary myoepithelial cells in situ can transdifferentiate into luminal cells in response to DNA damage or DTA-mediated ablation of the luminal population, and are capable of regenerating entire mammary glands after transplantation into the fat pads of recipient mice [24, 38, 75, 107, 109].

To probe mechanisms that might regulate this lineage switching, we isolated mammary glands from the 11.5kb-GFP transgenic mouse, which expresses GFP specifically in cap cells of terminal end buds but not in mature myoepithelial cells (Supplementary Fig S1 A). Myoepithelial cells were then sorted by FACS (Supplementary Figure S1 B) and cultured for 96 hrs on laminin-coated cover glasses in FAD media supplemented with ROCK inhibitor (Fig 1 A). Cell cultures were fixed every 24 hrs over 4 d, stained for the luminal marker K8 and imaged to determine whether cells expressed GFP and/or K8. GFP+ cells began to appear after 24hrs and increased over time. K8+ cells were detectable within 48 hrs. By 96 hrs in vitro ~20% of cells were GFP+, ~10% of cells expressed K8 and ~5% of cells were double positive (Fig 1 B,C,D).

To determine if this K8+ population represents the true luminal lineage or is a de-differentiated cell type that happens to express this keratin but not other luminal markers, we performed single cell RNAseq (scRNAseq) (Supplementary Fig S1 C). As an initial test we purified myoepithelial, luminal and cap cells from Tg11.5kb-GFP

Figure 1.


A.
Sort for Myoepithelial Cells (MC)
Time (hours):
0
72
96
MC
GFP+
K8+
Plate onto cover glass
Culture for 96 hours in FAD media and RI
Fix, stain and analyze
B.
0 Hours, Freshly Sorted
Hoechst
K14
GFP
K8
C.
0 Hours
72 Hours
96 Hours
HoechstGFPK8
D.
GFP+ only
K8+ only
GFP and K8+
Fraction of cells
Time in culture (hours)
E.
K14
K8
GFP
UMAP_1
UMAP_2
G.
96 Hours
Hoechst
GFP
K8
F.
Sort for Myoepithelial Cells (MC)
Image every 12 hours for 96 hours
Plate onto cover glass
Culture for 96 hours in FAD media and RI
Fix, stain and analyze
H.
Proportion of cells
Time in culture (hours)
0 Hrs
Myoepithelial Cells 402 (100%)
GFP+ Cells 282 (70.1%)
GFP- Cells 120 (29.9%)
72 Hrs +
K8+ Cells 240 (50.7%)
K8+ Cells 36 (8.9%)
96 Hrs +

mouse mammary glands, then pooled equal numbers of each population and processed them for inDrop scRNAseq. Uniform manifold approximation and projection (UMAP) identifies 4 separate clusters, as expected (Supplementary Fig S1 D). We then isolated mature myoepithelial cells and grew them in culture for 96 hrs prior to encapsulation and scRNAseq. In this case UMAP revealed a large cluster that expresses the myoepithelial marker K14 (and other myoepithelial cell markers (Fig 1 E, Supplementary Fig S2 A,B)) and a separate cluster mostly positive for K8 and other luminal markers (Fig 1 E, Supplementary Fig S2 C,D), while GFP+ cells partially overlapped with the K14+ population (Fig 1 E). The large myoepithelial cluster is composed of clusters of cells that may be in different cell states of conversion (Supplementary Fig S2 G). These data suggest that the GFP+ cells do not fully de-differentiate into cap cells, but that transdifferentiation into the luminal lineage is relatively robust.

We considered the possibility that the luminal cells could be contaminants from the original FACS sort, but there are several reasons that confirm that this is not the case. Firstly, luminal cells proliferate rapidly in culture. If a contaminating luminal cell(s) were cultured with the basal cells, the proportion of luminal cells would be much greater than the basal cells resulting in a larger UMAP projection cluster. Secondly, the experimental luminal cells express moderate levels of K14, which is not the case for true luminal cells. Lastly, in freshly sorted luminal cells there are ESR1+ and ESR1- cell clusters, and this is not the case with the experimental luminal cells (Supplementary Fig S2 C,D).

We next asked if the GFP+ state is an obligatory intermediate between the myoepithelial and luminal identities, or whether different myoepithelial cells independently adopt either the luminal state or the GFP+ state. To distinguish these hypotheses, we used cell tracking. Freshly isolated myoepithelial cells from Tg11.5kb-GFP mice were plated sparsely on gridded coverslips and imaged every 12 hrs for 96 hrs then fixed and stained for K8 (Fig 1 F,G). Each cell was then traced backwards in time to determine if/when cells became GFP+ and if they progressed or not to become K8+. Data were visualized as a Sankey plot (Fig 1 H). Of 402 cells that were analyzed, 70% began to express GFP within 72 hrs and of this population ~85% eventually

Figure 1. Myoepithelial cells from transgenic mice (TG11.5kb-GFP) express GFP prior to expressing luminal marker during transdifferentiation

A. Diagram of experimental approach used to follow transdifferentiation of myoepithelial cells (MC) into luminal marker-expressing K8+ cells.
B. Representative image of freshly sorted myoepithelial cells at 0 hrs in culture, expressing only myoepithelial cell marker K14. Scale bar = 100 μm
C. Representative images of myoepithelial cells cultivated in FAD medium with ROCK inhibitor Y-27632 (RI) over 96 hrs. Scale bar = 100 μm
D. Quantification of cap cell and luminal cell marker expression during culture of myoepithelial cells with Y-27632. A fraction of cells become GFP+, K8+ or express both markers, n= 5 biological replicates (each a different shape), 10 fields/experiment. Error bars = Mean +/- SD.
E. UMAP projections of myoepithelial cells after 96 hrs culture. Single Cell RNA sequencing was used to observe myoepithelial cell gene expression changes into GFP+ and K8+ cells over 96 hours.
F. Diagram of experimental approach to track transdifferentiation of single myoepithelial cell (MC) conversion in vitro toward the luminal lineage (K8+).
G. Representative images of myoepithelial cell transdifferentiation at 96 hrs, and differential conversion paths. Images only showing GFP and K8 expression. Arrows indicate cells that express GFP prior to expressing luminal marker K8. Arrowheads point to cells that express K8 without ever expressing GFP, and the star indicates a cell that becomes GFP+ but did not become K8+. Scale bar = 100 μm.
H. Sankey Plot displaying cell gene expression decisions in vitro over time. Majority of cells express the cap cell marker GFP prior to expressing K8 during luminal transdifferentiation.

expressed K8. In contrast, only 30% of the GFP- cells eventually expressed K8 (Fig 1 H). It is possible that some of these cells transiently expressed GFP and were missed during the periodic imaging. Moreover, we noted that transdifferentiation of this sparse culture of myoepithelial cells was much more efficient than occurs when the cells are plated more densely, as in Fig 1. The reason for this difference remains obscure, but we can conclude, nonetheless, that the principal pathway for transdifferentiation involves an initial, partial dedifferentiation towards a progenitor, cap cell state prior to converting to the luminal lineage.

2.3.2 RNAseq identifies multiple genes that are differentially expressed between cap cells and mature myoepithelial cells

Myoepithelial cells and GFP+ cap cells were isolated from 5 weeks old female Tg11.5kb-GFP mice and subjected to RNAseq. Principle component analysis confirmed that like-samples had the least variance and clustered together, clearly segregating the two cell types (Supplementary Fig S3 A). We validated the RNAseq data by qRT-PCR of chosen mRNAs that were either less expressed or more highly expressed in cap cells as determined from the RNAseq (Supplementary Fig S3 B). We were also able to validate differentially expressed genes in the single cell RNA seq data, a few shown in Supplementary Fig S2 E and Supplementary Fig S2 F.

A volcano plot identified multiple genes that were up or down regulated in the GFP+ cap cells versus myoepithelial cells (Fig 2 A). Using a p-value of ≤ 0.05 and a fold change of ≥ 2, 63 genes were significantly upregulated in cap cells and 106 were downregulated, as compared to mature myoepithelial cells. Unsupervised hierarchical clustering also identified a cap cell gene signature corresponding to GO terms associated with organ, tissue and embryonic development, which is consistent with the status of cap cells as progenitors (Fig 2 B, C). TheTg11.5kb-GFP transgenic line drives GFP from an alternate promoter for the SHIP-1 (INPP5D) gene, which expresses a shorter variant of SHIP-1 called s-SHIP [98]. Although most of the transcribed s-SHIP sequence is identical to that of full-length SHIP-1, there is a unique 42 nt sequence in the 5' UTR that differentiates it from the canonical SHIP-1 mRNA sequence. We expected that GFP+ cells would also express this s-SHIP sequence. Indeed, analysis of the RNAseq data identified about 300 (+/- 200 fkpm; n=6) reads of the unique s-SHIP

sequence in purified GFP+ cells and zero reads in the mature myoepithelial cells (Fig 2 D).

Figure 2.

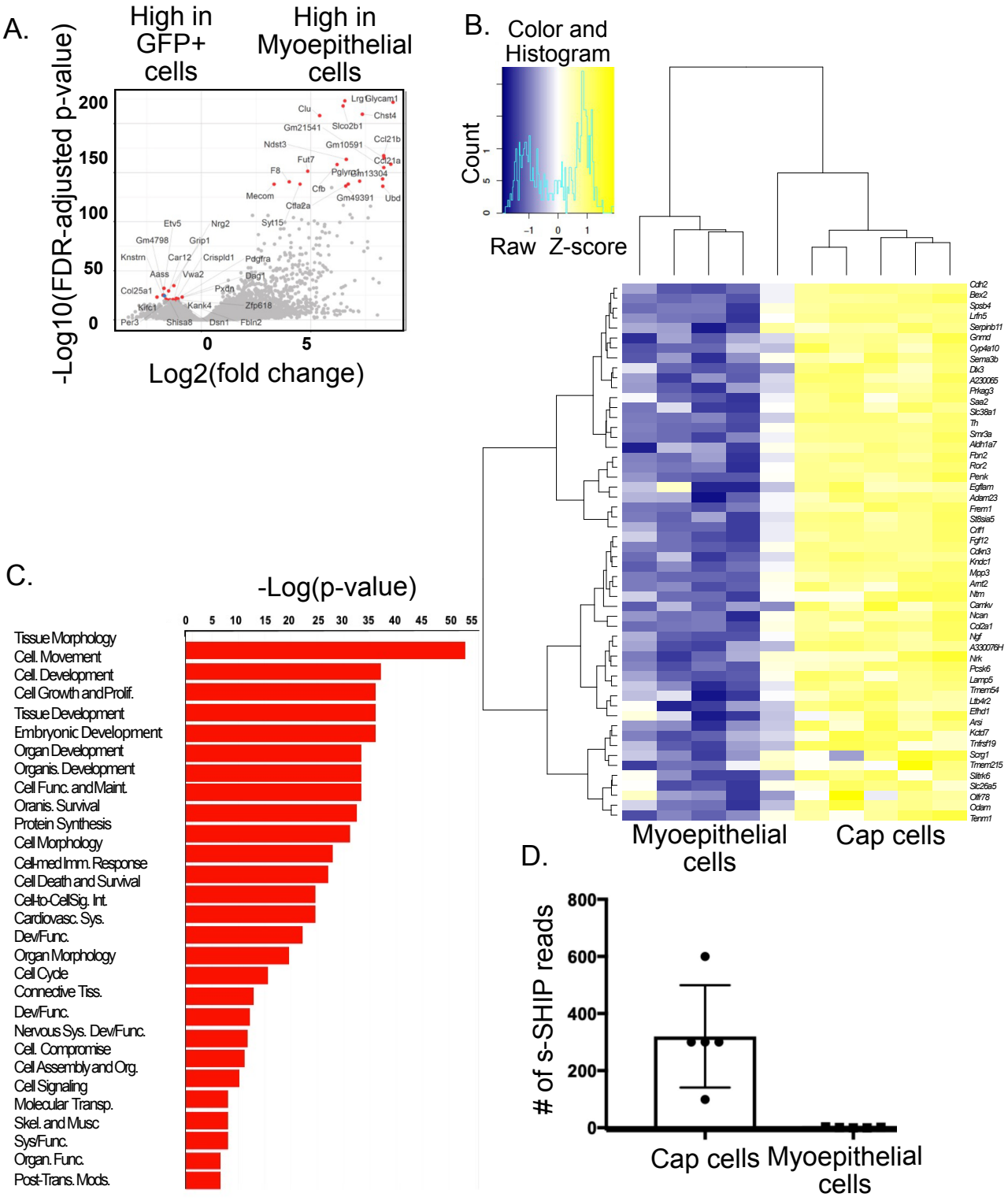

A.
High in GFP+ cells
High in Myoepithelial cells
-Log10(FDR-adjusted p-value)
Log2(fold change)
B.
Color and Histogram
Count
Raw Z-score
Myoepithelial cells
Cap cells
C.
-Log(p-value)
Tissue Morphology
Cell. Movement
Cell. Development
Cell Growth and Prolif.
Tissue Development
Embryonic Development
Organ Development
Organis. Development
Cell Func. and Maint.
Oranis. Survival
Protein Synthesis
Cell Morphology
Cell-med Imm. Response
Cell Death and Survival
Cell-to-Cell Sig. Int.
Cardiovasc. Sys.
Dev/Func.
Organ Morphology
Cell Cycle
Connective Tiss.
Dev/Func.
Nervous Sys. Dev/Func.
Cell. Compromise
Cell Assembly and Org.
Cell Signaling
Molecular Transp.
Skel. and Musc
Sys/Func.
Organ. Func.
Post-Trans. Mods.
D.
# of s-SHIP reads
Cap cells
Myoepithelial cells

Figure 2. Differential gene expression in myoepithelial cells vs. cap cells

A. Volcano plot displaying differential gene expression with the top 20 most highly expressed genes for each sample in red. Genes filtered with a p-value <0.005.
B. Heatmap displaying ~50 of the top genes differentially expressed between myoepithelial cells and cap cells. Genes filtered with a p-value < .05 and fold change > 2.
C. GO ontology graph of biological functions of genes most highly expressed in cap cells. Genes filtered with a p-value of < 0.05 and fold change > 2.
D. Dot plot displaying read count values from RNA sequencing data which show that s-SHIP is differentially expressed between GFP+ cap cells and myoepithelial cells. Each dot is a biological replicate. n=5. Error bars = Mean +/- SD.

