MOLECULAR DISSECTION OF BREAST LUMINAL TRANSCRIPTION NETWORKS

Francesca Bargiacchi

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of doctor of philosophy in the department of Pharmacology in the school of Medicine.

Chapel Hill
2015

Approved by:

H. Shelton Earp

Charles Perou

Gary Johnson

Michael Emanuele

Qing Zhang

ABSTRACT

Francesca Bargiacchi: Molecular Dissection of Breast Luminal Transcription Networks
(Under the direction of H. Shelton Earp)

During mammary development, cellular differentiation and lineage commitment to various epithelial and mesenchymal cell types are driven by hormonal and paracrine signaling mechanisms. Understanding mechanisms that govern differentiation into distinct cell populations is critically important for a complete understanding of development and breast tumorigenesis. Previous studies have shown that retroviral transduction of fibroblasts with four transcription factors can initiate the conversion of a somatic cell into an embryonic stem cell-like state with capabilities of differentiating into cell types of all three germ layers. The goal of the proposed study was to determine whether mammary specific transcription factors (TFs) could directly induce transdifferentiation to an ER+/luminal cell phenotype in mouse embryonic fibroblasts via an iPS-type approach. After screening 9 candidate TFs for their abilities to induce various epithelial-specific and breast-specific attributes, we focused subsequent efforts on ESR1, FOXA1, PGR, and GATA3. In human mammary epithelial cells, ectopic overexpression of these TFs had both unique and overlapping contributions toward inducing the expression of genes responsible for luminal differentiation. Combining these TFs in fibroblasts created combinations that produced a luminal progenitor phenotype as defined by gene expression histological markers for epithelial/breast development and functionality. Examination of the gene expression patterns induced by these combinations using a classifier of differentiation status identified several genes

that drive a transition toward the luminal subtype. Since there are presently no cell lines or mouse models of luminal A/ER+ breast cancers, the creation of such of a line would be of tremendous value.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1: MOLECULAR DISSECTION OF BREAST LUMINAL TRANSCRIPTION NETWORKS

Introduction

The molecular characterization of human breast cancers and normal mammary developmental processes has revealed commonalities that have proven to be useful in clinical cancer diagnostics and in understanding normal and neoplastic breast biology. During mammary development, cellular differentiation and lineage commitment are driven by distinct growth cycles under the control of local epithelial and mesenchymal paracrine signaling mechanisms. Multipotent stem cells give rise to various luminal and myoepithelial cell types organized in lobular and ductal architectures that ultimately comprise normal breast tissue. Likewise, our lab and others have established that human breast cancers consist of multiple "intrinsic" subtypes, several of which are defined by their similarity to normal mammary cell counterparts on the basis of gene expression patterns.

These tumor subtypes differ in their responses to therapy and survival outcomes. Importantly, luminal A tumors are the most common and diagnostically are ER+ and PR+. Luminal A tumors tend to grow slowly, and patients with these tumors have an overall better prognosis than those with the other subtypes; yet these tumors do not respond well to chemotherapeutic agents. In order to better elucidate the phenotypic and genotypic variations between the tumor subtypes, researchers have used breast cancer cell lines extensively to investigate specific pathobiologies for each, which may help explain clinical variables such as prognosis, response to therapies, and sites of metastasis. Unfortunately, luminal A tumors have

yet to be successfully propagated in vitro and to be captured as an immortal cell line (with MCF7 being a luminal B line). Thus, creating new cell lines that capture the pertinent aspects of the luminal A cell-type biology would have tremendous value for a better understanding of these tumors and for enabling new drug development.

To address the goal of creating a ER+/luminal A cell line, we propose an approach similar to the one taken to develop induced pluripotent stem cells (i.e., iPS cells). In this approach, we propose using two different lineage cell lines and transducing them with candidate lineage determining genes, in an attempt to transdifferentiate this basal-like cell line into a luminal A–like cell line. The nine candidate transcription factors chosen are all highly expressed in ER+/luminal A cells and are thought to be luminal lineage-specific factors. Our preliminary data suggested that ectopic overexpression of these transcription factors in HMECs have both unique and overlapping contributions toward inducing the expression of genes responsible for luminal differentiation. One of these transcription factors, GATA3, is expounded upon in Chapter 3. The process that results in the development of a luminal-like cell line requires a thorough understanding of mammary development in both mice and humans. In addition, knowledge of which mouse mammary models are utilized for the study of development is key, particularly with regard to gene knockout models, which result in either no mammary system or an abrogated one. Concluding the introduction with a discussion regarding reprogramming/transdifferentiation emphasizes not only the technical approach of creating the cell line but also the biological process necessary to achieve the change in the cell line lineage. The aim of this introduction is to give a relevant and meaningful understanding of the complexities surrounding reprogramming and defining breast luminal transcriptional regulatory mechanisms.

1.1 Overview of Mammary Development

1.1.1 General Process of Development

In 1758, zoologist Carl Linneaus noted the unique features of the mammary gland and, as the father of modern day taxonomy, created a new group he called Mammalia. This new group, Mammalia, was based on the merger of two groups: quadrupeds, or animals with 4 feet, and aquatic cetaceans, which include dolphins and whales (Oftedal, 2002). It is thought that the mammary gland is derived from apocrine glands, particularly ones in the skin. Darwin and others postulated that the complex processes involving the development of a lactating gland confer a survival benefit for offspring (Oftedal, 2002). In the early 1900s it was discovered that the ovary and pituitary glands are involved in mammopoiesis, particularly with regard to the secretion of mammotropic substances. Later, in 1928, castrated virgin rabbits were injected with pituitary extracts from lactating animals, resulting in lactation and demonstrating that lactogenic stimuli could induce lactation in the absence of ovarian hormonal regulation (Stricker and Grueter, 1928). In order to better understand development in vitro, Klaus Kratochwil isolated mammary gland rudiments of 12- to 14-day mouse embryos, typically developed in organ culture, forming a developed breast, including supporting adipocytes; however, rudiments taken from 16-day embryos failed to develop under the same conditions, suggesting that mammary development in mice is a time-specific process (Kratochwil, 1969). While the mid-20th century was a time focused on understanding the key players involved in mammopoiesis, as well as the ability to grow mammary tissue in vitro, it was not until the late 20th century that the cloning of steroid receptors, mechanisms of downstream signaling, and generation of genetically engineered mouse models truly matured, offering researchers a plethora of valuable information (Hennighausen and Robinson, 2001). The mammary gland, like other reproductive organs, is one that develops

principally after birth. Typically under the control of hormones, development occurs through distinctive stages throughout the life span of a mammal, starting with the embryo, continuing with puberty, and concluding with reproduction. At each developmental stage, there are cues specific to that developmental stage that initiate changes in both the mammary cells and the surrounding stroma. Functionally speaking, the mammary gland's role is to provide ample nutrition for offspring. Comprised of milk protein and fat, milk also provides immune factors such as antibodies that offer some protection to offspring against disease (Rogier et al., 2014). The mammary gland halts development after birth, until puberty, with the growth and expansion of milk ducts. It matures via lobuloalveolar development during pregnancy. How these intricate signaling processes come together at different stages of mammary development will be discussed further, with a primary focus on mouse mammary development. There are several reasons for this: First, the ability to genetically engineer mice makes them an attractive model to study. Second, there are developmental similarities between humans and mice, although it should be noted that there are some subtle structural and hormonal differences between the two (Watson and Khaled, 2008).

1.1.2 Mammary Development in the Embryo

During embryonic development in mice, five pairs of ectodermal placodes appear between Days 10 and 11. These placodes then form buds and continually grow until Day 15. These buds turn into epithelial cells and once embedded in the mesenchyme, elongate to form sprouts. At this point the sprout develops a lumen, or a hollow opening that connects with the skin. At the end of pregnancy, the sprout, called the nipple sheath, becomes a gland that invades the adjacent fat pad. For an illustration, please see Figure 1.1.