2.3.3 Transcription factors predicted to regulate the GFP+ cap cell signature

To determine what transcription factors are regulating the expression of the genes highly expressed in GFP+ cap cells we performed Gene Regulatory Network Analysis using iRegulon, which is a computational method developed to reverse-engineer a transcriptional network underlying a co-expressed gene set using *cis*-regulatory sequence analysis [110]. This program bases its predictions on databases of ~10,000 TF motifs and 1000 Chip-Seq data sets. The analysis identified 15 transcription factors predicted to regulate the majority of genes most highly expressed in GFP+ cap cells (Fig 3 A). We determined from the RNAseq data that of these transcription factors RFX3, MTF1, PDX1, and GATA2 are consistently upregulated in cap cells (Fig 3 B). RFX3 expression was also observed in the myoepithelial cell cluster in the single cell RNA seq data (Supplementary Fig S2 H).

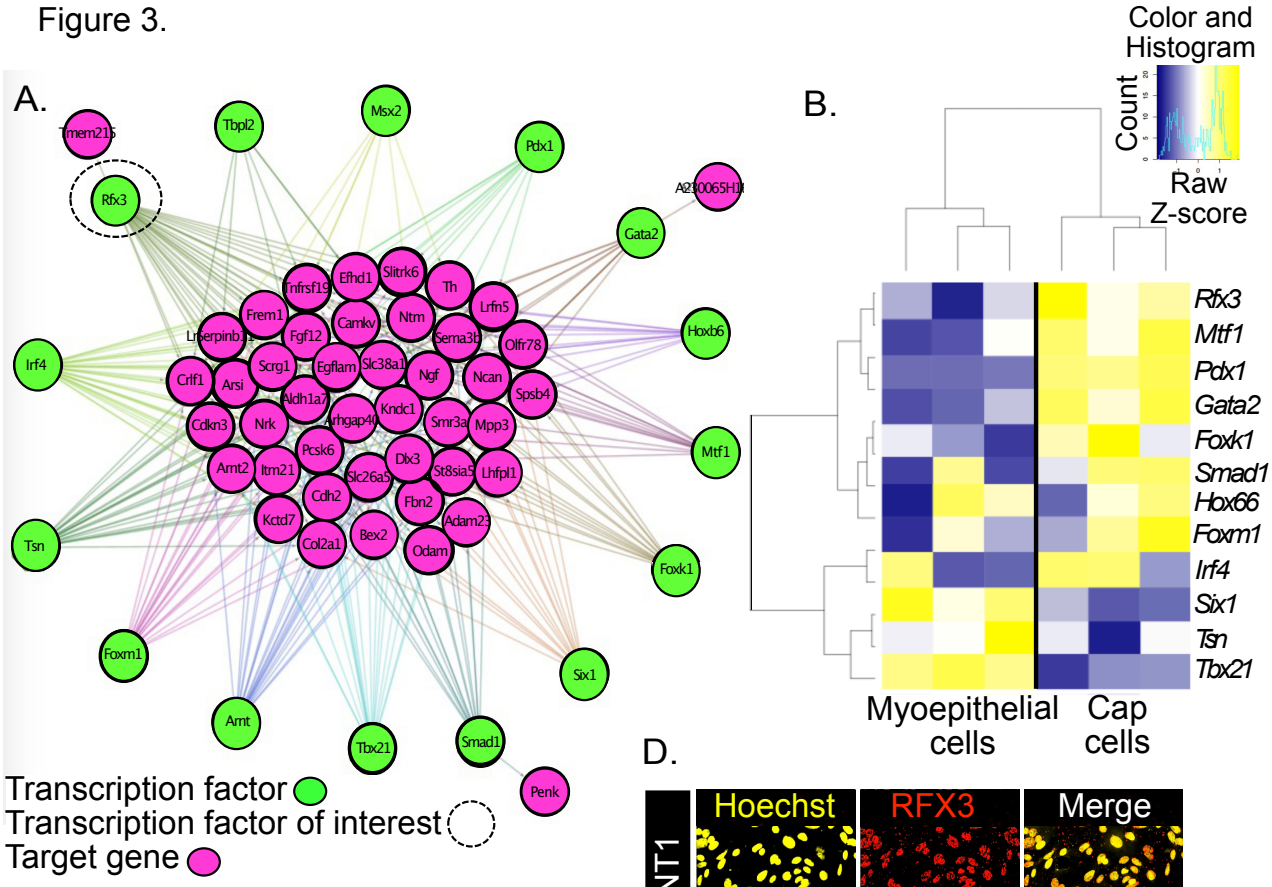
2.3.4 RFX3 stabilizes basal cell identities

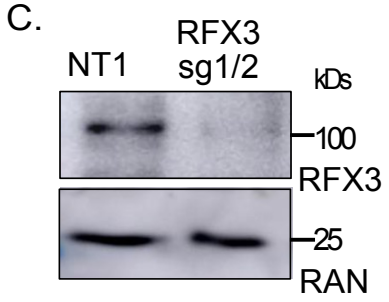
RFX3 is a key factor in ciliogenesis [111]. Primary cilia have been reported to regulate branching morphogenesis during mammary gland development [112]. It was also shown that cilia are enriched in mammospheres and cilia defects decreased the ability to form mammospheres. Loss of IFT88, a target of RFX3, reduces mammosphere growth [113].

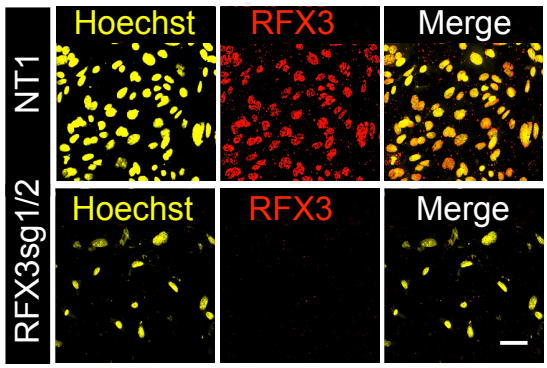
To determine whether RFX3 is an important regulator of the conversion of myoepithelial cells into GFP+ cap cells, we used Cas9-mediated gene editing with 2 gRNAs to target and splice out the RFX3 DNA binding domain in purified mature myoepithelial cells from the Tg11.5kb-GFP mice. A non-targeting gRNA (NT1) was used as the negative control. A high efficiency of knockout was confirmed on uncloned cultures of the infected cells, as compared to cells transduced with the NT1 gRNA, by immunoblot and immunofluorescence (Fig 3 C,D,E). RT-PCR of RFX3 and of two known targets of RFX3, IFT88 and DYNC2U, were also significantly diminished by the knockout (Fig 3 F). The knockout cells were also grown in culture for 96 hrs then stained for K8. The number of GFP+ cells in the RFX3- culture was significantly lower than in the NT1 culture, while K8+ cell numbers were increased (Fig 4 A,B,C). Moreover, when we performed time-lapse imaging and traced isolated myoepithelial

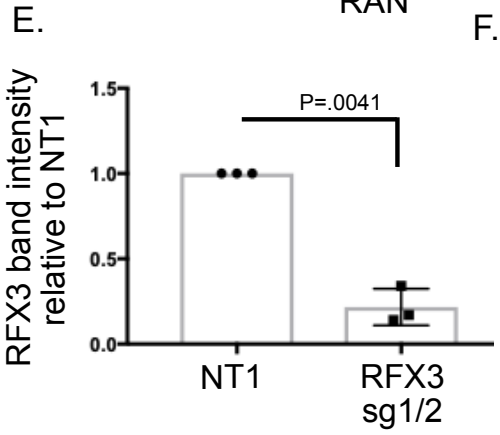
cells back to when they were initially plated, fewer of the RFX3- cells became GFP+ (30% versus 60% for the

Figure 3.


A.
Rfx3
Tbpl2
Msx2
Pdx1
Gata2
Hoxb6
Mtf1
Foxk1
Six1
Tbx21
Smad1
Penk
Foxm1
Arnt
Tsn
Irf4
Transcription factor
Transcription factor of interest
Target gene
B.
Color and Histogram
Count
Raw Z-score
Rfx3
Mtf1
Pdx1
Gata2
Foxk1
Smad1
Hox66
Foxm1
Irf4
Six1
Tsn
Tbx21
Myoepithelial cells
Cap cells


C.
NT1
RFX3 sg1/2
kDs
100
RFX3
25
RAN


D.
NT1
RFX3sg1/2
Hoechst
RFX3
Merge


E.
RFX3 band intensity relative to NT1
P=.0041
1.5
1.0
0.5
0.0
NT1
RFX3 sg1/2

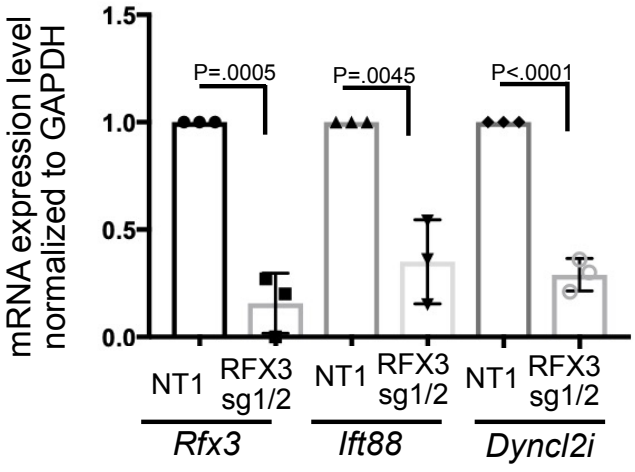

F.
mRNA expression level normalized to GAPDH
P=.0005
P=.0045
P<.0001
1.5
1.0
0.5
0.0
NT1
RFX3 sg1/2
Rfx3
Ift88
Dyncl2i

Figure 3. Gene Regulatory Network Analysis identifies Regulatory Factor X3 (RFX3) as a potential upstream regulator of cap cell gene expression

A. iRegulon Web showing the top 15 transcription factors (green) predicted to regulate the differentially expressed genes (DEGs) (pink) most highly expressed cap cells. RFX3 was a transcription factor of interest (black dashed circle).
B. Heat map of differentially expressed transcription factors predicted to regulate genes most highly expressed in cap cells.
C. Western blot confirming RFX3 KO in commaD Beta cells. anti-RAN is used as the loading control.
D. Immunocytochemistry confirming RFX3 knockout (KO) in primary cap cells. NT1 is the non-targeting 1 control lentivector and RFX3sg1/2 are the single guide lentivectors 1 and 2 used to excise the RFX3 DNA binding domain. Scale bar = 100μm
E. Quantification of Western blot. RFX3 band intensity relative to NT1. P-value = .0041, statistics calculated using unpaired t-test. Error bars = Mean +/- SD.
I. qRT-PCR confirms RFX3 loss results in a decrease in target gene (IFT88 and DYNC2LI) expression. P-value = .0005, .0045, <.0001, statistics calculated using unpaired t-test. Error bars = Mean +/- SD.

NT1 control cells), and – surprisingly - more GFP- cells transdifferentiated into the K8+ luminal lineage (Fig 4 D). These data suggest a reduced frequency of de-differentiation towards the cap cell lineage and/or increased transdifferentiation into K8+ luminal cells. We also observed that there was a delay in the onset of GFP expression, and more cells reverted to a non-GFP+ state (Fig 4 E).

To further investigate mechanism, we repeated the experiment using RFX3 knockout in purified GFP+ cap cells. As shown in Fig 4 F,G,H the loss of RFX3 reduced the GFP+ population and increased the fraction of K8+ cells. This result suggests that RFX3 normally functions to stabilize cap cell identity and prevent transdifferentiation into the luminal lineage. Because luminal cells are not competent to regenerate mammary glands, a prediction of this model is that deletion of RFX3 would reduce the capacity of basal cells (cap or myoepithelial cells) to form mammospheres, or to regenerate mammary glands in vivo after transplantation into the cleared fat pads of recipient mice.

Indeed, we observed a significant reduction in mammosphere growth of myoepithelial cells grown in 3D culture (Fig 5 C,D). Importantly, this effect was not the result of a decreased ability of RFX3 knockout cells to proliferate, as we could detect no significant difference in BrdU incorporation into these cells versus an NT1 control that expresses a non-targeting gRNA (Fig 5 A,B).

We next tested the efficiency of mammary gland regeneration in a transplantation assay. Five, ten or twenty thousand unsorted mammary epithelial cells were transplanted into the cleared fat pads of isogenic recipient mice. After 8 weeks the mice were euthanized and the glands were removed, fixed, and stained with Carmine Alum. Analysis of these whole mounts showed that mammary gland outgrowth was substantially reduced for transplanted RFX3-negative cells compared to the NT1 control (Fig 5 E,F,G). We infer that the RFX3 transcription factor functions to stabilize the identities of basal cell populations, which is necessary for efficient ductal regeneration.

2.4 Discussion

It has become clear in recent years that the concept of distinct, stable cell states that change in a uni-directional manner during development is incorrect, and that most cells in most animal tissues are in metastable states that – often in response to stress –

can revert to earlier states or change lineages. For example, following injury, differentiated airway epithelial cells and intestinal epithelial Dll1+ secretory progenitors can become functional stem cells [114, 115]. Injury to the heart in zebrafish results in dedifferentiation and proliferation of cardiomyocytes [116]. Differentiated cells can also be forced to revert to a stem cell state by the over-expression of various transcription factors, most famously for mesenchymal fibroblasts, which in the context of the Yamanaka factors are reprogrammed into pluripotent stem cells. Mature pancreatic acinar cells can be converted directly into beta cells by expression of Ngn3, Pdx1 and Mafa transcription factors [117]; and mammary luminal cells can be driven to a multipotent stem-like cells by expression of Sox9 and Slug [88].

On the other hand, mature mammary myoepithelial cells, which in vivo normally form a discrete, self-renewing population with no evidence of multipotency, can spontaneously convert to a stem-like state after isolation from the mammary gland and grown in culture, or transplanted into a recipient fat pad [60, 75]. This conversion suggests that the micro-environment plays an important role in stabilizing the differentiated state of mature myoepithelial cells, a hypothesis consistent with recent demonstrations that DNA damage to the mammary gland or the ablation of luminal cells by diphtheria toxin in vivo results in transdifferentiation of myoepithelial cells to the luminal lineage.

However, the mechanism for this conversion remains unclear. Do myoepithelial cells directly switch lineage, or do they pass through an intermediate, stem-like state? And what gene regulatory network controls lineage stability versus instability?
To address these questions, we examined the transdifferentiation of mature mammary myoepithelial cells towards the luminal lineage when they are isolated from the mammary gland and grown in culture. Single cell RNAseq showed that over a period of 96 hrs, a fraction of the myoepithelial cells take on a luminal identity in vitro, and cell tracking showed that the majority of such cells pass through an intermediate state in which they express a marker of cap cells, which are myoepithelial cell progenitors. Differential gene expression analysis and regulatory network analysis using iRegulon identified several transcription factors that potentially control expression of multiple genes that are upregulated in cap cells. One of these, RFX3, proved to be necessary for

the stability of basal (cap and myoepithelial) cell identities. Knockout of RFX3 strongly reduced the number of cells that expressed the cap cell GFP marker.

Moreover, while loss of RFX3 had no impact on proliferation of myoepithelial cells in vitro, it significantly reduced mammosphere growth and reduced mammary ductal outgrowth in a transplantation assay. These data suggest that RFX3 is not required for cap cell identity, or for conversion of myoepithelial cells to the luminal lineage, but instead plays a role in stabilizing cell identity.

We speculate that state stabilization is an important process in development, and that multiple transcription factors, regulated by signaling from the micro-environment, might operate in different contexts and tissues to ensure phenotypic stability. We also speculate that the mechanism behind the decreased phenotypic instability is due to a loss of *Sonichedgehog* signaling, which is critical for normal mammary development. A loss of RFX3 results in a decrease in Sonichedgehog target genes preventing normal signaling and myoepithelial cell phenotypic stability.

Figure 4.

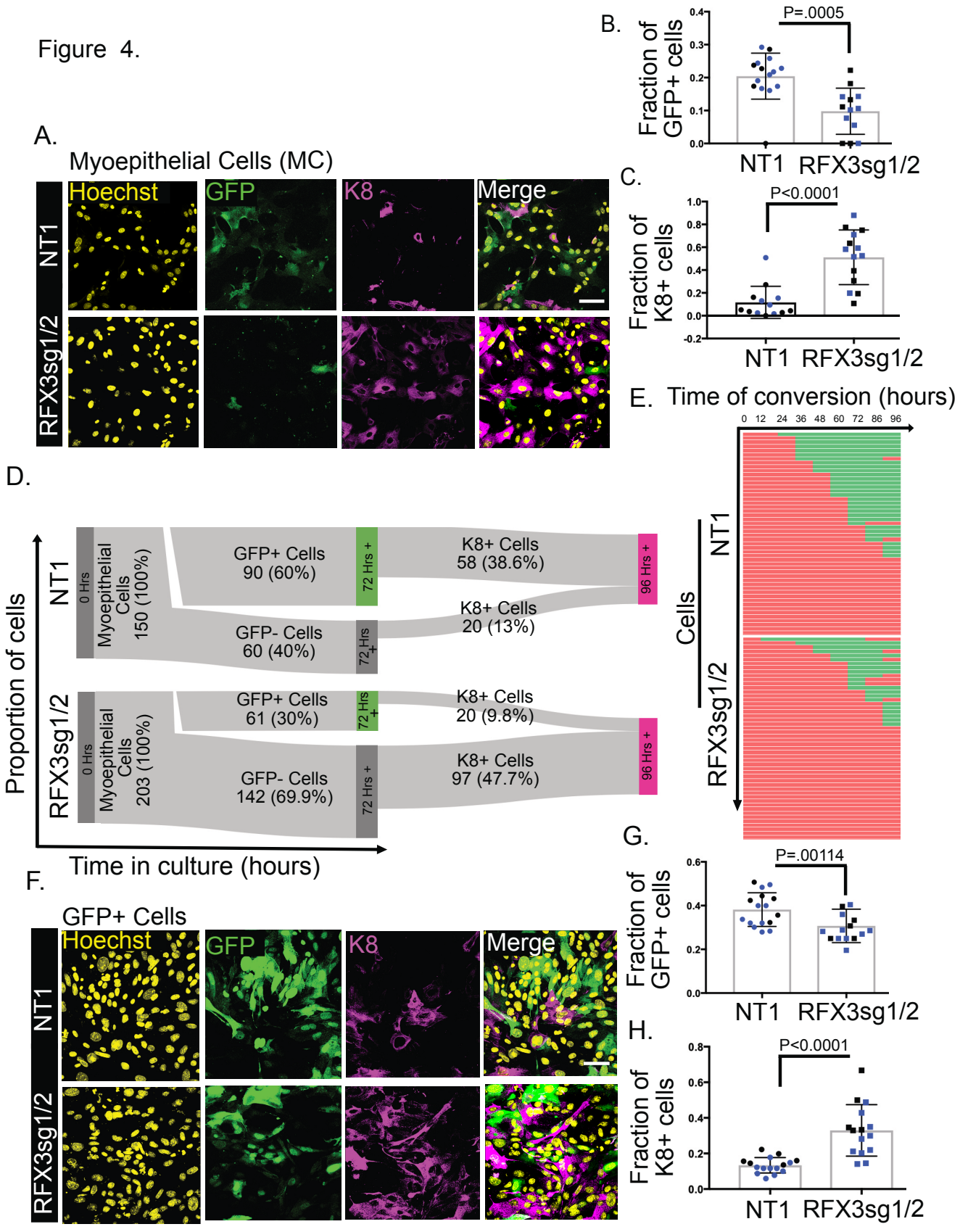
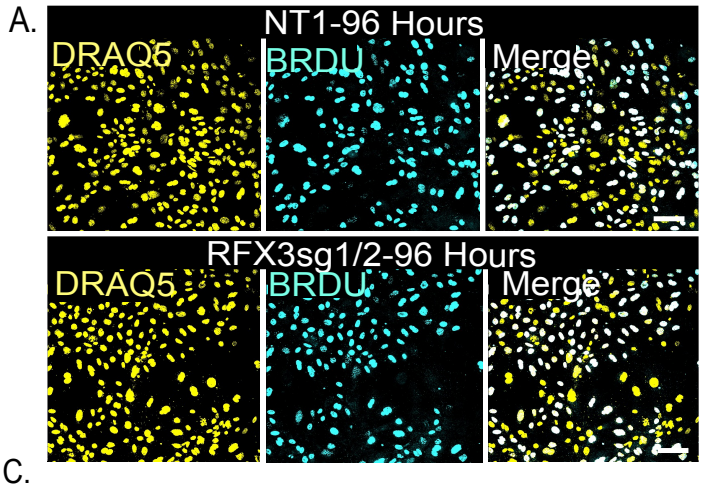
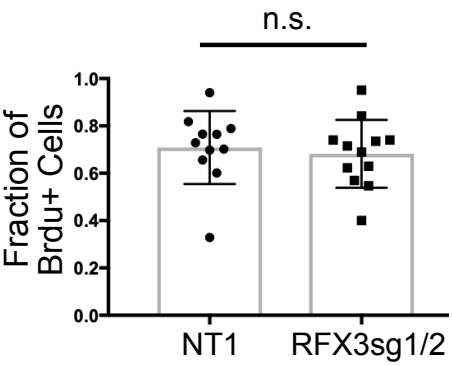

A.
Myoepithelial Cells (MC)
Hoechst
GFP
K8
Merge
NT1
RFX3sg1/2
B.
P=.0005
Fraction of GFP+ cells
NT1
RFX3sg1/2
C.
P<0.0001
Fraction of K8+ cells
NT1
RFX3sg1/2
D.
Proportion of cells
NT1
0 Hrs
Myoepithelial Cells 150 (100%)
GFP+ Cells 90 (60%)
GFP- Cells 60 (40%)
72 Hrs +
K8+ Cells 58 (38.6%)
K8+ Cells 20 (13%)
96 Hrs +
RFX3sg1/2
Myoepithelial Cells 203 (100%)
GFP+ Cells 61 (30%)
GFP- Cells 142 (69.9%)
K8+ Cells 20 (9.8%)
K8+ Cells 97 (47.7%)
Time in culture (hours)
E. Time of conversion (hours)
0 12 24 36 48 60 72 86 96
Cells
NT1
RFX3sg1/2
F.
GFP+ Cells
G.
P=.00114
H.
P<0.0001

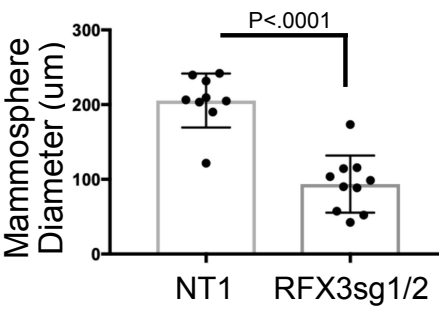
Figure 4. Loss of RFX3 promotes luminal transdifferentiation and reduces GFP+ cap cell abundance.

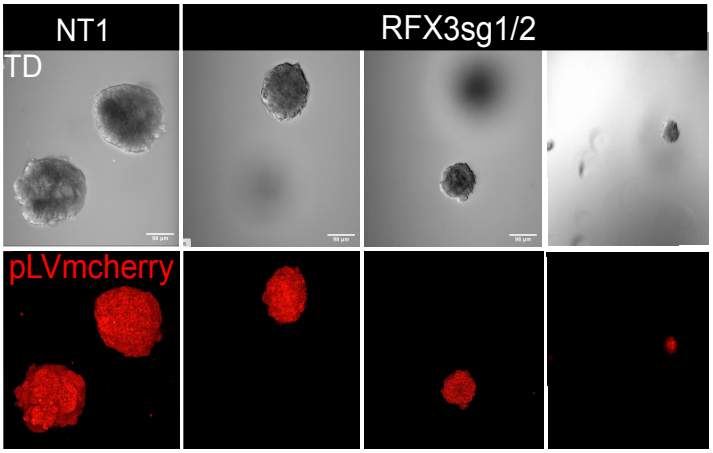
A. Representative images of RFX3 knockout (RFX3sg1/2) versus a negative control (non-targeting gRNA) in myoepithelial cells cultured in vitro for 96 hrs. Scale bar = 100μm
B. Quantification showing a decrease in the number of myoepithelial cells that become GFP+ (NT1: n= 14, 2 experiments, RFX3sg1/2: n=13, 2 experiments) when RFX3 is lost. P-value = .0005, statistics calculated using mixed model ANOVA. Error bars = Mean +/- SD.
C. Quantification graph showing an increase in the fraction of luminal cells (K8+) (NT1:n= 13, 2 experiments, RFX3sg1/2:n=14, 2 experiments) that arise. Individual data points show values for technical replicates; different colors identify biological replicates. Statistics calculated using mixed model ANOVA. Error bars = Mean +/- SD.
D. Sankey Plots of NT1 and RFX3 knockout (RFX3sg1/2) transdifferentiation behavior in vitro after 96 hrs. Loss of RFX3 results in a majority of myoepithelial cells converting towards a luminal cell fate, bypassing a GFP+ phase.
E. Kymograph showing times of conversion of myoepithelial cells into GFP+ cells and when/if GFP expression turned off. n= 50 cells
F. Representative images of RFX3 knockout (RFX3sg1/2) or the non-targeting control in FACS-purified GFP+ cap cells cultured in vitro for 96 hrs. Scale bar = 100μm
G. Quantification showing a decrease in the number of myoepithelial cells that become GFP+ (NT1: n= 16, 2 experiments, RFX3sg1/2: n=14, 2 experiments) when RFX3 is lost. Individual data points show values for technical replicates; different colors identify biological replicates. Statistics calculated using mixed model ANOVA. Error bars = Mean +/- SD.
H. Quantification graph showing an increase in the number of luminal cells (NT1: n= 16, 2 experiments, RFX3sg1/2: n=15, 2 experiments) that arise. P-value < 0.0001, statistics calculated using mixed model ANOVA. Error bars = Mean +/- SD.

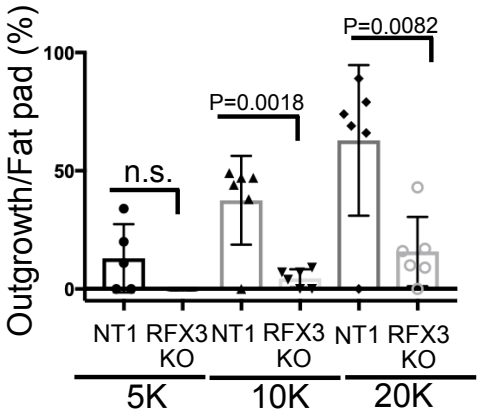
Figure 5.

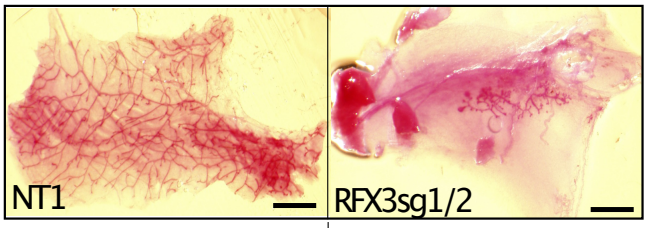












F.

Cell #	NT1	RFX3sg1/2
20K 10K 5K		
Stem Cell Frequency (95%CI)	**1/7381** (1/14017-1/3887)	**1/41704** (1/108614-1/16013)
	LR P-value = 0.00144	

H.