Studies have shown that the mesenchyme from the mammary region is responsible for the mammary bud's epithelial development. Initial signals for epithelial development come from the mesenchyme and include transcription factors' estrogen and androgen receptors, Lef-1, Msx1, and the growth factors BMP4 and FGF7. These factors are induced in a paracrine fashion by the developing bud and are silenced at later stages (Heuberger et al., 1982). As mentioned earlier, mammary gland development was derived from apocrine glands from the skin. Hence, some of these transcription factors are not unique to the development of the mammary gland. Once the bud is formed, Lef-1 and Msx1/2 are turned off, leading to a lack of other ectodermal appendages such as teeth and hair (Hennighausen and Robinson, 2001). With the formation of the bud comes the elongation of the primary spout, driven largely by the mammary bud's expression of parathyroid hormone related protein (PTHrP). PTHrP signals to the adjacent mesenchyme (which expresses PTHrP receptors) to give rise to the nipple. By Day E18.5, several small, tree-like glands have developed and are dependent on factors supplied by the fat pad. Signaling from PTHrP continues, facilitating the activation of downstream marker BMP4.

From a mouse model perspective, PTHrP knockout mice exhibit greatly reduced ductal branching morphogenesis. Finally, PTHrP and BMP4 induce Msx2 to regulate the suppression of hair follicle formation (Watson and Khaled, 2008). The development of the mammary gland elicits signals from the nascent mesenchyme, which is the primary driver of epithelial bud development, while working synergistically with the epithelial sprout to further differentiate into other ectoderm lineages by rapid changes in transcription factor expression. Interestingly, development in the mouse arrests at around Day E18.5 and will not continue until puberty, which is when branching morphogenesis and lumen formation predominate.

1.1.3 Mammary Development During Puberty

During this time, the development of the mammary gland relies heavily on hormones, leading to the creation of a complex ductal network. This complex network is a result of branching morphogenesis, which is regulated by the various components of the mammary gland. This includes cellular players such as the stromal compartment-fibroblasts, adipocytes, and immune cells, as well as the mammary epithelial cells themselves. Hormones and growth factors act in both a paracrine and autocrine fashion to control the growth of the ductal network, assisting the invasion into the fat pad. At the tip of these ducts are terminal end buds (TEBs), which comprise of two layers of cells. The outer layer consists of cap cells, which are highly proliferative, hormone receptor-negative cells involved in the remodeling and extension of the bud. In contrast, the body cells, which line the inner portion of the TEB and are responsible for lumen development, are less proliferative than the cap cells, and are responsive to hormonal/lactogenic stimulus. Please see Figure 1.2. In addition, ductal elongation will result in bifurcation of the existing ducts to create branches. In the mouse, this process will be complete by 10–12 weeks of age. At that point, TEBs have invaded the fat pad to the full extent and growth of the branches has ceased (Watson and Khaled, 2008). One of the key regulators of ductal morphogenesis is estrogen. Estrogen has two receptors, ESR1 (ERα) and ESR2 (ERβ). ERα has been shown to be critical for mouse mammary development, as ERα-null mice fail to develop TEBs and the ducts that do form do not invade the fat pad (Mallepell et al., 2006).

Along with the process of morphogenesis comes lumen formation. Lumen formation is a process that occurs early in the development of mammals and continues through organ development, particularly in the lungs and kidneys. Lumen formation in the breast is thought to form via the death of luminal cells in the space of the TEB. In addition to undergoing apoptosis,

cells within the TEB become polarized, recognizing via contact with the basement membrane just prior to lumen formation. Once polarity is established, cells that do not become polarized undergo apoptosis (Debnath et al., 2002). Some key proteins that regulate polarity and lumen formation include ROBO1 and SLIT2. Knockout mouse models with either ROBO1 or SLIT2 deficiency not only exhibit defects in TEB structure but also ductal abnormalities that result in separation between the luminal epithelial and myoepithelial cell layers (Macias et al., 2011). These data highlight the complex signals needed for lumen development.

There are several transcription factors that drive this stage of mammary development. As mentioned before, estrogen signaling is critical for branching morphogenesis and becomes an invaluable signal for alveologenesis during pregnancy. ERα-null mice not only have reduced levels of prolactin, which augments milk production, but also have insufficient progesterone synthesis, which is necessary for ductal development (Hennighausen and Robinson, 2001). Progesterone mediates its effects on the development of the mammary gland through two main isoforms, PR-A and PR-B. Mice lacking both isoforms become anovulatory and develop a severely growth-retarded mammary gland. Progesterone signaling controls cell proliferation through the activation of the Wnt pathway, promoting alveolar formation (Brisken and Duss, 2007). Although estrogen mediates much of the branch morphogenesis, its regulation of progesterone receptors (PRs) results in the elaborate extension of the branches, known as side branching, in the virgin mouse (Fernandez-Valdivia et al., 2009; Ismail et al., 2003). The role of estrogen and progesterone signaling will be further discussed in Section 3.

While a significant portion of mammary development is largely regulated by estrogen and progesterone, certain growth factors mediate their actions. The TGF-β family of receptors and ligands was first thought to be involved in mammary development in the late 1980s and

activation of TGF-β was thought to correlate with hormonal, differentiation status, as well as proliferation/apoptosis (Silberstein and Daniel, 1987). Normally TGF-β is expressed throughout all phases of mammary development. TGF-β1 is regulated by estrogen and progesterone, and the downstream effectors of TGF-β1 regulate morphogenesis in a variety of ways, including apoptosis, matrix remodeling, and loss of proliferation (Moses and Barcellos-Hoff, 2011). Although TGF-β signaling can induce apoptosis during morphogenesis, within the mammary gland it can exert its effects on the stroma via the production of extracellular matrix (ECM) (Moses and Barcellos-Hoff, 2011). This production of ECM may be an attempt to control the growth of mammary epithelial cells during development (Gomm et al., 1991).

1.1.4 Pregnancy and Lactation

During pregnancy, under the regulation of progesterone, the mammary gland must undergo morphological change to prepare for lactation. This includes the differentiation of specific luminal epithelial cells that can synthesize and secrete milk. Alveologenesis is the production of alveoli, or milk sacs, that will store and secrete milk and is under the regulation of both progesterone and prolactin (PRL). As mentioned earlier, PGR is critical for this stage of development, as PGR-null mice do not form alveoli (Brisken et al., 1998). Because of the massive expansion in mammary epithelial cell growth, signals are sent to further define the milk-producing luminal cells from the contractile-competent myoepithelial cells. The factors that control the fates of these cells have not been well defined; however, some studies have shed light on the potential regulatory mechanisms. There are studies suggesting that mammary luminal development mimics the development and differentiation of T-helper (Th) cells (Khaled et al., 2007; Watson and Khaled, 2008). Briefly, the idea is that the polarization of Th cells into either Th1 or Th2 lineages occurs via the action of cytokines. These cytokines then activate Stat4 and

Stat6, whose downstream effector molecule is GATA3. GATA3 is necessary for Th2 development due to its role in chromatin remodeling and facilitates the expression of IL-4, IL-13, and IL-5 cytokines (Watson and Khaled, 2008). Thus, the regulation of these cytokines promotes Th2 while suppressing Th1 differentiation. Mouse knockout models of IL-4/IL-13 (responsible for Th2 regulation) resulted in reduced proliferation and milk protein production, as well as delayed alveolar development during gestation, a similar phenotype to the Stat6 knockout mouse model (Khaled et al., 2007). These data also suggested that under lactogenic stimulation, expression of Th1 cytokines was suppressed and that the mechanisms that mediate Th2 development also determine the fate of the mammary luminal epithelial lineage. The IL-4/IL-13/Stat6/GATA3 pathway in mammary development is illustrated in Figure 1.3. GATA3, a potent regulator of alveolar differentiation, is discussed in further detail in Chapter 3. In addition to Stat6 signaling, Stat5, a potent downstream signal from prolactin receptor (PRLR), is critical for the differentiation of mammary alveoli during pregnancy (Liu et al., 1997). Although the mechanisms are still unclear, it appears that Stat5a and Stat6 work in parallel to induce IL-4/IL-13 expression.