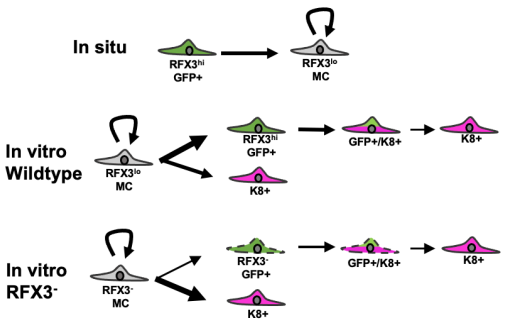
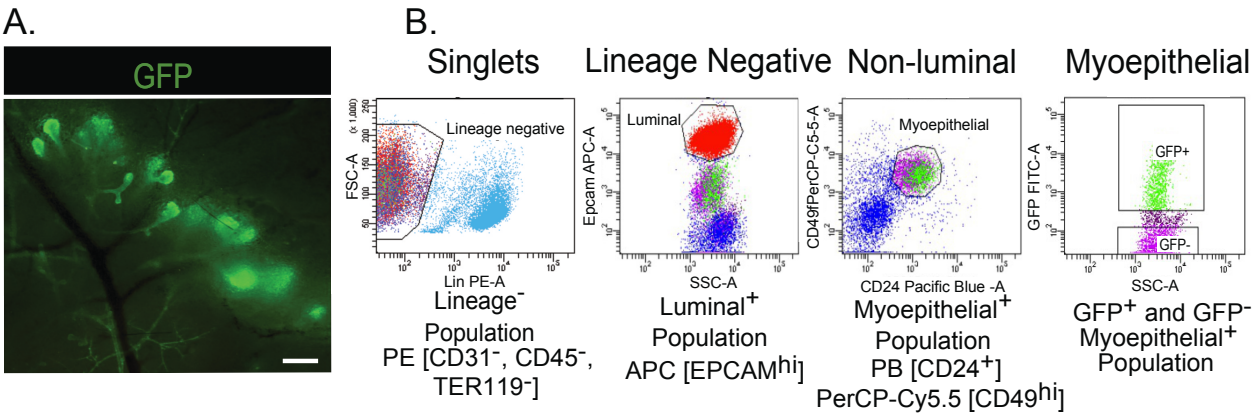


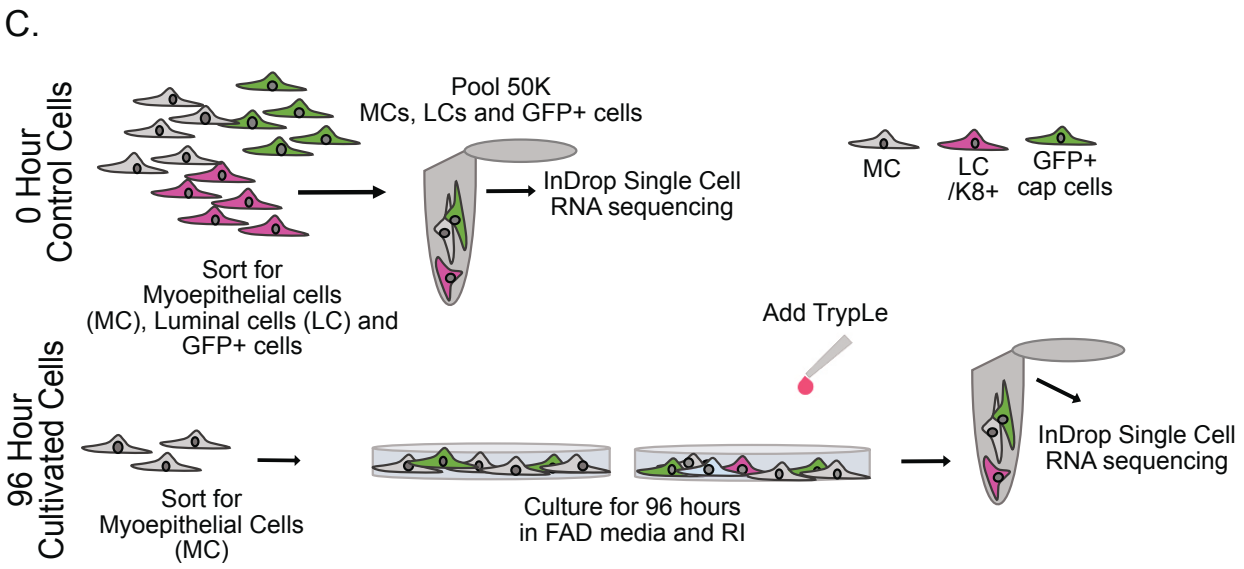
Figure 5. RFX3 stabilizes basal cell identities

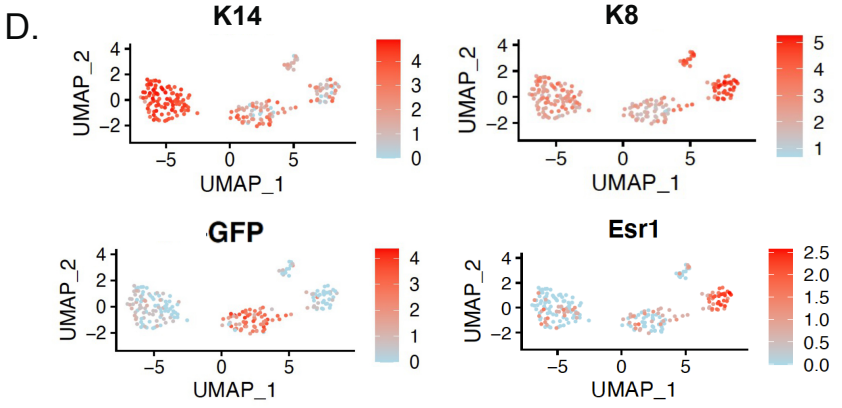
A. Representative images of myoepithelial cells transduced with either NT1 or RFX3 knockout lentivirus (RFX3sg1/2), pulse-labeled for 2 hrs with BrdU and stained for DNA (DRAQ5) and BrdU incorporation. Scale bar = 100μm.
B. Quantification of fraction of BrdU+ myoepithelial cells. P-value= n.s., calculated using unpaired t-test. Error bars = Mean +/- SD.
C. Representative images of myoepithelial cells transduced with either NT1 or RFX3 knockout lentivirus (RFX3sg1/2) and grown as mammospheres. Cherry+ mammospheres confirm transduction. Scale bar = 98μm
D. Quantification of mammosphere size. P-value = < 0.0001, calculated using unpaired t-test. Error bars = Mean +/- SD.
E. Representative images of outgrowths regenerated from transplantation of 20K NT1- or 20K RFX3sg1/2–transduced myoepithelial cells. Scale bar = 3 mm.
F. Circle graphs displaying the percentage of outgrowth that resulted (outgrowth/fat pad area) from the number of cells transplanted. Stem cell frequency was calculated and generated a likelihood ratio (LR) of P-value = 0.00144.
G. Quantification of transplantation assay outgrowth. P-values were calculated using unpaired t-test. P-value = n.s. (5K cells) 0.0018 (10K cells), 0.0082 (20K cells). Error bars = Mean +/- SD.
H. Model for the role of RFX3 in stabilizing mammary basal cell identities.
In situ, cap cells, the myoepithelial progenitors (GFP+), give rise only to mature myoepithelial cells (MC) which maintain their lineage through self-renewal. In vitro, wildtype myoepithelial cells de-differentiate predominantly towards a progenitor state (GFP+) before changing lineage into K8+ luminal cells (K8). A subset of myoepithelial cells can apparently bypass the GFP+ state and transdifferentiate directly into K8+ cells. RFX3 KO myoepithelial cells have a reduced frequency of conversion into a GFP+ state (thin arrow) and increased transdifferentiation directly into a luminal cell lineage (thick arrow).

SUPPLEMENTARY FIGURES AND FIGURE LEGENDS

Supplementary Figure 1.



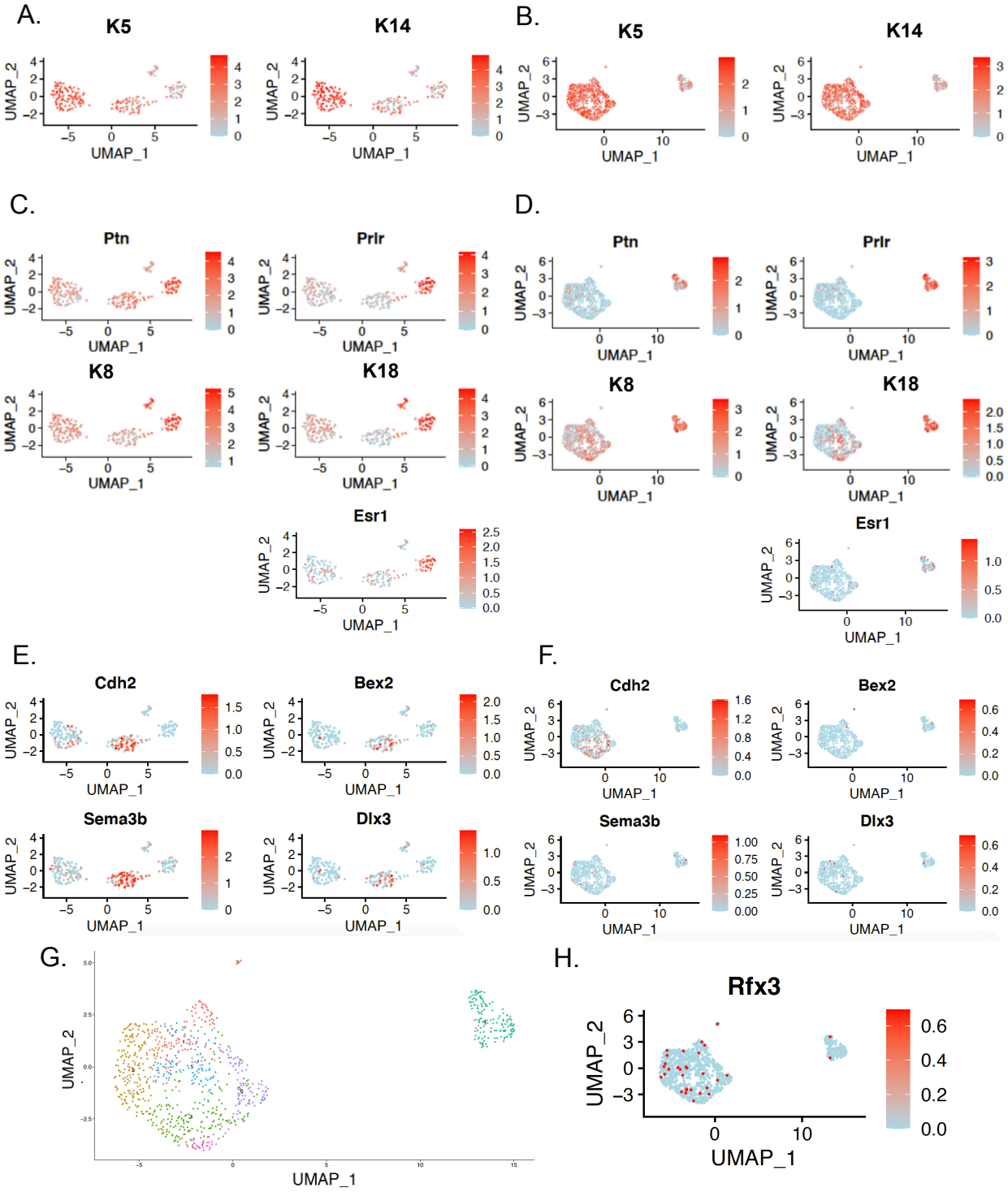




Supplemental Figure 1. Identification, isolation and single-cell RNA sequencing of GFP+ cap cells and myoepithelial cells

A. Whole mount of mammary glands from transgenic mice (TG11.5kb-GFP) showing GFP+ terminal end buds. Scale bar = 1 mm
B. Gating strategy for isolation of myoepithelial cells (GFP-) from TG11.5kb-GFP transgenic mice.
C. Diagram of experimental approach to prepare cells for single-cell RNA sequencing.
D. UMAP projections of control sample displaying the 3 sorted and distinctly clustered cell populations with K14 (myoepithelial cells), K8 (luminal cells), Esr1 (Esr1+ luminal cells) and GFP (cap cell) gene expression levels.

Supplementary Figure 2.

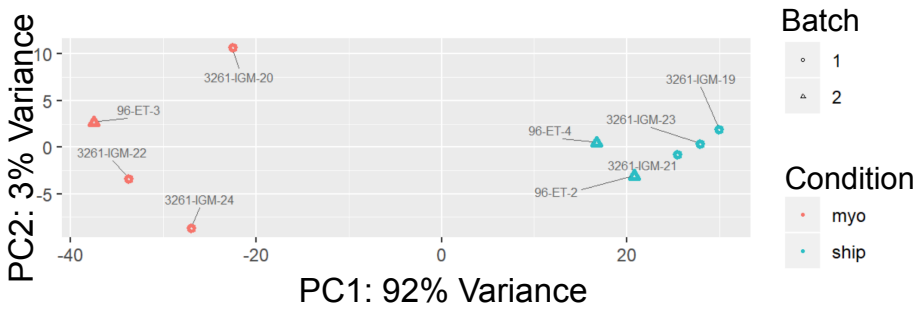


Supplemental Figure 2. Single-cell RNA sequencing differential gene expression analysis

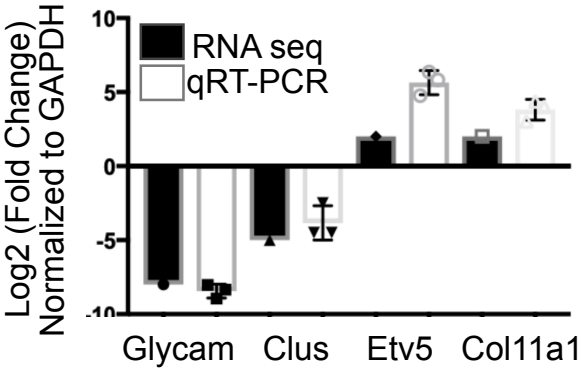
A. UMAPs of myoepithelial cell specific markers, K5 and K14 in fresh control sample
B. UMAPs of myoepithelial cell specific markers, K5 and K14 in 96 hr cultivated sample
C. UMAPs of luminal cell specific markers, Ptn, Prlr, K8, K18 and Esr1 in fresh control sample
D. UMAPs of luminal cell specific markers, Ptn, Prlr, K8, K18 and Esr1 in 96 hr cultivated sample
E. UMAPs of 4 genes highly expressed in cap cells (based on bulk RNA seq differential gene expression data) in fresh control sample
F. UMAPs of genes highly expressed in cap cells (based on bulk RNA seq differential gene expression data) in 96 hr cultivated sample
G. UMAP of 96 hr cultivated myoepithelial cells showing the different sub clusters within the myoepithelial cluster.
H. UMAP of RFX3 expression within the 96 hr cultivated myoepithelial cell sample

Supplementary Figure 3.

A.



B.



Supplemental Figure 3. Bulk RNA sequencing validation

A. Principle Component Analysis of GFP+ cap (ship) and myoepithelial (myo) cells show distinct clustering (data has been batch corrected). Shapes indicate different experiments. ship, n=5, myo, n=4.
B. Validation of RNA seq differential gene expression by qRTPCR. Comparison of the Log2 (Fold Change) (mean) of RNA Seq and qRT-PCR data expression of myoepithelial cells vs. cap cells, which are consistent with one another, n=3

2.5 Materials and Methods

Table 1. Key Resources Table

Key Resources Table				
Reagent type (species) or resource	**Designation**	**Reagent type (species) or resource**	**Identifiers**	**Additional information**
Strain	C57BL/6J	Strain	Cat#000664B6, B6J, B6/J	
Strain	s-SHIP-GFP transgenic (Tg11.5kb-GFP)	Strain		Fred Hutchinson Cancer Research Center, Seattle, Washington
Antibody	Chicken polyclonal anti-Cytokeratin 14 (1:500)	Antibody	Cat#906001; RRID: AB_2565055	
Antibody	Rat monoclonal anti-Cytokeratin 8 (1:500)	Antibody	Cat#TROMA-I; RRID: AB_531826	
Antibody	Goat polyclonal anti-Chicken Alexa Fluor 488 (1:1000)	Antibody	Cat#A-11039; RRID: AB_2534096	
Antibody	Goat polyclonal anti-Chicken Alexa Fluor 647 (1:1000)	Antibody	Cat#A-21449; RRID: AB_2535866	
Antibody	Goat polyclonal anti-Rabbit Alexa Fluor 488 (1:1000)	Antibody	Cat#A-11034; RRID: AB_2576217	
Antibody	Goat polyclonal anti-Rabbit Alexa Fluor 594 (1:1000)	Antibody	Cat#A-11037; RRID: AB_2534095	
Antibody	Donkey polyclonal anti-Goat Alexa Fluor 594 (1:1000)	Antibody	Cat#A32758; RRID: AB_2762828	
Antibody	Donkey polyclonal anti-Goat Alexa Fluor 647 (1:1000)	Antibody	Cat#A-21447; RRID: AB_2535864	
Antibody	Donkey polyclonal anti-Mouse Alexa Fluor 594 (1:1000	Antibody	Cat#A-21203; RRID: AB_141633	
Antibody	Donkey polyclonal anti-Mouse Alexa Fluor 647 (1:1000	Antibody	Cat#A-31571; RRID: AB_162542	
Antibody	Donkey polyclonal anti-Rat Alexa Fluor 594 (1:1000)	Antibody	Cat#A-21209; RRID: AB_2535795	
Antibody	Goat polyclonal anti-Rat Alexa Fluor 647 (1:1000)	Antibody	Cat#A-21247; RRID: AB_141778	
Antibody	PE rat anti-mouse CD31 (1:200)	Antibody	Cat#554656AB;RRID:AB_394819	
Antibody	PE rat anti-mouse TER-119 (1:200)	Antibody	Cat#12-5921-82; RRID: AB_466042	
Antibody	PE rat anti-mouse CD45 (1:200)	Antibody	Cat#103105; RRD: AB_312970	
Antibody	PB rat anti-mouse CD24 (1:400)	Antibody	Cat#101819; RRID: AB_572010	
Antibody	APC rat anti-mouse CD326 (1:600)	Antibody	Cat#50-152-15; RRID: 175791	
Antibody	PerCP-Cy5 rat anti-human/mouse CD49f (1:200)	Antibody	Cat#50-166-772; RRID: AB_313618	
Antibody	rabbit anti-RFX3 (1:500)	Antibody	Cat#NBP1-86301	
Antibody	rabbit anti-Ran13 (1:1000)	Antibody		
Other	Hoechst 33342 (1:1000)	Other	Cat#H3570	
Other	anti-BrdU (1:1000)	Abcam	Cat#Ab6326	
Recombinant DNA reagent (*Mus musculus*)	Rfx3 KO 1	This paper		ATCATGCAGACTTCAGAGA
Recombinant DNA reagent (*Mus musculus*)	Rfx3 KO 2	This paper		GCAAGTGCCAGTGCAGCAGC
Recombinant DNA reagent (*Mus musculus*)	plentiCRISPRv2 puro	Addgene	Cat#98290;RRID:Addgene_98290	
Recombinant DNA reagent (Mus musculus)	pLentiCRISPR-mCherry	Addgene	Cat#75161;RRID:Addgene_75161	
Recombinant DNA reagent (Mus musculus)	control nontargeting (NT)sgRNA	This paper		GCGAGGTATTCGGCTCCGCG
Software	GraphPad Prism	GraphPad Software	RRID: SCR_002798	https://graphpad-prism.software.informer.com/5.0/;
Software	Fiji	Fiji	RRID: SCR_002285	
Software	NIS Elements	Nikon	RRID: SCR_014329	https://www.nikoninstruments.com/Products/Software;
Software	Ingenuity Pathway Analysis	Qiagen	RRID: SCR_002798	http://www.ingenuity.com/products/pathways_analysis.html
Software	CLC Workbench	Qiagen	RRID:SCR_011853	https://digitalinsights.qiagen.com/
Software	SankeyMATIC	GitHub	https://sankeymatic.com	
Software	R	The R Project	https://www.r-project.org	
Sequence-based reagent	GAPDHF	This paper		5'-TCTCCACACCTATGGTGCAA-3'
Sequence-based reagent	GAPDHR	This paper		5'-TGCCGTGAGTGGAGTCATA-3'
Sequence-based reagent	GLYCAMF	This paper		5'-GTGCCACCATGAAATTCTTC-3'
Sequence-based reagent	GLYCAMR	This paper		5'- TCTTCATGACTTCGTGATAC-3'
Sequence-based reagent	CLUSTERINF	This paper		5'- GCATACCTGCATGAAGTTCTAT-3'
Sequence-based reagent	CLUSTERINR	This paper		5'- GTAGAAGGGTGAGCTCTGGTTT-3
Sequence-based reagent	COLLAGEN11A1F	This paper		5'- GACTACTCAGATGGCATG-3'
Sequence-based reagent	COLLAGEN11A1R	This paper		5'- ACTTCCTGGTTTCTCCTT-3'
Sequence-based reagent	ETSVARIANTFACTOR5F	This paper		5'- GCAGTTTGTCCCAGATTTTCA-3
Sequence-based reagent	ETS VARIANTFACTOR5R	This paper		5' GCAGCTC CCGTTTGATCTT-3'
Sequence-based reagent	RFX3F	This paper		5'-GGACAGCCGCTTTCAGAGAA-3'
Sequence-based reagent	RFX3R	This paper		5'-TCTCTACAGCCCAGCAAGGA-3'
Sequence-based reagent	IFT88F	This paper		5'-TGGCCAACGACCTGGAGATTAACA-3
Sequence-based reagent	IFT88R	This paper		5'-ATAGCTGCTGGCTTGGGCAAATTC-3',
Sequence-based reagent	DYNC2LI1F	This paper		5'-GGTGAGCCGGAATACAGAGAA-3'
Sequence-based reagent	DYNC2LI1R	This paper		5'-TGTTTGGTAGGATCTGGGACA-3'.
Chemical compound, drug	Laminin	Milipore Sigma	Cat#L2020; CAS: 114956-81-9	
Chemical compound, drug	BrdU	Millipore Sigma	Cat#B5002; Beilstein Registry#: 30395	
Chemical compound, drug	ROCK Inhibitor (Y-27632)	Sigma-Aldrich	Cat#Y0503-1MG	
Deposited Data	Raw and analyzed RNA sequencing data	This paper	GEO: *pending*	
Cell Lines	Eph4, mouse	Dr. Jürgen Knoblich	N/A	Institute of Molecular Biotechnology, Vienna, Austria
Cell Lines	HEK293T, human	ATTC	RRID: CVCL_0063	
Cell Lines	CommaD, mouse	Dr. Javier F. Medina	N/A	Baylor college of Medicine, Houston, TX

2.5.1 Mice

C57BL/6J (JAX stock # 000664) female mice were purchased from Jackson Laboratory (Bar Harbor, ME). The s-SHIP–GFP transgenic (Tg11.5kb–GFP) mice were provided by Dr. Larry Rohrschneider from Fred Hutchinson Cancer Research Center, Seattle, Washington. Mice were housed in the Vanderbilt mouse facility with a standard 12hrs light/12hrs dark cycle and provided with normal laboratory chow and water. The mice were monitored daily by the Vanderbilt Division of Animal Care (DAC). s-SHIP-GFP primer sequences used for genotyping were as described previously (Bai and Rohrschneider, 2010). All mouse experiments were performed with approval from the Vanderbilt Institutional Animal Care and Use Committee (IACUC).

2.5.2 Cell Lines and Cell Culture

Mouse mammary EpH4 cells were provided by Dr. Juergen Knoblich (Institute of Molecular Biotechnology, Vienna, Austria). HEK293T (ATCC CRL-3216) cells were obtained from ATCC. CommaDβ cells were provided by Dr. Medina (Baylor college of Medicine, Houston, TX). Cell lines were cultured in Dulbecco's modified Eagle medium (Life Technologies), supplemented with 10% Fetal Bovine Serum (R&D Systems, Minneapolis, MN) and 1X penicillin/streptomycin (Life Technologies) and maintained in culture at 37°C with 5% CO_2. WPI, psPAX2 and pMD2.G were provided by Didier Tron. LentiCRISPRv2 puro was a gift from Brett Stringer (Addgene plasmid # 98290; http://n2t.net/addgene:98290;RRID:Addgene_98290). pLentiCRISPR-mCherry was a gift from Beat Bornhauser. (Addgene plasmid #75161;http://n2t.net/addgene:75161;RRID:Addgene_75161). Guide RNAs for RFX3 DNA binding-domain knock out (Key Resources Table) were cloned into plentiCRISPRv2 puro or pLentiCRISPR-mCherry. sgRNAs were ligated into lentivectors at the *BsmBI* restriction site as described in the Zhang lab protocol [118, 119].

2.5.3 Lentiviral Production

Lentivirus was produced by transfecting 80% confluent 15 cm dishes of HEK293T cells. Lentiviral plasmid, protein plasmid (pMD2.G) and packaging plasmid (psPAX2) were transfected into HEK293T cells at a 1.5:1:1 ratio using calcium phosphate precipitation. Virus-conditioned medium was collected from cells 48 hrs post transfection, stored at −80 °C, and titered using HEK293Tcells.

2.5.4 Primary Cell Isolation

To isolate mammary epithelial cells the third and fourth pairs of mammary glands were removed from 4 to 6-week-old s-SHIP–GFP transgenic (Tg11.5kb–GFP) mice, minced with scissors and digested in digestion medium (DMEM/F12, 2 mg ml^{-1}collagenase I (Roche), 5 mg ml^{-1} insulin (Sigma), 100 U ml^{-1} penicillin/streptomycin) for 1 hr at 37 °C. Epithelial organoids were collected by centrifugation at 1500 rpm for 5 min. The cell pellet was washed for 1 min in 5 ml of DMEM/F12 containing DNase I and centrifuged at 1500 rpm for 5 min. The pellet was resuspended in 5 ml DMEM/F12 10% fetal bovine serum followed by 15 sec of centrifugation at 1500 rpm five times. Cells were resuspended in 1 ml of fresh trypsin/EDTA (Invitrogen) and incubated at 37 °C for 15 min, dissociated into a single-cell suspension and passed through a cell strainer (BD) to obtain a single-cell suspension of mammary gland cells. Primary cells were transduced with lentivector (RFX3 KO lentivectors or NT1) by incubating for 1-4 hours at 37 °C.

2.5.5 BrdU Incorporation Assay

72 hours post-transduction with lentivector (RFX3 KO or NT1), myoepithelial cells were pulse-labeled for 2 hrs with BrdU (Millipore Sigma, Cat # L2020). Cells were then stained with anti-BrdU (Abcam, Cat# Ab6326) according to the antibody protocol.

2.5.6 Flow cytometry, antibodies, and cell sorting

Single mammary epithelial cells were blocked in 5% goat serum for 5 min on ice, stained with antibodies for 30 min on ice, washed, and resuspended in 1X PBS, 2mM EDTA, 2%FBS. Antibodies used were PE rat anti-mouse CD31 (1:200; BD Biosciences), PE rat anti-mouse TER-119 (1:200; Invitrogen), PE rat anti-mouse CD45

(1:200; BioLegend), PB rat anti-mouse CD24 (1:400; BioLegend), APC rat anti-mouse CD326 (Epcam) (1:600; Ebioscience), and PerCP-Cy5 rat anti-human/mouse CD49f (1:200; BioLegend). The single live cells were gated and sorted on 5-laser FACS Arialll flow cytometers.

2.5.7 Myoepithelial Cell Conversion Assays

Freshly sorted mature myoepithelial cells were plated on 16-well chambered coverglasses (Grace Biol-Labs Cat. # 112358) in FAD media supplemented with 10 μm ROCK Inhibitor (Y-27632) (Sigma-Aldrich Cat.# Y0503-1MG) and cultivated for 96 hrs. Media was changed every day and cells were fixed, stained and analyzed at 96 hrs.

2.5.8 Cell Fate Mapping

Myoepithelial cells were sparsely plated onto gridded coverslips (ibidi catalogue #80826-G500) in FAD media and live imaged on a Nikon A1R inverted confocal microscope (Nikon Instruments Inc.) every 12 hrs for 96 hrs using 20X/0.75 NA. At 96 hrs cells were fixed and stained for K8 and nuclei. Cells were annotated at 24 hrs and tracked up to hr 96. Cells that were tracked over the 96 hrs were scored for becoming GFP+ and whether the cells were expressing luminal marker (K8) at hr 96. Sankey plots were generated using SankeyMATIC by Steve Bogart (https://github.com/nowthis/sankeymatic/blob/main/README.md). Kymographs were generated by analyzing images from 12 hrs to 96 hrs and marking the timepoint at which GFP expression is visible and the timepoint when GFP expression turns off, if within the 96 hrs.

2.5.9 Mammosphere Assays

Myoepithelial cells were embedded in growth-factor-reduced Matrigel (BD) on16-well chambered coverglasses (Grace Biol-Labs Cat. # 112358) and cultured in mammosphere medium supplemented with 10 μm Y-27632 and cultivated for 10 days. Medium was changed every day and cells were fixed, stained and analyzed on day 10.