1.1.5 Involution

As with pregnancy and lactation, the immune-regulated genes are involved in mouse mammary involution, a process of tissue remodeling and cell death. There are two phases of involution: the first occurs within hours of weaning and entails the accumulation of milk within the alveolar lumen and subsequent decrease of lactogenic hormones, while the second phase is a signal initiated by systemic hormonal regulation. This results in a remodeling of the ECM and the basement membrane of the ductal-lobular architecture due to the activation of matrix metalloproteases. Ultimately, the immune system is then recruited to clear the “debris” (Stein et

al., 2004). All of this occurs within 2 weeks post weaning, making involution a quick and efficient process in the mouse. Examining Phase 1 in more detail, we determined that this phase is responsible for the apoptotic events within the alveolar lumen and is initiated by Stat3 signaling. Stat3 is activated by the leukemia initiation factor (LIF). Downstream targets of Stat3, IGFPB-5 and C/EBPδ have been shown to be instrumental in the apoptotic response. Knockout models of these two genes have suggested a diminished ability to undergo involution (Thangaraju et al., 2005; Tonner et al., 2002). Cells with activated Stat3 signaling begin to shed, filling the lumen and expressing proapoptotic markers such as Caspase-3, Bax, and decreased levels of Bcl-x. Other downstream targets of Stat3 ensure that transition to the second phase of involution ensues.

Approximately 48 hours after Phase 1, Phase 2 begins, primarily by remodeling the lobulo-ductal architecture. This is achieved by the stromal secretion of matrix metalloproteases, which not only degrade the lobulo-ductal supportive matrix but also initiate further apoptosis through the detachment of the ECM to the epithelial cells. Adipocytes migrate to the newly remodeled matrix and differentiate via MMP3 and plasminogen (Watson, 2006). Ultimately, NF-κB signaling recruits immune cells, primarily macrophages, to phagocytose and remove the cells and debris, a process that involution mediates.

The four main stages of mouse mammary development have unique roles and complex signaling mechanisms regulated by hormones, growth factors, and lactogenic signals. Interestingly, certain phases of mammary development, unlike the phases of development of most organs in the body, are induced by the voluntary events of pregnancy and involution. Such events, as some studies have suggested, may lower a woman's overall risk for developing premenopausal breast cancer (Newcomb et al., 1994), highlighting how development may

influence the risk of neoplastic events during child-bearing years. These different stages of development also complement the mouse mammary differentiation hierarchy, as the earlier stages correspond to a less differentiated (more stem cell-like) phenotype, while lactation is more in line with a luminal differentiated phenotype. It is essential in understanding the developmental phases especially when trying to recapitulate a mammary epithelial cell line de novo. This will be touched upon in Chapter 2.

1.2 Current Mouse Models of Mammary Development

1.2.1 Genes Required for Mammary Function

Mammary gland development and function are tightly controlled processes that are heavily regulated by hormones and growth factors both produced by the epithelial and/or stromal compartments, as well as by more distant locations such as the ovaries and/or pituitary gland. Understanding the critical mediators of development and function involve multiple cell types, particularly within the lobulo-ductal network. The focus will be on the myoepithelial and more differentiated basal/luminal epithelial cells. There are several considerations when determining which mouse model to utilize for the scientific question at hand. First, in which phase of development does the scientific interest rest? For example, as mentioned in Section 1, hormonal dependence and necessity become apparent during puberty, when such signals are required for ductal elongation and morphogenesis. An ERα-knockout model would thus be a poor choice for studying embryonic development of the mammary gland.

Another important consideration is the phenotype of the mouse model. GATA3, a pioneer factor known to be necessary for breast luminal development, is also embryonically lethal (Lim et al., 2000). An effective way to circumvent this issue is the creation of a conditional knockout, in which a mammary specific promoter (MMTV for example) in a Cre-lox transgenic expression

system will delete GATA3 when the MMTV promoter is turned on, presumably in breast tissue. This method not only assures expression in the tissue types being studied but also prevents the effects of the knockout on the overall development of the embryo. More recently, there has been interest in comparing normal breast tissue to breast tissue in tumor development. Here, one may want to focus on certain cell types, whether myoepithelial in origin or the cell source that gave rise to a luminal tumor. Inducible Cre knock-in models for these questions may include the use of a KRT14, KRT8, or KRT18 promoter, depending on the mammary cell type desired (Keymeulen et al., 2011).

Some of the key regulators of mammary development and their phenotypic characteristics when knocked out are shown in Table 1.1. The essential gene knockout developmental models either result in a breast phenotype that is severely abrogated or completely defective in morphology and/or function. It is not surprising to see that loss of ERα results in not only branching morphogenesis loss but also a loss in overall reproductive function. Interestingly, knocking out ERβ in place of ERα results in the normal development of the mammary gland, which suggests that both isoforms have unique downstream targets and that functional ERβ cannot compensate for the loss of ERα (Couse and Korach, 1999). Loss of both reproductive and lactation function is also observed with progesterone receptor and prolactin knockout models, suggesting that genes involved with specific developmental cues across multiple tissue types may have a profound effect on those systems when lost. Reduced functions are even seen in a PRLR-heterozygous knockout mouse model (Ormandy et al., 1997).
LEF-1, a transcription factor involved in canonical WNT signaling, is a key initiator of placode development in embryonic mice. It is also a key regulator of ectodermal lineages, including teeth and hair (Kratochwil et al., 2003). Two other LEF-1 downstream targets, MSX1 and MSX2, also

arrest mammary development at the embryonic bud stage, however only when both are knocked out is bud development arrested. This suggests that both genes may have redundant functions when either one is knocked out (Cowin and Wysolmerski, 2010).

During each stage of development, signaling cues give rise to particular cellular types, be it multilineage, as seen in embryonic development, or within a specific lineage, such as epithelial differentiation, as seen in pregnancy. During the embryonic stage, the formation of a rudimentary gland includes initial bud formation, which gives rise to the mammary epithelial cells. These epithelial cells are poorly differentiated and are not hormone dependent at this stage of development. Exhibiting a "basal" phenotype, embryonic cells at this phase express Keratin 14 and have multipotent capabilities. Keratin 14 positive progenitor cells arise from the embryonic epidermis at embryonic Day 17 (E17). Studies have shown that at birth, mouse mammary glands consist of a small network of rudimentary ducts comprising of myoepithelial and luminal cells. The myoepithelial cells are positive for Keratins 5 and 14 and smooth muscle actin (SMA). These are considered markers for the basal phenotype seen in both mammary development and cancer and will be further discussed later in this chapter.

Likewise, luminal cells at birth are positive for a different set of keratins, mainly Keratins 8 and 19. Their multipotent abilities are worth noting since Keratin 14–positive cells isolated from the embryonic phase of development can recapitulate all mammary epithelial lineages (Keymeulen et al., 2011). The definition of cellular subsets is made possible by in vivo reconstitution assays. Such assays allow for the ability to determine mammary-repopulating capabilities. This is achieved by injecting cells in limiting dilutions into a cleared mammary fat pad (Visvader, 2009). Cells with the ability to regenerate a mammary gland were also shown to have self-renewal properties. As mammary epithelial cells become more defined, particularly

during puberty and/or pregnancy, such cells lose their self-renewal and multipotent functions. Studies have provided evidence for the abilities of both bi- and unipotent progenitor cells that could contribute to the expansion of the mammary gland during puberty and pregnancy (Keymeulen et al., 2011). Interestingly, others have shown that while prospective alveolar progenitors fail to form outgrowths in vivo, when such cells are coinjected with Matrigel into mammary fat pads, they in fact can create small, branched mammary structures (Jeselsohn et al., 2010; Vaillant et al., 2011).

There are several markers used to determine cellular plasticity and differentiation in the mouse. The four main markers are CD24, CD29 (ITGβ1), CD61 (ITGβ3), and CD49f (ITGα6). Initial studies have shown that CD24 is an excellent marker in delineating between epithelial and stromal cells (Sell, 2013). Within the mammary epithelial lineage, CD24 expression also delineates between the basal ($CD24^{mod/+}$) and luminal ($CD24^{hi}$) subpopulations (Sleeman et al., 2006). Interestingly, the same study also recognized that sorted mouse mammary epithelial cells with high CD24 expression had limited repopulation capacity, suggesting that the $CD24^{low}$ cell fraction was responsible for the ability to repopulate a mammary gland. $CD29^{hi}$ and $CD49f^{hi}$ have been shown to represent cells with repopulation capacity (Wang, 2006).

While mammary stem cells (MaSC) exhibit the $CD24^{mod/+}/CD29^{hi}/Cd49f^{hi}$ phenotype, these three markers are difficult for use in defining the luminal progenitor phenotype. In addition, they also include the myoepithelial and basal progenitor cells. The highest expression levels of CD24, as confirmed by fluorescence-activated cell sorting (FACS), occurs in more differentiated mammary epithelial cells. Although neither CD29 nor CD49f were good markers for capturing the luminal progenitor population, two markers have been determined to define the luminal progenitor population, CD61 and CD133 (prominin-1), although CD61 is the more

commonly used marker of the two (Asselin-Labat et al., 2007; Sleeman et al., 2006). The mature luminal population ($CD24^{+}/CD29^{low}/CD49f^{low/-}/CD61^{-}$) will have loss of expression of the integrins CD29, CD49f, and CD61 and have higher expression of CD24 than the MaSC/progenitor cell types (Lim et al., 2010). There will also be few progenitor cells within this population (Lim et al., 2010). As a result, luminal cells are considered to be unipotent with no self-renewal capabilities (see Figure 1.4).

Studies suggest that EpCAM is superior to CD24 as a marker for mouse mammary luminal cells. EpCAM is used in place of CD24 in human mammary studies due to its specificity for luminal/luminal progenitor populations, as CD24 expression is found in basal progenitor/ MaSC and even some stromal cells (Oakes et al., 2014; Smalley, 2015; Visvader, 2009). Depending on the experimental question, either marker is considered acceptable for the isolation of progenitor/luminal mammary cells; however, if the desire is to study the basal/myoepithelial component of the mammary gland, then EpCAM is a better marker for separating those cells from the luminal component (Smalley, 2015).

Recently, there has been uncertainty with respect to the roles of stem versus progenitor cells. Progenitor cells in the mammary gland are thought to be uni-/bipotent and have little to no self-renewal capabilities (Rios et al., 2014; Stingl et al., 2001). Meanwhile, mammary stem cells are multipotent and have self-renewal capabilities, resulting in asymmetrical division when activated. Mammary transplantation assays have shown the presence of multipotent stem cells within the TEBs, and while their frequency is low, questions remain regarding their role in development during puberty/pregnancy and homeostasis (Keymeulen et al., 2011; Smith, 1996). Therefore, the multipotent stem cell is considered to be at the top of the mammary cellular hierarchy, giving rise to either bipotent MaSC or to luminal and basal-limited progenitor cells.

Rios and colleagues (2014) determined that bipotent MaSCs could collectively be called multipotent MaSCs due to the fact that these cells contributed to the development of alveolar branches during pregnancy. Notably, these cells were traced over a couple of pregnancy and involution cycles, suggesting that they were involved in ductal remodeling (Rios et al., 2014; Visvader et al., 2014). These bipotent MaSCs may also give rise to lineage-specific progenitor cells (basal, luminal). Ultimately, these cells have been found to maintain mammary gland homeostasis and development, postnatally, instead of the rare multipotent stem cell that had been shown to give rise to a completely functional mammary gland (Mark et al., 2006; Visvader et al., 2014). It is worth mentioning while discussing the subtle nuances between the multi-/bipotent progenitor cells and their functions that the characteristics and functions of the differentiated luminal cell have been better characterized. Transplantation studies with mammary stem cells typically have basal-like characteristics, perhaps in part due to their basal apical location in the TEBs and their role of giving rise to a functional ductal network (Smith and Boulanger, 2003).

As mentioned earlier, lineage-committed progenitor cells have features that contain either basal or luminal characteristics, although some lineage-tracing studies suggest a bipotent progenitor may contain features that are passed onto and maintained in its differentiated counterparts, as has been seen using Keratin 5–labeled cells (Rios et al., 2014). Keratin 8–labeled stem cells have been shown to give rise to a luminal progenitor cell that can give rise to both luminal and milk-producing cells. In addition, these labeled cells were found to be self-renewing, after as many as 3 pregnancy/involution cycles, indicating that these cells are not being replaced by a mammary stem cell (Keymeulen et al., 2011).

Luminal cells in the mouse tend to exhibit heterogeneous characteristics; hence only a subset of these cells express the Estrogen Receptor (ESR1) (Petersen and Deurs, 1997). It is

widely thought that cells expressing ESR1 within the luminal component are rarely activated (Clarke et al., 1997); however, others have discovered rare ESR1-positive cycling cells (Booth and Smith, 2006). This suggests that not all ESR1-positive cells are quiescent; they may represent a minor fraction of the luminal progenitor population. Luminal progenitor (LP) cells expressing ESR1 tend to be $CD24^{+}/Sca1^{+}/c\text{-}kit^{+}$. This is in contrast to LP cells without ESR1, whose expression of lineage markers' $CD24^{+}/Sca1^{-}/c\text{-}kit^{+}$ only varies by the loss of Sca1 (Sleeman et al., 2007).

Both LP populations exhibit bipotent capabilities; however, in the mouse at least, ESR1-expressing LPs appear to have a survival advantage over ESR1-negative LPs in the presence of oestrogens. The basal level of oestrogens in the mammary gland is generally low and thus shifts the LP population from ESR1-negative to ESR1-expressing luminal cells. The ESR1-positive LP cells are likely to be ductal-restricted progenitor due to their expression of ESR1 and FOXA1. ESR1 and FOXA1 are known to be involved in ductal, not alveolar, development, which may explain why ESR1-negative LP cells have been shown to be alveolar-restricted progenitor cells Shehata et al., 2012). These cells are long living and prepare the gland for pregnancy and involution, giving rise to functional differentiated luminal cells (Keymeulen et al., 2011).

As luminal cells are composed of different cell types, lineage-tracing studies have been performed using an inducible KRT18 reporter (Keymeulen et al., 2011). Like KRT8, KRT18 is a marker for luminal cells. KRT18-positive cells in the Keymeulen et al. (2011) study suggest that KRT18 cells do not clonally expand, even during pregnancy and lactation. Therefore, it is plausible that KRT18-positive cells are committed luminal cells with a low proliferative profile. KRT8 and KRT18 are markers that also delineate basal from luminal subtypes in mammary tumors (Herschkowitz et al., 2007; Perou et al., 2000; Prat et al., 2010; Ross and Perou, 2001).

Differentiated luminal cells will have high CD24 expression, low to moderate CD29 expression, and low to no CD61 expression (Sleeman et al., 2006). The functions of $CD24^{+}/CD29^{low/mod}/CD61^{-}$ luminal epithelial cells are well defined, as they exhibit two main responsibilities: First, ductal luminal cells are responsible for the lining of the duct, which assists in draining the lobuloalveolar structures during lactation. Second, during pregnancy, signaling cues (more on this in the next section) initiate the production of milk in an otherwise quiescent cell, which explains why such differentiated cells are difficult to culture in vitro (Clarke et al., 1997; Smalley et al., 1998). Consequently, human studies with reduction mammoplasty tissue grown in vitro tend to lose some of the key luminal subpopulations, leaving behind a largely proliferating mammary epithelial cell with basal-like features (Gordon et al., 2000). More recently, new lineage markers used for the separation of subpopulations have been discovered within the stem cell hierarchy. While CD24/CD29/CD61 have been instrumental in defining populations with repopulation capacity, metabolic activity, and milk-producing function, new markers have been shown to help further define the subpopulations, expanding our understanding of the layers involved in plasticity, self-renewal, and function (see Table 1.2).