2.5.10 Tissue Processing, Staining and Analysis

For immunocytochemistry, cells were fixed in 4% paraformaldehyde at room temperature for 15 min. Cells were permeabilized with 0.2% Triton X-100. For immunohistochemical analyses mammary tissue samples were cryo-embedded in O.C.T. (Fisher Scientific, Hampton, NH) and cryo-sectioned on a Leica C1950 cryostat generating 50 µm sections of mammary tissue. Tissue sections were fixed for 15 min in 4% paraformaldehyde or for 5 min in -20°C Acetone. Both cells and tissues were blocked in 1x Western Blocking Reagent (Roche). The primary antibodies used in this study included: chicken α-cytokeratin 14, rabbit α-S100A4 (BioLegend, San Diego, CA), rat α-cytokeratin 8 (Developmental Studies Hybridoma Bank, Iowa City, Iowa), rabbit α-rfx3 (Novus Biologicals, Centennial, CO). Secondary antibodies used in this study included: Alexa Fluor 405, 488, 594 and 647 conjugates of anti-chicken, -mouse, -rat, -rabbit and -goat (ThermoFisher Scientific, Waltham, MA). Hoechst 33342 (ThermoFisher Scientific, Waltham, MA) was used to stain DNA. Slides were mounted in Fluoromount G and sealed with nail varnish. Laser scanning confocal Images were acquired on a Nikon A1R inverted confocal microscope (Nikon Instruments Inc.). 20X/0.75 numerical aperture (NA) and 40X/1.30 NA Plan Apochromat objectives were use. Type B Immersion oil (Cargille Laboratories, Cedar Grove, NJ) was used. Image post-acquisition processing was done using Nikon NIS-Elements imaging software and Fiji (ImageJ) software. The images in the manuscript are maximum intensity projections of z-stacks and merged images are composites of individual color channels.

2.5.11 RNA Sequencing

Total RNA (50-200ng) isolation was performed using either TRIzol (Life Technologies) or the RNeasy Mini Kit (Qiagen, Hilden, Germany). 5 samples of myoepithelial cells and 5 samples of s-SHIP GFP+ cap cells were submitted and subjected to quality analysis using a 2100 Bioanalyzer (Agilent Technologies, Santa Clara, CA). All Samples had an average RNA integrity number (RIN) value ~8. Libraries for whole transcriptome analysis were generated following Illumina's TruSeq RNA v2 sample preparation protocol. Libraries were sequenced on an Illumina HiSeq 3000 at the Vanderbilt Technologies for Advanced Genomics (VANTAGE). 150 million reads at 75 basepairs

were obtained for each of the samples. Data processing, analysis and plotting were performed using R software, CLC Genomics Workbench and Ingenuity Pathway Analysis (Qiagen, Hilden, Germany). Heatmaps were generated using R graphics on RNAseq gene lists filtered with a p-value < .05 and fold change >2. Volcano plots were generated using CLC workbench with genes filtered with a p-value of <0.005. The GO ontology graph was generated using Ingenuity Pathway Analysis and genes were filtered with a p-value of <.05 and fold change >2. Reads for s-Ship were quantified manually by analyzing the read alignment data to the 42 nucleotide unique sequence present in this variant.

2.5.12 Single Cell RNA Sequencing

After 96 hrs cultivation, converted myoepitheial cells were dissociated using TrypLE and cell viability was determined using Trypan Blue. The control sample (fresh s-SHIP GFP+ cap cells, myoepithelial cells and luminal cells) and cultivated cells were encapsulated and barcoded using the inDrop platform (1CellBio) with an in vitro transcription library preparation protocol (Klein *et al*., 2015). As per Klein *et al*., the CEL-Seq workflow is summarized: 1) RT, 2) ExoI, 3) SPRI purification (SPRIP), 4) SSS, 5) SPRIP, 6) T7in vitro transcription linear Amplification, 7) SPRIP, 8) RNA Fragmentation, 9) SPRIP, 10) primer ligation, 11) RT, 12) library enrichment PCR. The number of cells encapsulated was calculated by the density of cells arriving at the device multiplied by the duration of encapsulation. After library preparation, the samples were sequenced using Nextseq 500 (Illumina) using a 150bp paired-end sequencing kit. After sequencing, reads were filtered, sorted by their barcode of origin and aligned to the reference transcriptome using the inDrops pipeline (https://github.com/indrops/indrops). Mapped reads were quantified into UMI-filtered counts per gene, and barcodes that correspond to cells were retrieved based on previously established methods (Klein *et al*., 2015).

2.5.13 Real-time qPCR

cDNA was reverse transcribed using the SuperScript III First-Strand Synthesis System (Invitrogen). qPCR was performed with triplicate replicates on a BioRad CFX96 Thermocycler and analyzed using the ΔΔCt method. Expression levels were calculated

relative to GAPDH. Primer sequences used were: GAPDHF 5'-TCTCCACACCTATGGTGCAA-3', GAPDHR 5'-TGCCGTGAGTGGAGTCATA-3', GLYCAMF 5'-GTGCCACCATGAAATTCTTC-3', GLYCAMR 5'-TCTTCATGACTTCGTGATAC-3', CLUSTERINF 5'- GCATACCTGCATGAAGTTCTAT-3', CLUSTERINR 5'- GTAGAAGGGTGAGCTCTGGTTT-3', COLLAGEN11A1F 5'-GACTACTCAGATGGCATG-3', COLLAGEN11A1R 5'- ACTTCCTGGTTTCTCCTT-3', ETSVARIANTFACTOR5F 5'- GCAGTTTGTCCCAGATTTTCA-3', ETS VARIANTFACTOR5R 5' GCAGCTC CCGTTTGATCTT-3',RFX3F 5'-GGACAGCCGCTTTCAGAGAA-3', RFX3R 5'-TCTCTACAGCCCAGCAAGGA-3',IFT88F 5'-TGGCCAACGACCTGGAGATTAACA-3',IFT88R 5'-ATAGCTGCTGGCTTGGGCAAATTC-3', DYNC2LI1F 5'-GGTGAGCCGGAATACAGAGAA-3', DYNC2LI1R 5'-TGTTTGGTAGGATCTGGGACA-3'.

2.5.14 Gene Regulatory Network Analysis

iRegulon (Janky et.al., 2014) was used to predict transcription factor regulation. RNA seq data was filtered for genes highly expressed in GFP+ cap cells with a p-value < .05. Prediction data was generated by surveying transcription factor binding motifs present 20 kb around transcription start site (TSS) [TSS−10 kb,TSS+10 kb]. Gene regulatory networks were generated within Cytoscape.

2.5.15 Transplantation/ Limited Dilution Assays

Myoepithelial cells were FACS sorted and transduced with either non-targeting or RFX3-targeting lentivirus. 24 hrs later cells were dissociated in trypLE Select (Gibco) for 8 minutes. Dilutions of 5, 000, 10,000, or 20,000 cells were resuspended in 10 ul of injection medium (10 μm Y-27632 (Sigma-Aldrich Cat.# Y0503-1MG), 20% Matrigel (Corning), DMEM/F12, 40 ng/mL EGF, 20 ng/mL (R&D Systems), FGF2 (R&D Systems) and injected into the cleared fat pads of 3-week-old female C57Bl6 mice (Jackson Laboratories) using a 26 gauge needle and Hamilton syringe. Mice were sacrificed 8 weeks post transplantation and whole mounts analyzed. Outgrowths were detected by carmine alum whole-mount. Pictures were acquired with an Olympus

SZX16. A positive outgrowth was described as 10% or greater percentage of outgrowth/fatpad.

2.5.16 Western Blots

Cells were washed with 1x PBS and lysed in 1 ml of lysis buffer containing 20 mM HEPES; pH 7.4, 50 mM NaCl, 2 mM EDTA, and 0.1% Triton X-**100**, supplemented with cOmplete^{TM}mini EDTA-free protease inhibitor cocktails (Roche) and PhosStop (Roche). Cell lysates were briefly centrifuged at 16,000 x g and the soluble fraction was boiled with SDS sample buffer for 5 min. Antibodies used for Western blotting include: rabbit anti-RFX3 (Novus Biologicals, Centennial, CO) and rabbit anti-Ran13 (Macara Lab; [120].

2.5.17 Statistics

All cell counts were performed using Hoechst staining and a fluorescent antigen marker in NIS-Elements imaging software or Fiji (ImageJ) software. All measurements for mammosphere diameter, transplant outgrowth area and western blot analysis were done using ImageJ. All statistical analyses were performed using unpaired Student's t-test or two-way ANOVA test. Stem cell frequencies were calculated using Extreme Limiting Dilution Analysis (ELDA) [121].

2.6 Data and Code Availability

This study generated Rfx3 sgRNAs (KO 1: ATCATGCAGACTTCAGAGA, KO 2: GCAAGTGCCAGTGCAGCAGC) within the pLentiCRISPR-mCherry (Addgene plasmid #75161;http://n2t.net/addgene:75161;RRID:Addgene_75161) backbone that target the Mus musculus Rfx3 (Gene ID: 19726) DNA-binding domain. The RNA sequencing datasets generated in this study will be deposited to the GEO repository on the NCBI website.

2.7 Acknowledgements

This work was supported by grant R35 CA197571 to IGM, and T32 HD07502 to EMT. We thank members of the Macara laboratory for advice and support.

III. CONCLUSIONS

We studied the transdifferentiation of myoepithelial cells to luminal cells to understand mammary epithelial cell plasticity and the factors that regulate lineage conversion. We found that during conversion the principal pathway for transdifferentiation involves an initial, partial dedifferentiation towards a progenitor, cap cell state prior to converting to the luminal lineage. We also found that the transcription factor Regulatory Factor X 3 (RFX3) is consistently upregulated in cap cells and has a critical role in stabilizing the basal cell state. When RFX3 is knocked out of myoepithelial cells there is a reduced frequency of de-differentiation towards the cap cell lineage and/or increased transdifferentiation into K8+ luminal cells. There was a delay in the onset of GFP expression, and more cells reverted to a non-GFP+ state. We also observed a significant reduction in mammosphere growth of myoepithelial cells grown in 3D culture and substantially reduced mammary gland outgrowth for transplanted RFX3- cells compared to the NT1 control. We conclude that the RFX3 transcription factor functions to stabilize the identities of basal cell populations, which is necessary for efficient ductal regeneration.

The above findings support the idea that mature mammary epithelial cells and other cell types can exist in metastable states that allow them to revert to earlier cell states and change lineages. Microenvironment, stress and cell damage have been shown to induce cell state changes. This behavior has been observed in other animal tissues. Upon injury intestinal epithelial Dll1+ secretory progenitors can revert into functional stem cells. This is also observed in lung where differentiated airway epithelial cells can dedifferentiate into stem cells [114, 115].

In our study we identify a transcription factor, RFX3, that stabilizes basal cell behavior. Basal cell behavior includes the potential to revert to a less differentiated state in certain contexts. Several studies have shown that over-expression of certain transcription factors can force reversion in differentiated cells into stem cells. Most famously, Yamanaka showed that over-expression of Oct3/4, Sox2, Klf4 and c-Myc in human and mouse somatic cells induced pluripotency [117].

Differential signaling from micro-environments may regulate the transcription factors that determine more or less stringent cell state stabilization. Lineage tracing reveals that myoepithelial cells and the less differentiated cap cells never give rise to luminal cells *in vivo*, in the unperturbed tissue [75, 108]. This may be due to repressive signals from the luminal cell population *in situ*. Le Guelte's studies support this idea as she shows that when luminal and cap cells are co-cultured, cap cell multipotency is suppressed [107, 108]. Le Guelte also showed that TGFβ is important in cap cell behavior as it promotes transdifferentiation to the luminal lineage.

In my study of the mechanism behind mammary epithelial cell plasticity, I speculate that Sonic hedgehog signaling may play an important role in regulating transcription factors that mediate phenotypic stability. RFX3 is a master regulator of ciliogenesis and cilia are critical for Sonic hedgehog signaling [122]. A loss of RFX3 results in a decrease in ciliogenesis factors and cilia; this leads to perturbed Sonic hedgehog signaling and mammary basal cell instability [122].

This study leads to many questions about cell state stability in other contexts and tissues. Most instances of multipotency and cell plasticity are observed embryonically, under cell stress and in diseased states. Looking more closely at a cell's potential during normal development, could give further insight into the onset of diseased states. While adult stem cells do not exist in the mammary gland we have found that stem cell potential is maintained in post-embryonic development and can be unleashed in certain contexts like DNA damage [109].

IV. FUTURE DIRECTIONS AND CONSIDERATIONS

To further elucidate the role of Regulatory Factor X 3 (RFX3) in stabilizing basal cell identity there are additional studies and themes to consider concerning cell plasticity in the mammary gland and mammary epithelial cells.

Cell plasticity is defined as the ability of cells to change their phenotypes without genetic mutations, in response to environmental cues [123]. Cell plasticity is observed in many animal tissues including intestine and lung. Following injury, intestinal epithelial Dll1+ secretory progenitors can revert into functional stem cells. In the lung differentiated airway epithelial cells dedifferentiate into stem cells following damage or tissue loss [114, 115]. Mammary gland myoepithelial cells display plasticity in their ability to convert to luminal cells during cultivation.

In early development all cells of the early embryo have the same development potential, but as development progresses epigenetic modifications restrict the genomic landscape. As a future direction, It would be interesting to look at the epigenetic differences between freshly isolated myoepithelial cells, cultivated myoepithelial cells and RFX3- myoepithelial cells. I hypothesize that in freshly isolated myoepithelial cells, RFX3 target genes would be poised, with repressive modifications inhibiting gene expression, but during myoepithelial cell cultivation, there would be a switch from repressive to permissive modifications during dedifferentiation and then a re-accumulation of repressive modifications as cells differentiate into luminal cells.

Gene expression during physiological cell fate changes is controlled by spatiotemporal regulation of enhancers and target promoters [124]. Cis-regulatory elements within the promoter, whether proximal or distal, can be bound by a single – or several transcriptional proteins forming a complex [125]. Studies using DNA Fluorescence In Situ Hybridization (FISH) or Chromosome Capture (3C)-based approaches were first used to identify these complex interactions [126]. FISH and 3C have also been used to identify enhancer ‘hubs’ or ‘cliques’ which are areas enriched for interacting transcriptional machinery components and transcription factors. Hubs are very dynamic during development in cell

differentiation and dedifferentiation [127]. Though RFX3 is a single transcription factor predicted to regulate the cap cell gene signature and found to stabilize basal cell identity, most transcription factors do not independently regulate gene expression, and it is possible that RFX3 may be acting in a complex with other factors [128]. RFX3 could be acting with other members of the RFX transcription factor family, or other factors, even those whose expression RFX3 may control. To determine whether this is the case, I would isolate cap cells and myoepithelial cells and do a Co-Immunoprecipitation (Co-IP) and pull-down assay. Using RFX3 antibodies, this would reveal any additional proteins that form a complex with RFX3 and regulate cap cell gene expression and basal cell behavior. An additional control experiment that would strengthen my work would be to use an additional set of RFX3 KO single guide RNAs. My current guides target the n terminus and c terminus of the DNA-binding domain, but using a single guide or a set of guides targeting other exons and show a similar phenotype would strengthen my claims.