The prepubertal and pubertal/gestational development phases are largely defined by their dependence on hormones. Unlike the prepubertal phase, hormone dependence occurs after puberty and results in ductal elongation and side branching, and in pregnancy, inducing alveologenesis and milk production. Mouse experiments involving the surgical ablation of hormone-secreting organs from mature female mice have resulted in lack of mammary gland development. The organs involved in mammary development, mainly the pituitary gland and the ovaries, are responsible for the secretion of the hormones GH, PRL, 17β-estradiol, and progesterone.

Subsequent studies have demonstrated that when the pituitary and/or ovaries were removed, treatment of mice with 17β-estradiol, progesterone, growth hormone, cortisol, and prolactin could recapitulate the development of the gland (Brisken and O'Malley, 2010). The steroid 17β-estradiol is one of the most potent and prevalent of the estrogens. Other members of the estrogen family include, estrone and estriol, are also present in the mammary gland. The effects of these ligands are mediated through two estrogen receptors, ESR1 and ESR2, and when bound to a ligand, ESR1 undergoes a conformational change, resulting in its interaction with other transcription factors such as AP-1, SP-1, STAT3, or NF-κB (Tanos et al., 2012). ESR1 also recruits coactivators and/or corepressors to help promote these transcriptional changes (Moggs and Orphanides, 2001). Due to the action of 17β-estradiol, this genomic mechanism of ESR1 signaling induces expression of amphiregulin (AREG) which is then acted on by ADAM17, a protease that cleaves AREG to its secreted form (Sternlicht and Sunnarborg, 2008). Via paracrine mechanisms, AREG subsequently binds to the epidermal growth factor receptor (EGFR) on fibroblasts, further inducing the expression of various growth factors, including several members of the fibroblast growth factor (FGF) family of ligands (Coleman-Krnacik and Rosen, 1994). Within the luminal cell, ESR1 signaling is also highly influenced by coregulators, which include coactivators and corepressors. This is also the case for progesterone receptor (PGR) signaling, which will be discussed next.

A fine balance of ESR1 coactivators and corepressors within the cell determines the fate of ESR1 signaling. The recruitment of either coactivators/corepressors to the receptor site is predicated on the mechanisms of ESR1 activation. Phosphorylation of ESR1 is necessary for particular interactions with coactivators, and the consequences can lead to the recruitment of other coactivators, thereby increasing the transcriptional activity of ESR1 (Dutertre and Smith,

2003; Leo and Chen, 2000). While ESR1 is necessary for mammary development in both mice and humans, ESR1 levels in luminal progenitor cells are lower in mice than in humans (Asselin-Labat et al., 2007; Asselin-Labat et al., 2010; Lim et al., 2009), although the reasons for this are not well understood. This may explain, in part, why there are no real ESR1-positive mammary models.

It is also difficult to study ESR1 signaling in mice, as both the stroma and epithelium express ESR1. This is in contrast to the human mammary gland, where ESR1 is uniquely expressed in the epithelium. As a result, studies surrounding the role of ESR1 signaling in development have been focused on human normal and breast cancer cell lines. Two of the critical mediators of ESR1 signaling are FOXA1 and GATA3. Both are considered pioneer factors: FOXA1 has been shown to mediate ESR1 binding to a subset of gene promoters whose genes are involved in proliferation, such as Cyclin D_1 (CCND1) (Eeckhoute et al., 2006, 2007). CCND1 is required for proliferation in ESR1/PGR-positive cells in response to progesterone stimulation during alveologenesis (Beleut et al., 2010). GATA3, like FOXA1, initiates the tethering to promoter regions adjacent to estrogen responsive elements (ERE) prior to ESR1 binding (Liu et al., 2011). In return, ESR1 binds to EREs in the GATA3 promoter, upregulating GATA3 expression (Grober et al., 2011; Jin et al., 2004). GATA3 is also involved in mouse mammary differentiation, as Gata3 deletion in alveolar cells during pregnancy leads to a loss in alveolar differentiation and failed lactogenesis (Asselin-Labat et al., 2007). ESR1-mediated gene repression is controlled through the interaction of ESR1 with NF-κB and C/EBPβ, resulting in the down regulation of the proinflammatory gene IL-6 (Ray et al., 1994; Stein and Yang, 1995). Treatment with 17β-estradiol in human mammary cells in vitro leads to upregulation in genes involved in proliferation (GREB1), secretory development (TFF1, TFF3), or hormonal

regulation (PIP) (Wilson et al., 2006). Interestingly, several of these genes are also overexpressed in human luminal A/B tumors, which highlights the remarkable similarities between normal and tumorigenic genomic profiles.

In addition to mammary development, estrogens act on the pituitary gland by stimulating prolactin synthesis and secretion. In turn, prolactin controls the synthesis of progesterone and induces ESR1 expression in various tissues (Bachelot and Binart, 2007; Frasor and Gibori, 2003). In order to better understand the direct versus indirect mechanisms behind hormonal regulation of the mammary gland, hormone receptor-deficient mouse strains have been developed. For the cognate hormone receptors ESR1, ESR2, and PGR, and prolactin receptor (PRLR), mice without any of these receptors are in fact viable but have severe reproductive abnormalities, including sterility (Bole-Feysot et al., 1998; Hamilton et al., 2014; Ismail et al., 2003). In the mouse mammary gland, ESR1 is expressed in both the epithelium and the stroma. Epithelial ESR1 expression is required for ductal elongation and for subsequent side branching in alveologenesis. Studies addressing the necessary role of ESR1 in the stroma were performed by grafting $ESR1^{-/-}$ fat pads along with wild-type ESR1 mammary epithelium into the muscle wall of mice. The presence of stromal ESR1 was not required for mammary gland development (Mallepell et al., 2006).

Similar to ESR1 signaling, PGR is also expressed in both the epithelia and the stromal tissue of the mammary gland (Tanos et al., 2012). There are two protein isoforms of PGR, PR-A and B, which are expressed as a result of transcription from two alternative promoters (Conneely et al., 2001). Functional characterization of isoforms A and B have shown that PR-B is required for mammary development while PR-A is necessary for reproductive responses to fertility (Conneely et al., 2003). The ligand for PGR, progesterone (P4), is produced by the ovaries

during the luteal phase of the reproductive cycle, whose role in the mammary gland is the promotion of ductal side branching and alveolar development (Graham and Clarke, 1997). PGR receptor signaling is thought to be mediated in the mammary gland via a series of signaling events. First, progesterone activates PGR-negative cells in the TEBs, undergoing proliferation to promote side branching and alveolar development, while the adjacent PGR-positive cells in the TEB remain quiescent. Downstream targets of PGR have been implicated as effector genes for proliferation and side branching morphogenesis (Obr and Edwards, 2012). These genes include Wnt4, amphiregulin, and RANKL, among others. During pregnancy, progesterone suppresses differentiation by inhibiting the expression of tight junction proteins until parturition in order to assume the proliferative role of alveolar expansion up until lactation. During lactation, progesterone levels will decline, allowing for milk production and secretion (Obr and Edwards, 2012). Currently, the molecular mechanisms regulating tight junctions between luminal cells, which progesterone mediates, are unknown.

The hormonal regulation of milk production falls under the regulation of two key lactogenic hormones: prolactin (PRL) and glucocorticoids. As mentioned earlier, PRL induction is regulated by ESR1. Prolactin receptor (PRLR) signaling activates STAT5A via phosphorylation by JAK-2. In addition to ESR1, glucocorticoids may also potentiate PRL signaling by recruiting glucocorticoid receptors (GR) near STAT5 binding sites. Other coactivators of milk protein production, such as C/EBPβ will form a complex with STAT5A and GR to help further recruit other proteins involved in chromatin remodeling, allowing for transcriptional activity at the β-casein (CSN2) promoter site (Doppler et al., 1990; Kazansky et al., 1999; Wyszomierski and Rosen, 2001). ESR1 and PGR-positive cells in the mammary gland tend to be quiescent, and a similar phenomenon is seen with STAT5A and PRLR-positive cells

(Brisken and Rajaram, 2006). In the presence of progesterone, a STAT5A-PGR complex can inhibit activation of downstream genes, importantly β-casein (Buser et al., 2013), highlighting the timely importance of progesterone stimulation during pregnancy and lactation.

1.2.2 Comparing Mammary Development to the Five Intrinsic Subtypes of Breast Cancer

Gene expression microarray analysis of human breast cancers has revealed five intrinsic molecular subtypes: luminal A, luminal B, basal-like, HER2-enriched, and claudin-low (Perou et al., 2000; Prat et al., 2010). Each of these subtypes has unique histological features and different responses to therapy and clinical outcomes. The term "intrinsic" subtype was used to characterize these tumors in part due to their developmental characteristics (Perou et al., 2000). Concerted efforts have been underway for several years now to understand where along the breast developmental hierarchy cellular transformation occurs. One of the more prevailing thoughts is that a normally quiescent stem cell undergoes a "genomic" hit, or series of hits, causing oncogenic transformation. These stem cells are found at different stages of development; however they may likely give rise to a less differentiated tumor type, such as a basal or claudin-low tumor.

Alternatively, an oncogenic hit can occur within a well-defined, mature luminal cell, resulting in a HER2-enriched or luminal tumor type (Visvader, 2009; Visvader and Stingl, 2014). As an attempt to exploit the relationship between mammary development and the intrinsic subtypes, Lim et al. (2009) sought to answer two key questions. First, they wanted to determine if they could isolate and characterize human mammary cells from reduction mammoplasties using two markers: EpCAM and CD49f. Each of the four populations—$EpCAM^-/CD49f^-$, $EpCAM^-/CD49f^+$, $EpCAM^+/CD49f^+$, and $EpCAM^+/CD49f^-$—represents the stromal, MaSC,

progenitor, and luminal fractions, respectively. The subpopulations were introduced into NOD-SCID mice to determine their mammary regenerative capacity.

The EpCAM$^-$/CD49f$^+$, MaSC population was determined to have such capacity (Lim et al., 2009). Furthermore, each of the sorted fractions were characterized by various basal and luminal markers. The EpCAM$^-$/CD49f$^+$ MaSC group exhibited expression of KRT14, p63, and Vimentin—all basal markers—while the progenitor group, EpCAM$^+$/CD49f$^+$, expressed KRTs 5/6, 8/18, 19, GATA3, and MUC1. As expected, the MaSC group lacked hormone receptor expression. The mature luminal group, EpCAM$^+$/CD49$^-$, stained positive for KRTs 8/18, 19, GATA3, and MUC1, although at higher frequencies than the progenitor. Hormone receptor detection of ESR1 and PGR was highest in the EpCAM$^+$/CD49f$^-$ group, indicating this is the mature luminal population.

In terms of mammary development, membrane markers EpCAM and CD49f also segregate the mouse mammary hierarchy of development similarly, which is important to know, as the genomic classification of human breast EpCAM and CD49f subpopulations discovered in the Lim et al. (2009) study will be used to extrapolate the differentiation status of various mouse models in Chapter 2 (Shehata et al., 2012). There are other aspects of development in both species that appear to be conserved. For example, hormone expression is lacking in the EpCAM$^-$/CD49f$^+$ MaSC subpopulation. The MaSC subpopulation expresses KRTs 5/6, p63, and EGFR while there is no expression of ERBB2/HER2, all of which are characteristics of the basal-like tumors (Bertucci et al., 2012; Kreike et al., 2007). However, the MaSC genomic signature was most enriched in both normal mammary and claudin-low tumors. This is likely due to the low proliferative profile and mesenchymal-associated genes found in both tissue types (Herschkowitz, 2007).

The progenitor subpopulation contains both basal and luminal features; yet this group was found to have the greatest enrichment in the basal-like tumors, including the expression of ESR1 in a small subset of cells (Asselin-Labat et al., 2010; Nielsen, 2004). This may explain why approximately 10%–15% of basal-like tumors are ESR1 positive (Nielsen, 2004). It has been established that BRCA1 mutant breast tumors in both mouse and human are basal-like, which makes for a good model in understanding the origin of basal-like tumors (Akslen et al., 2003; Liu, 2008; Turner et al., 2006). The study of the origin of BRCA1 tumors may be best studied in human subjects with germline BRCA1 mutations. Lim et al. (2009) compared the EpCAM and CD49f subpopulations in both BRCA1 wild-type and mutant normal breast samples from reduction mammoplasties. Curiously, the BRCA1-mutant normal progenitor fraction expanded significantly more than the matched BRCA1 wild-type group. Moreover, the growth properties of the mutant fraction were significantly greater than that of the wild-type group in vitro, confirming the consequences of BRCA1 loss in regulating cell cycle checkpoints (Lim et al., 2009).

Finally, the mature luminal subpopulation, $EpCAM^{+}/CD49f^{-}$, described in the Lim et al. (2009) study as ESR1, PGR, and KRTs 8/18 positive, are also found to be highly enriched in luminal A, luminal B, and HER2-enriched tumor types. To further characterize the functional aspects of this population, there was lack of growth in a 3-D Matrigel assay, a typical growth pattern seen in differentiated, quiescent cells, in the absence of lactogenic hormone stimulus (Gordon et al., 2000).

A more recent characterization of human breast cancer cell lines and normal mammary tissue was performed to identify their genomic correlation with one of the five intrinsic subtypes (Prat et al., 2013). Breast cancer cell lines have afforded the scientific community a valuable

resource for understanding the molecular mechanisms behind diseases and have given us an ability to effectively study the effectiveness and potency of drugs. While there are no luminal A models of breast cancer, MCF-7, a luminal B cell line, has allowed the elucidation of ESR1 signaling through study of the effects of tamoxifen, an ESR1 antagonist. A reason for the lack of luminal A cell lines is that they are slow-growing tumors and are extremely difficult to propagate in culture. Basal cell lines such as SUM149 and HCC1143 have demonstrated expression of dual populations of MaSC ($EpCAM^-$) and progenitor ($EpCAM^+$) subpopulations.

These results have highlighted others' work confirming that basal-like tumors may, in fact, exhibit both MaSC and progenitor phenotypes and growth characteristics (Liu, 2008; Prat et al., 2013). More importantly, these data have helped adopt the theory as to why basal tumors initially respond well to chemotherapy and then relapse, as such subpopulations are slower cycling and resistant to chemotherapy (Bertucci et al., 2012; Fillmore and Kuperwasser, 2008). The claudin-low tumors and normal mammary cell lines were classified in the Prat et al. (2013) study as representing the MaSC subpopulation, as these cell types are $EpCAM^-$, express mesenchymal genes, and are slower cycling than their progenitor/basal $EpCAM^+$ counterparts (Prat et al., 2013). These data and others confirm that the current repertoire of human breast normal and cancer cell lines recapitulate all of the hierarchical states seen in breast tumors, with the exception of luminal A tumors (Kenny et al., 2007; Neve et al., 2006; Prat et al., 2013; Prat et al., 2010). Figure 1.5 illustrates a summary of commonly used normal and cancerous cell lines and their respective locations along the developmental axis of differentiation.

The ability to transform breast cancer cell lines and tumor data into a format that represents their developmental status is crucial for interrogating the origin of the tumor. These data also represent the relative conservation of mammary development between mouse and

human, stressing the importance of reliably using mouse models in interpreting normal and oncogenic events in humans. The knowledge regarding where tumors are derived from yields valuable insight into how developmental cues can influence the characteristics of a tumor, including its response to therapy and metastatic potential. In addition, these studies have shown that the cell lines characterized closely resemble the tumors/normal tissue they are derived from, reinforcing their vital roles as reliable models for research. A critical issue that remains, however, is the lack of a cell line that represents the most commonly diagnosed breast cancer subtype, luminal A. Laboratories, including our own, are trying to utilize innovative ways to develop a working luminal A cell line for in vitro use. Some of these technical advances include the reprogramming of cells from one type to another, and such techniques have demonstrated success in a limited number of tissue types.