Communication between enhancers and promoters can happen over long distances, skipping several intervening genes [123]. With the knowledge that RFX3 loss results in a decrease in GFP+ cells and a default towards the luminal lineage, overexpression of RFX3 may result in stabilization of the basal cell identity, rescuing the phenotype. Overexpression studies come with some barriers; the expression levels necessary for rescue and whether expression would produce the same, opposite or different results would be difficult to determine. But to show the ability of RFX3 to rescue the lineage switch phenotype would further corroborate the importance of RFX3 in stabilizing the basal cell identity.

Ablating RFX3 from myoepithelial cells also results in a loss of stabilization of the basal cell identity which results in an otherwise multipotent cell defaulting into a differentiated luminal cell. It is as if during the conversion, because we have removed RFX3 the cell forgets the molecular mechanism that determines the cell fate decision, and the cell has an identity crisis. A term that has generated much controversy, ‘cell lineage infidelity’ is when a differentiated cell forgets or loses its

original identity and either reverts to a less differentiated multipotent cell state or transdifferentiates into a different cell type [129]. This is the phenotype that we observe in cultivated myoepithelial cells. We observe an increase in the proportion of luminal cells in culture in response to loss of RFX3. Is there also an increase in luminal cells in mammosphere and transplantation assays? A future experiment that may answer this question would include imaging mammospheres after fixing and staining and comparing the K8 fluorescent intensities between RFX3+ and RFX3- images. Fixing and cryosectioning tissue from transplantation assays would allow us to see if outgrowths have normal ductal architecture or if there is an abundance of luminal cells in the RFX3- tissue. In addition to determining the composition if RFX3- tissue, one of the observations made in transplantation assays was that after several weeks of growth RFX3+ outgrowths filled the fat pad, while RFX3$^-$ had retarded growth, but had active terminal end buds. If RFX3- outgrowths were left to grow for a extended time, would they ultimately fill the fatpad similarly to RFX3+ tissue?

The cell marker, s-SHIP, was originally found to be specifically expressed in hematopoietic and embryonic stem cells [100]. Though s-SHIP does not mark stem cells in the mammary gland, it does transiently mark a population of basal stem cells in the terminal ductal tips of the prostate in Tg11.5kb-GFP mice. To determine whether RFX3 functions to stabilize the basal cell identity in other contexts and if we observe cell lineage infidelity, it would be revealing to isolate prostate basal cells and knockout RFX3 to determine whether this effects prostate basal cell behavior.

I speculate that *Sonichedgehog (Shh)* signaling may have an important role in basal cell plasticity. Looking more closely at how *Shh* may be mediating changes in cell state and whether it has a direct or indirect relationship with RFX3 would elucidate the mechanism behind cell plasticity in the mammary gland. To bridge the gap between *Shh* and basal cell stabilization I would first do RTqPCR to determine whether Shh pathway genes have decreased expression when RFX3 is knocked out. Secondly, I would use the Shh inhibitor, cyclopamine to see if the phenotype was similar to that of an RFX3 knock out.

Considerations

Generalists vs. Specialists in Mammary Gland

The phenomenon of specialization involves a gradient that at one end or extreme, represents organisms that have similar fitness across a variety of environments. The other extreme includes organisms with high fitness in one environment, and lower fitness in other environments [130]. To think of this idea on a cellular level, one could consider a stem cell as a generalist – having fitness to become several cell types and a differentiated cell as a specialist having a very defined unique roll and limited abilities. Within the mammary gland, a cap cell (myoepithelial progenitor) and a body cell (luminal progenitor) would fall on the gradient, not as a generalist, but closer to that end of the spectrum than a specialist which would include myoepithelial and luminal cells. Mammary epithelial cells that would surpass the generalist-like cap and body cells, and fall on the spectrum closer to the generalist extreme would only be observed during embryonic development when multipotent cells are present driving mammary gland development. Outside of embryonic development, generalist-like cells may be observed during DNA damage, tissue ablation, oncogenesis or cultivation.

Stochasticity in Cell Fate Decisions

Cell fate decisions are tightly regulated in order to produce highly reproducible results. Developmental programs often incorporate stochastic mechanisms to diversify cell types [131]. Stochasticity refers to the randomization of choices within a population of cells or tissue. During cultivation of normal myoepithelial cells, different mechanisms are observed in the cell conversion path. We observe both dedifferentiation then redifferentiation and transdifferentiation. Is it possible that stochasticity plays a role in this conversion path decision? I was not alarmed when I observed that myoepithelial cells could take different conversion paths into luminal cells. There are several examples of ways that cells change or adapt to accomplish a goal. In some cases cells can choose one of multiple mechanisms to complete a biological process. Tostevin *et*

al., found that there are four different mechanisms that a cell can adapt when switching cell polarity [132]. To maintain tissue homeostasis a normally differentiated call can dedifferentiate into a multipotent stem cell [114, 115]. In mammary epithelial cells, the observation that myoepithelial cells have the capacity to not only change their cell fate, but via different mechanisms underscores the innate intelligence of each cell that drives the orchestrated biological processes of an organism.

V. REFERENCES

1. Hens, J.R. & Wysolmerski, J.J. Key stages of mammary gland development: molecular mechanisms involved in the formation of the embryonic mammary gland. *Breast Cancer Res* **7**, 220-224 (2005).
2. Veltmaat, J.M., Mailleux, A.A., Thiery, J.P. & Bellusci, S. Mouse embryonic mammogenesis as a model for the molecular regulation of pattern formation. *Differentiation* **71**, 1-17 (2003).
3. Propper, A.Y. Wandering epithelial cells in the rabbit embryo milk line. A preliminary scanning electron microscope study. *Dev Biol* **67**, 225-231 (1978).
4. Robinson, G.W. Cooperation of signalling pathways in embryonic mammary gland development. *Nat Rev Genet* **8**, 963-972 (2007).
5. Chu, E.Y. *et al.* Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. *Development* **131**, 4819-4829 (2004).
6. Veltmaat, J.M., Van Veelen, W., Thiery, J.P. & Bellusci, S. Identification of the mammary line in mouse by Wnt10b expression. *Dev Dyn* **229**, 349-356 (2004).
7. Jerome-Majewska, L.A. *et al.* Tbx3, the ulnar-mammary syndrome gene, and Tbx2 interact in mammary gland development through a p19Arf/p53-independent pathway. *Dev Dyn* **234**, 922-933 (2005).
8. Watson, C.J. & Khaled, W.T. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**, 995-1003 (2008).
9. Robinson, G.W., Karpf, A.B. & Kratochwil, K. Regulation of mammary gland development by tissue interaction. *J Mammary Gland Biol Neoplasia* **4**, 9-19 (1999).
10. Hogg, N.A., Harrison, C.J. & Tickle, C. Lumen formation in the developing mouse mammary gland. *J Embryol Exp Morphol* **73**, 39-57 (1983).
11. Lyons, W.R. Hormonal synergism in mammary growth. *Proc R Soc Lond B Biol Sci* **149**, 303-325 (1958).
12. Nandi, S. Endocrine control of mammarygland development and function in the C3H/ He Crgl mouse. *J Natl Cancer Inst* **21**, 1039-1063 (1958).
13. Silberstein, G.B. & Daniel, C.W. Glycosaminoglycans in the basal lamina and extracellular matrix of the developing mouse mammary duct. *Dev Biol* **90**, 215-222 (1982).
14. Williams, J.M. & Daniel, C.W. Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev Biol* **97**, 274-290 (1983).
15. Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D. & Werb, Z. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127**, 1041-1055 (2006).
16. Nelson, C.M. & Bissell, M.J. Of extracellular matrix, scaffolds, and signaling: tissue architecture regulates development, homeostasis, and cancer. *Annu Rev Cell Dev Biol* **22**, 287-309 (2006).
17. Gallego, M.I. *et al.* Prolactin, growth hormone, and epidermal growth factor activate Stat5 in different compartments of mammary tissue and exert different and overlapping developmental effects. *Dev Biol* **229**, 163-175 (2001).

18. Ruan, W. & Kleinberg, D.L. Insulin-like growth factor I is essential for terminal end bud formation and ductal morphogenesis during mammary development. *Endocrinology* **140**, 5075-5081 (1999).
19. Zhou, Y. *et al.* A mammalian model for Laron syndrome produced by targeted disruption of the mouse growth hormone receptor/binding protein gene (the Laron mouse). *Proc Natl Acad Sci U S A* **94**, 13215-13220 (1997).
20. Wang, C., Christin, J.R., Oktay, M.H. & Guo, W. Lineage-Biased Stem Cells Maintain Estrogen-Receptor-Positive and -Negative Mouse Mammary Luminal Lineages. *Cell Rep* **18**, 2825-2835 (2017).
21. Rodilla, V. *et al.* Luminal progenitors restrict their lineage potential during mammary gland development. *PLoS Biol* **13**, e1002069 (2015).
22. Van Keymeulen, A. *et al.* Lineage-Restricted Mammary Stem Cells Sustain the Development, Homeostasis, and Regeneration of the Estrogen Receptor Positive Lineage. *Cell Rep* **20**, 1525-1532 (2017).
23. Shackleton, M. *et al.* Generation of a functional mammary gland from a single stem cell. *Nature* **439**, 84-88 (2006).
24. Stingl, J. *et al.* Purification and unique properties of mammary epithelial stem cells. *Nature* **439**, 993-997 (2006).
25. Visvader, J.E. & Stingl, J. Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev* **28**, 1143-1158 (2014).
26. Hennighausen, L. & Robinson, G.W. Information networks in the mammary gland. *Nat Rev Mol Cell Biol* **6**, 715-725 (2005).
27. Muschler, J. & Streuli, C.H. Cell-matrix interactions in mammary gland development and breast cancer. *Cold Spring Harb Perspect Biol* **2**, a003202 (2010).
28. Howard, B.A. & Lu, P. Stromal regulation of embryonic and postnatal mammary epithelial development and differentiation. *Semin Cell Dev Biol* **25-26**, 43-51 (2014).
29. Liu, X. *et al.* ROCK inhibitor and feeder cells induce the conditional reprogramming of epithelial cells. *Am J Pathol* **180**, 599-607 (2012).
30. Makarem, M. *et al.* Developmental changes in the in vitro activated regenerative activity of primitive mammary epithelial cells. *PLoS Biol* **11**, e1001630 (2013).
31. Betterman, K.L. *et al.* Remodeling of the lymphatic vasculature during mouse mammary gland morphogenesis is mediated via epithelial-derived lymphangiogenic stimuli. *Am J Pathol* **181**, 2225-2238 (2012).
32. Gouon-Evans, V., Rothenberg, M.E. & Pollard, J.W. Postnatal mammary gland development requires macrophages and eosinophils. *Development* **127**, 2269-2282 (2000).
33. O'Brien, J., Martinson, H., Durand-Rougely, C. & Schedin, P. Macrophages are crucial for epithelial cell death and adipocyte repopulation during mammary gland involution. *Development* **139**, 269-275 (2012).
34. Gregor, M.F. *et al.* The role of adipocyte XBP1 in metabolic regulation during lactation. *Cell Rep* **3**, 1430-1439 (2013).
35. Hovey, R.C. & Aimo, L. Diverse and active roles for adipocytes during mammary gland growth and function. *J Mammary Gland Biol Neoplasia* **15**, 279-290 (2010).

36. Hinck, L. & Silberstein, G.B. Key stages in mammary gland development: the mammary end bud as a motile organ. *Breast Cancer Res* **7**, 245-251 (2005).
37. Brisken, C. & O'Malley, B. Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol* **2**, a003178 (2010).
38. Bai, L. & Rohrschneider, L.R. s-SHIP promoter expression marks activated stem cells in developing mouse mammary tissue. *Genes Dev* **24**, 1882-1892 (2010).
39. Brisken, C. Hormonal control of alveolar development and its implications for breast carcinogenesis. *J Mammary Gland Biol Neoplasia* **7**, 39-48 (2002).
40. Jena, M.K., Jaswal, S., Kumar, S. & Mohanty, A.K. Molecular mechanism of mammary gland involution: An update. *Dev Biol* **445**, 145-155 (2019).
41. Watson, C.J. & Kreuzaler, P.A. Remodeling mechanisms of the mammary gland during involution. *Int J Dev Biol* **55**, 757-762 (2011).
42. Talhouk, R.S., Chin, J.R., Unemori, E.N., Werb, Z. & Bissell, M.J. Proteinases of the mammary gland: developmental regulation in vivo and vectorial secretion in culture. *Development* **112**, 439-449 (1991).
43. Fata, J.E., Werb, Z. & Bissell, M.J. Regulation of mammary gland branching morphogenesis by the extracellular matrix and its remodeling enzymes. *Breast Cancer Res* **6**, 1-11 (2004).
44. Kratochwil, K. In vitro analysis of the hormonal basis for the sexual dimorphism in the embryonic development of the mouse mammary gland. *J Embryol Exp Morphol* **25**, 141-153 (1971).
45. Howard, B.A. & Gusterson, B.A. Human breast development. *J Mammary Gland Biol Neoplasia* **5**, 119-137 (2000).
46. Visvader, J.E. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev* **23**, 2563-2577 (2009).
47. Silberstein, G.B. Postnatal mammary gland morphogenesis. *Microsc Res Tech* **52**, 155-162 (2001).
48. Deome, K.B., Faulkin, L.J., Jr., Bern, H.A. & Blair, P.B. Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. *Cancer Res* **19**, 515-520 (1959).
49. Kordon, E.C. & Smith, G.H. An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**, 1921-1930 (1998).
50. Daniel, C.W. Regulation of cell division in aging mouse mammary epithelium. *Adv Exp Med Biol* **61**, 1-19 (1975).
51. Daniel, C.W., Aidells, B.D., Medina, D. & Faulkin, L.J., Jr. Unlimited division potential of precancerous mouse mammary cells after spontaneous or carcinogen-induced transformation. *Fed Proc* **34**, 64-67 (1975).
52. Rosen, J.M. On murine mammary epithelial stem cells: discovery, function, and current status. *Cold Spring Harb Perspect Biol* **4**, a013268 (2012).
53. Smith, G.H. & Medina, D. A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. *J Cell Sci* **90 (Pt 1)**, 173-183 (1988).
54. Visvader, J.E. & Smith, G.H. Murine mammary epithelial stem cells: discovery, function, and current status. *Cold Spring Harb Perspect Biol* **3** (2011).
55. Joshi, P.A. *et al.* Progesterone induces adult mammary stem cell expansion. *Nature* **465**, 803-807 (2010).

56. Lindeman, G.J., Visvader, J.E., Smalley, M.J. & Eaves, C.J. The future of mammary stem cell biology: the power of in vivo transplants. *Breast Cancer Res* **10**, 402; author reply 403 (2008).
57. Daniel, C.W., De Ome, K.B., Young, J.T., Blair, P.B. & Faulkin, L.J., Jr. The in vivo life span of normal and preneoplastic mouse mammary glands: a serial transplantation study. *Proc Natl Acad Sci U S A* **61**, 53-60 (1968).
58. Hoshino, K. Regeneration and Growth of Quantitatively Transplanted Mammary Glands of Normal Female Mice. *Anat Rec* **150**, 221-235 (1964).
59. Hoshino, K. Morphogenesis and growth potentiality of mammary glands in mice. II. Quantitative transplantation of mammary glands of normal male mice. *J Natl Cancer Inst* **30**, 585-591 (1963).
60. Prater, M.D. *et al.* Mammary stem cells have myoepithelial cell properties. *Nat Cell Biol* **16**, 942-950, 941-947 (2014).
61. Sleeman, K.E. *et al.* Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. *J Cell Biol* **176**, 19-26 (2007).
62. Stingl, J., Eaves, C.J., Zandieh, I. & Emerman, J.T. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67**, 93-109 (2001).
63. Badders, N.M. *et al.* The Wnt receptor, Lrp5, is expressed by mouse mammary stem cells and is required to maintain the basal lineage. *PLoS One* **4**, e6594 (2009).
64. Spike, B.T. *et al.* A mammary stem cell population identified and characterized in late embryogenesis reveals similarities to human breast cancer. *Cell Stem Cell* **10**, 183-197 (2012).
65. Zeng, Y.A. & Nusse, R. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6**, 568-577 (2010).
66. Plaks, V. *et al.* Lgr5-expressing cells are sufficient and necessary for postnatal mammary gland organogenesis. *Cell Rep* **3**, 70-78 (2013).
67. Wang, D. *et al.* Identification of multipotent mammary stem cells by protein C receptor expression. *Nature* **517**, 81-84 (2015).
68. dos Santos, C.O. *et al.* Molecular hierarchy of mammary differentiation yields refined markers of mammary stem cells. *Proc Natl Acad Sci U S A* **110**, 7123-7130 (2013).
69. Machado, H.L. *et al.* Separation by cell size enriches for mammary stem cell repopulation activity. *Stem Cells Transl Med* **2**, 199-203 (2013).
70. Alexander, C.M. The Wnt Signaling Landscape of Mammary Stem Cells and Breast Tumors. *Prog Mol Biol Transl Sci* **153**, 271-298 (2018).
71. Tharmapalan, P., Mahendralingam, M., Berman, H.K. & Khokha, R. Mammary stem cells and progenitors: targeting the roots of breast cancer for prevention. *EMBO J* **38**, e100852 (2019).
72. Kretzschmar, K. & Watt, F.M. Lineage tracing. *Cell* **148**, 33-45 (2012).
73. Rulands, S. & Simons, B.D. Tracing cellular dynamics in tissue development, maintenance and disease. *Curr Opin Cell Biol* **43**, 38-45 (2016).
74. Carlone, D.L. Identifying Adult Stem Cells Using Cre-Mediated Lineage Tracing. *Curr Protoc Stem Cell Biol* **36**, 5A 2 1-5A 2 18 (2016).

75. Van Keymeulen, A. *et al.* Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**, 189-193 (2011).
76. Rios, A.C., Fu, N.Y., Lindeman, G.J. & Visvader, J.E. In situ identification of bipotent stem cells in the mammary gland. *Nature* **506**, 322-327 (2014).
77. Wuidart, A. *et al.* Quantitative lineage tracing strategies to resolve multipotency in tissue-specific stem cells. *Genes Dev* **30**, 1261-1277 (2016).
78. Sale, S., Lafkas, D. & Artavanis-Tsakonas, S. Notch2 genetic fate mapping reveals two previously unrecognized mammary epithelial lineages. *Nat Cell Biol* **15**, 451-460 (2013).
79. Lafkas, D. *et al.* Notch3 marks clonogenic mammary luminal progenitor cells in vivo. *J Cell Biol* **203**, 47-56 (2013).
80. van Amerongen, R., Bowman, A.N. & Nusse, R. Developmental stage and time dictate the fate of Wnt/beta-catenin-responsive stem cells in the mammary gland. *Cell Stem Cell* **11**, 387-400 (2012).
81. Chang, T.H. *et al.* New insights into lineage restriction of mammary gland epithelium using parity-identified mammary epithelial cells. *Breast Cancer Res* **16**, R1 (2014).
82. Sreekumar, A. *et al.* WNT-Mediated Regulation of FOXO1 Constitutes a Critical Axis Maintaining Pubertal Mammary Stem Cell Homeostasis. *Dev Cell* **43**, 436-448 e436 (2017).
83. Elias, S., Morgan, M.A., Bikoff, E.K. & Robertson, E.J. Long-lived unipotent Blimp1-positive luminal stem cells drive mammary gland organogenesis throughout adult life. *Nat Commun* **8**, 1714 (2017).
84. Shehata, M. *et al.* Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res* **14**, R134 (2012).
85. Jopling, C., Boue, S. & Izpisua Belmonte, J.C. Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. *Nat Rev Mol Cell Biol* **12**, 79-89 (2011).
86. Liao, M.J. *et al.* Enrichment of a population of mammary gland cells that form mammospheres and have in vivo repopulating activity. *Cancer Res* **67**, 8131-8138 (2007).
87. Vaillant, F., Lindeman, G.J. & Visvader, J.E. Jekyll or Hyde: does Matrigel provide a more or less physiological environment in mammary repopulating assays? *Breast Cancer Res* **13**, 108 (2011).
88. Guo, W. *et al.* Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* **148**, 1015-1028 (2012).
89. Panciera, T. *et al.* Induction of Expandable Tissue-Specific Stem/Progenitor Cells through Transient Expression of YAP/TAZ. *Cell Stem Cell* **19**, 725-737 (2016).
90. Hein, S.M. *et al.* Luminal epithelial cells within the mammary gland can produce basal cells upon oncogenic stress. *Oncogene* **35**, 1461-1467 (2016).
91. De Craene, B. & Berx, G. Regulatory networks defining EMT during cancer initiation and progression. *Nat Rev Cancer* **13**, 97-110 (2013).
92. Lamouille, S., Xu, J. & Derynck, R. Molecular mechanisms of epithelial-mesenchymal transition. *Nat Rev Mol Cell Biol* **15**, 178-196 (2014).
93. Nassour, M. *et al.* Slug controls stem/progenitor cell growth dynamics during mammary gland morphogenesis. *PLoS One* **7**, e53498 (2012).

94. Phillips, S. & Kuperwasser, C. SLUG: Critical regulator of epithelial cell identity in breast development and cancer. *Cell Adh Migr* **8**, 578-587 (2014).
95. Damen, J.E., Liu, L., Cutler, R.L. & Krystal, G. Erythropoietin stimulates the tyrosine phosphorylation of Shc and its association with Grb2 and a 145-Kd tyrosine phosphorylated protein. *Blood* **82**, 2296-2303 (1993).
96. Kavanaugh, W.M. & Williams, L.T. An alternative to SH2 domains for binding tyrosine-phosphorylated proteins. *Science* **266**, 1862-1865 (1994).
97. Chacko, G.W. *et al.* Negative signaling in B lymphocytes induces tyrosine phosphorylation of the 145-kDa inositol polyphosphate 5-phosphatase, SHIP. *J Immunol* **157**, 2234-2238 (1996).
98. Rohrschneider, L.R., Custodio, J.M., Anderson, T.A., Miller, C.P. & Gu, H. The intron 5/6 promoter region of the ship1 gene regulates expression in stem/progenitor cells of the mouse embryo. *Dev Biol* **283**, 503-521 (2005).
99. Liu, Q., Shalaby, F., Jones, J., Bouchard, D. & Dumont, D.J. The SH2-containing inositol polyphosphate 5-phosphatase, ship, is expressed during hematopoiesis and spermatogenesis. *Blood* **91**, 2753-2759 (1998).
100. Tu, Z. *et al.* Embryonic and hematopoietic stem cells express a novel SH2-containing inositol 5'-phosphatase isoform that partners with the Grb2 adapter protein. *Blood* **98**, 2028-2038 (2001).
101. Rohrschneider, L.R., Fuller, J.F., Wolf, I., Liu, Y. & Lucas, D.M. Structure, function, and biology of SHIP proteins. *Genes Dev* **14**, 505-520 (2000).
102. Inman, J.L., Robertson, C., Mott, J.D. & Bissell, M.J. Mammary gland development: cell fate specification, stem cells and the microenvironment. *Development* **142**, 1028-1042 (2015).
103. Fu, N.Y., Nolan, E., Lindeman, G.J. & Visvader, J.E. Stem Cells and the Differentiation Hierarchy in Mammary Gland Development. *Physiol Rev* **100**, 489-523 (2020).
104. Wuidart, A. *et al.* Early lineage segregation of multipotent embryonic mammary gland progenitors. *Nat Cell Biol* **20**, 666-676 (2018).
105. Abdul-Manan, N. *et al.* Structure of Cdc42 in complex with the GTPase-binding domain of the 'Wiskott-Aldrich syndrome' protein. *Nature* **399**, 379-383 (1999).
106. Humphreys, R.C. *et al.* Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. *Development* **122**, 4013-4022 (1996).
107. Centonze, A. *et al.* Heterotypic cell-cell communication regulates glandular stem cell multipotency. *Nature* **584**, 608-613 (2020).
108. Le Guelte, A.a.M.I.G. A self-limiting circuit regulates mammary cap cell plasticity through TGF-beta signaling. *BioRxv* (2021).
109. Seldin, L. & Macara, I.G. DNA Damage Promotes Epithelial Hyperplasia and Fate Mis-specification via Fibroblast Inflammasome Activation. *Dev Cell* **55**, 558-573 e556 (2020).
110. Janky, R. *et al.* iRegulon: from a gene list to a gene regulatory network using large motif and track collections. *PLoS Comput Biol* **10**, e1003731 (2014).
111. Chen, B. *et al.* Auto-fatty acylation of transcription factor RFX3 regulates ciliogenesis. *Proc Natl Acad Sci U S A* **115**, E8403-E8412 (2018).

112. McDermott, K.M., Liu, B.Y., Tlsty, T.D. & Pazour, G.J. Primary cilia regulate branching morphogenesis during mammary gland development. *Curr Biol* **20**, 731-737 (2010).
113. Mitchell, E.H. & Serra, R. Normal mammary development and function in mice with Ift88 deleted in MMTV- and K14-Cre expressing cells. *Cilia* **3**, 4 (2014).
114. Tata, P.R. *et al.* Dedifferentiation of committed epithelial cells into stem cells in vivo. *Nature* **503**, 218-223 (2013).
115. van Es, J.H. *et al.* Dll1+ secretory progenitor cells revert to stem cells upon crypt damage. *Nat Cell Biol* **14**, 1099-1104 (2012).
116. Jopling, C. *et al.* Zebrafish heart regeneration occurs by cardiomyocyte dedifferentiation and proliferation. *Nature* **464**, 606-609 (2010).
117. Zhou, Q., Brown, J., Kanarek, A., Rajagopal, J. & Melton, D.A. In vivo reprogramming of adult pancreatic exocrine cells to beta-cells. *Nature* **455**, 627-632 (2008).
118. Sanjana, N.E., Shalem, O. & Zhang, F. Improved vectors and genome-wide libraries for CRISPR screening. *Nat Methods* **11**, 783-784 (2014).
119. Shalem, O. *et al.* Genome-scale CRISPR-Cas9 knockout screening in human cells. *Science* **343**, 84-87 (2014).
120. Richards, S.A., Lounsbury, K.M. & Macara, I.G. The C terminus of the nuclear RAN/TC4 GTPase stabilizes the GDP-bound state and mediates interactions with RCC1, RAN-GAP, and HTF9A/RANBP1. *J Biol Chem* **270**, 14405-14411 (1995).
121. Hu, Y. & Smyth, G.K. ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. *J Immunol Methods* **347**, 70-78 (2009).
122. Bonnafe, E. *et al.* The transcription factor RFX3 directs nodal cilium development and left-right asymmetry specification. *Mol Cell Biol* **24**, 4417-4427 (2004).
123. Shen, S. & Clairambault, J. Cell plasticity in cancer cell populations. *F1000Res* **9** (2020).
124. Kim, T.K. & Shiekhattar, R. Architectural and Functional Commonalities between Enhancers and Promoters. *Cell* **162**, 948-959 (2015).
125. Pennacchio, L.A., Bickmore, W., Dean, A., Nobrega, M.A. & Bejerano, G. Enhancers: five essential questions. *Nat Rev Genet* **14**, 288-295 (2013).
126. Giorgetti, L. & Heard, E. Closing the loop: 3C versus DNA FISH. *Genome Biol* **17**, 215 (2016).
127. Di Giammartino, D.C., Polyzos, A. & Apostolou, E. Transcription factors: building hubs in the 3D space. *Cell Cycle* **19**, 2395-2410 (2020).
128. Nakagawa, T., Yoneda, M., Higashi, M., Ohkuma, Y. & Ito, T. Enhancer function regulated by combinations of transcription factors and cofactors. *Genes Cells* **23**, 808-821 (2018).
129. Malinova, A., Veghini, L., Real, F.X. & Corbo, V. Cell Lineage Infidelity in PDAC Progression and Therapy Resistance. *Front Cell Dev Biol* **9**, 795251 (2021).
130. Vilanova, C., Baixeras, J., Latorre, A. & Porcar, M. The Generalist Inside the Specialist: Gut Bacterial Communities of Two Insect Species Feeding on Toxic Plants Are Dominated by Enterococcus sp. *Front Microbiol* **7**, 1005 (2016).

131. Johnston, R.J., Jr. & Desplan, C. Stochastic mechanisms of cell fate specification that yield random or robust outcomes. *Annu Rev Cell Dev Biol* **26**, 689-719 (2010).
132. Tostevin, F., Wigbers, M., Sogaard-Andersen, L. & Gerland, U. Four different mechanisms for switching cell polarity. *PLoS Comput Biol* **17**, e1008587 (2021).