1.3 Reprogramming/Transdifferentiation: What Has History Taught Us?

Cellular reprogramming, by definition, is the conversion from a somatic cell, such as a fibroblast, to a pluripotent stem cell (iPSC) by a number of mechanisms addressed in this section. The world went wild in 1996 when a laboratory at the Roslin Institute teamed up with a small biotech firm to create the first cloned animal, Dolly the Sheep. Ian Wilmut and colleagues transferred the nuclei of cultured epithelial cells into enucleated oocytes to create Dolly, who, ironically, lived a short 6 years (Edwards, 2010). The key to mammalian cloning success, scientists discovered, was the presence of nuclear factors. Such nuclear factors present within the nucleus blocked the effective transfer of genetic information; hence unfertilized meiotic stage II nuclei, where the nuclear factors reside in the cytoplasm, were used (Eggan et al., 2007). The question still remains, however, whether it can be utilized in human cells, as the successful reprogramming of human somatic cells has only been performed up to the blastocyst stage.

The discovery and isolation of mouse embryonic stem cells (mESCs) in the early 1980s, followed by the establishment of human embryonic stem cells (hESCs) in the 1990s by James Thomson at the University of Wisconsin, were incredible breakthroughs not only for understanding the concept of pluripotency but also as a tool to manipulate cells into different lineages. Expanding on the nuclear transfer methods mentioned earlier, scientists discovered that transferring mESCs into a blastocyst, implanting it in vivo, and subsequently breeding can create mice entirely from the ESC DNA fingerprint (Graf, 2011). Basic biological tools used in the lab have been fine-tuned and designed to assist with reprogramming to pluripotency. The use of growth factors was found to be essential for many specialized cell types.

Cytokines, in particular LIF, were found to be essential for ESC growth and maintenance (Williams et al., 1988). RNA tumor viruses were later found to be an indispensable tool for the insertion of genes into a host genome at high efficiencies (Yu et al., 2007). Retroviral gene transduction was instrumental in the first positive iPS studies as a way to introduce and overexpress ectopic genes essential for reprogramming (Graf, 2011; Stadtfeld and Hochedlinger, 2010; Takahashi and Yamanaka, 2006). Finally, it is worth mentioning that the creation of transgenic knock-in/out mouse models utilized in performing lineage-tracing experiments. The creation of a mouse line with a reporter construct inserted into Fbx15, a locus specific for ESC, was important for the discovery of iPSC reprogramming. Takahashi and colleagues used this line to select for Fbx15-positive cells, thus enriching for the propagation of iPS colonies in vitro (Takahashi and Yamanaka, 2006).

Takahashi and Yamanaka (2006) discovered the replacement of nuclear transfer lies in the ectopic overexpression of 4 transcription factors (OCT4, KLF4, SOX2, MYC) in a somatic fibroblast cell line. As a matter of fact, conversion of a somatic cell to an iPS cell was completed

in vitro, although the conversion is not always a complete one and has extremely low efficiency (< 1% of cells). Confirmation of the function of candidate iPS colonies was determined in two ways. First, colonies were transplanted into NOD-SCID mice and gave rise to all three lineages. Secondly, transfer of GFP-labeled iPS cells into a blastocyst resulted in embryos positive for GFP. After the pups were born, GFP-labeled cells gave rise to all three lineages, confirming that the induction of the four transcription factors in a somatic cell can create cells with pluripotent capabilities (Takahashi and Yamanaka, 2006).

Meanwhile, at the University of Wisconsin, James Thomson and colleagues created an OCT4 knock-in hESC cell line to screen for combinations of genes that could reprogram the cell line. Candidate colonies with transcription factors OCT4, SOX2, NANOG, and LIN28 that grew in geneticin (a selection marker on OCT4 promoter plasmid) were transplanted into NOD-SCID mice to demonstrate teratoma formation with all three lineages. Yu et al. (2007) also discovered that the iPS cells were demethylated at the OCT4 locus compared to controls, an epigenetic feature of embryonic stem cells. Both labs utilized similar mechanisms to demonstrate the feasibility of creating a functional iPS cell. The subtle differences in transcription factor selection were largely due to the species differences (mouse vs. human).

While their studies set off a wave of new research directions, limitations to reprogramming still persist. The low efficiency of conversion is a consequence of a couple of key technical and biological issues. Cellular transduction of all four transcription factors in every cell can be challenging as well. Choosing the most optimal viral delivery system is extremely important. Retroviral transgenes can be silenced toward the end of reprogramming due to methyltransferase activation, resulting in an incomplete repertoire of reprogrammed cells (Sridharan et al., 2009; Takahashi and Yamanaka, 2006). Lentiviruses are less likely to be

silenced than retroviruses; however, their constitutive activation can result in difficulties inducing differentiation of the iPS cell (Brambrink et al., 2008). A way to circumvent this issue is the creation of an inducible system, where induction/silencing of transcription factors can be controlled by the use of doxycycline. Moreover, reprogramming has been accomplished more recently without the use of viral vectors. Both human and mouse iPSCs have been derived by the delivery of recombinant proteins and whole-cell extracts isolated from ESCs, although the efficiency of this process is extremely low (Cho et al., 2010; Zhou et al., 2009). In order to overcome reprogramming inefficiencies, recent efforts have included the use of chemical compounds as another nonintegrating approach to iPSC development. While these agents, including histone deacetylase inhibitor valproic acid, increased reprogramming efficiencies, they have not been successful at replacing reprogramming factors (Desponts and Ding, 2010). Unlike the single iPS reprogramming experiment that occurred in 2006, the discovery of converting one cell type into another occurred in increments that started from the directed differentiation of fibroblasts into muscle cells by Myod, followed by the demonstration that committed and fully differentiated cells can be switched within the hematopoietic system and then finding that cell types from different germ layers could be interconverted (Davis et al., 1987; Graf and Frampton, 1995; Huang et al., 2011). Many of these conversions were purposely sought out for therapeutic reasons: for example, b-islet cell production to alleviate hyperglycemia caused by insulin deficiency and cardiomyocytes from fibroblasts in heart injury (Ubil et al., 2014; Melton et al., 2008). These studies also demonstrate that the phenotype of the new transdifferentiated cells might not be identical to that of their "normal" counterparts. While this may be the case, their confirmed function in vivo raises hopes that specific transcription factor perturbations can be harnessed to generate therapeutically useful cells.

Some of the key transdifferentiation studies mentioned earlier relied on mouse embryonic fibroblasts (MEFs) as the host cell line. These cells are considered to be genomically stable due to nonconforming chromatin configurations and DNA modifications, among other attributes that can make working with MEFs both advantageous and troublesome (Vierbuchen and Wernig, 2012). In 2011, Huang and colleagues published data on the successful conversion of mouse fibroblasts into functional hepatocytes. One consideration when starting with a primary cell line is its proliferative limitations. Alternatively, the generation of immortalized mouse fibroblasts has become an invaluable tool in studying transdifferentiation, although it depends greatly on the method of immortalization, as some can create genetic instabilities and become tumorigenic (Muhammad and Mohammad, 2013; Oh et al., 2007). MEFs derived from *cdkn2a*-null mice tend to be used due to their genetic stability despite continuous cycles of proliferation (Ieda et al., 2010). Both the primary tail-tipped fibroblasts (TTF) and the *cdkn2a*$^{-/-}$ MEFs used in the Huang study were transdifferentiated with three defined transcription factors important for liver function (FOXA2, HNF1α, and GATA4). Conversion into induced hepatocytes (iHEP) took 2–3 weeks, and while there was no selection based on a tagged reporter as seen in the iPS studies, "candidate" colonies were selected based on morphology and expression of key epithelial markers. Production of key liver enzymes was confirmed, and ultimately the test for true functionality came when the iHEP cells were injected into *Fah*$^{-/-}$ mice (a mouse strain with severe liver disease that ultimately causes liver failure and death). The iHEP cells were transplanted into this strain, resulting in over 80% survival rate after 60 days, while the control group were all dead within that time frame (Huang et al., 2011). The Huang study, as well as others, exhibited a decrease in exogenous expression of defined TFs, suggesting that despite the conversion, the unremitting overexpression was not necessary. These induced cells may not

express and/or produce proteins at the same level as their tissue counterparts, despite the fact they function similarly in vivo (Vierbuchen et al., 2010; Virmani et al., 2001). Finally, the conversion of *cdkn2a*$^{-/-}$ MEFs does not seem to promote tumor growth, suggesting that silencing a key component of the cellular senescence pathway does not increase risk of transformation. Transdifferentiation studies have been successfully executed across several tissue types. Similar to iPS, conversion events are rare and colony selection is critical, as conversion includes several intermediate stages comprising heterogeneous cells. Other considerations going forward include understanding the role of chromatin modifications and how they affect the rate or completion of cellular conversion. Besides the established technical limitations, it is crucial to understand why conversion so rare. Is there an underlying biological process that can explain this? Moreover, the creation of an induced cell line representing reproductive tissue and utilizing viral transduction methods has yet to have been achieved.

One approach in developing a functional prostate gland was tested several years ago with tissue recombination. Mesenchymal cues are necessary for the development of reproductive organs, including the prostate, uterus, and breast. Taylor et al. (2006) sought to create a functional prostate by combining spermatogonial stem cells (SSC), which are pluripotent, like human ESCs, with cells of mesenchymal origin—urogenital sinus mesenchyme (USM). Cells from the USM give rise to either prostate cells in males or uterine tissue in females, and early studies on the importance of mesenchyme dictating epithelial fate have been established (Cairns and Saunders, 1954). Ultimately, the process of tissue recombination resulted in the development of prostate tissue that expressed prostate-specific antigen (PSA), and expressed basal and luminal markers associated with the cellular organization of the prostate gland. The process of mature glandular development took 8–12 weeks, highlighting the arduous amount of time needed for

development, a stark contrast to the 3–4 week time frame for iPS conversion (Takahashi and Yamanaka, 2006; Taylor et al., 2006).

Mammary tissue, similar to prostate tissue, develops slowly relative to other organs. It is reliant on hormonal signaling cues from distant organs, as well as local mesenchyme (Cooke et al., 2010; Tanos et al., 2012). The origin of the mammary epithelial cell is poorly understood. The complicating factors include the time for reaching maturity, which in humans is not until puberty. It is therefore possible that well-controlled epigenetic mechanisms maintain the quiescence of the gland until puberty/pregnancy is reached. The ability of such mechanisms to successfully keep cells in check for years in humans may explain the difficulties in creating a mature luminal breast cell line. The next few chapters will discuss attempts to create a luminal breast line, as well as the study of a key pioneer factor that may be crucial in reversing the epigenetic regulation of breast stem cells, GATA3.

1.4 Summary

Mammary development requires stepwise changes in tissue interaction, hormones/growth factors, and tightly regulated transcription. Breast cancers appear to reflect the molecular aspects of the stage at breast differentiation from which they arise. Luminal A tumors appear to be the most differentiated and slowest growing; currently, there are no models for luminal A tumors. Finally, there is literature describing the utility of transdifferentiation from one somatic cell type to another, which has been largely successful due to the sentinel work in the development of iPS cells.

Table 1.1. Current Mammary Mouse Models and their Phenotypes

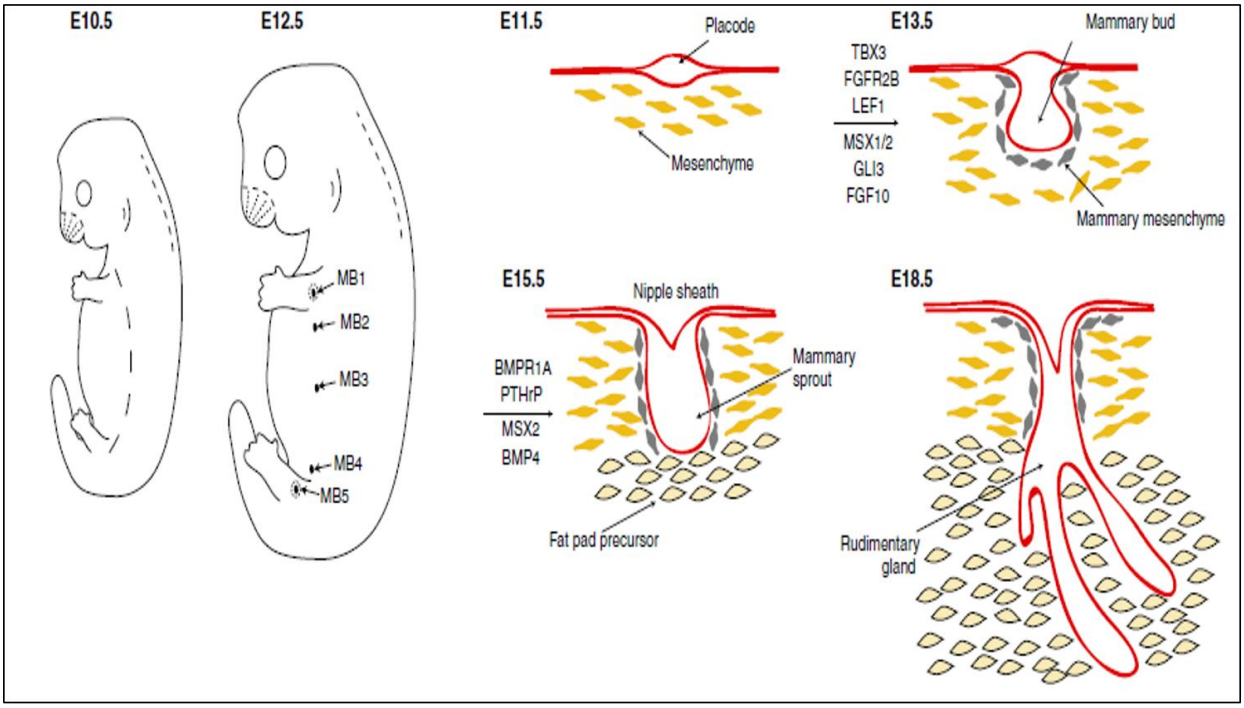
Gene Model	Mammary Defects	Reproductive System
Hormonal Regulation		
ERα	No ductal outgrowth	Infertile
PR	Impaired alveolar development	Infertile
SRC-1, SRC-3	Reduced ductal branching	Fertile
Lactogens		
Prolactin	Reduced ductal branching	Infertile
PRLR	Reduced ductal branching, alveolar development	N/F corpus luteum
Embryonic Specific		
PTHrP	Lack of embryonic mesenchyme and primary sprout growth	NA
Pregnancy/Involution Specific		
Stat3	Delayed involution	NA
Stat5a	Impaired differentiation, milk production	None
Stat5b	No defects	Nonfunctional corpus luteum
Bcl-x	No defects	Loss of primordial germ cells
RANKL or RANK	Impaired alveolar development	NA
Proliferation Related		
Cyclin D1	Impaired alveolar development	NA
p27	Impaired alveolar development	Infertile
Id2	Impaired alveolar development	NA
Transcription Factors		
C/EBPβ	Impaired ductal and alveolar development	Lack of corpus luteum
LEF-1	Arrest at embryonic bud stage	NA
GATA3	Defects in ductal elongation and branching	NA
MSX1/MSX2	Arrest at embryonic bud stage	NA

Note. NA = not applicable, or not described.

Table 1.2. Mammary Specific Lineage Markers Commonly Used in Mice

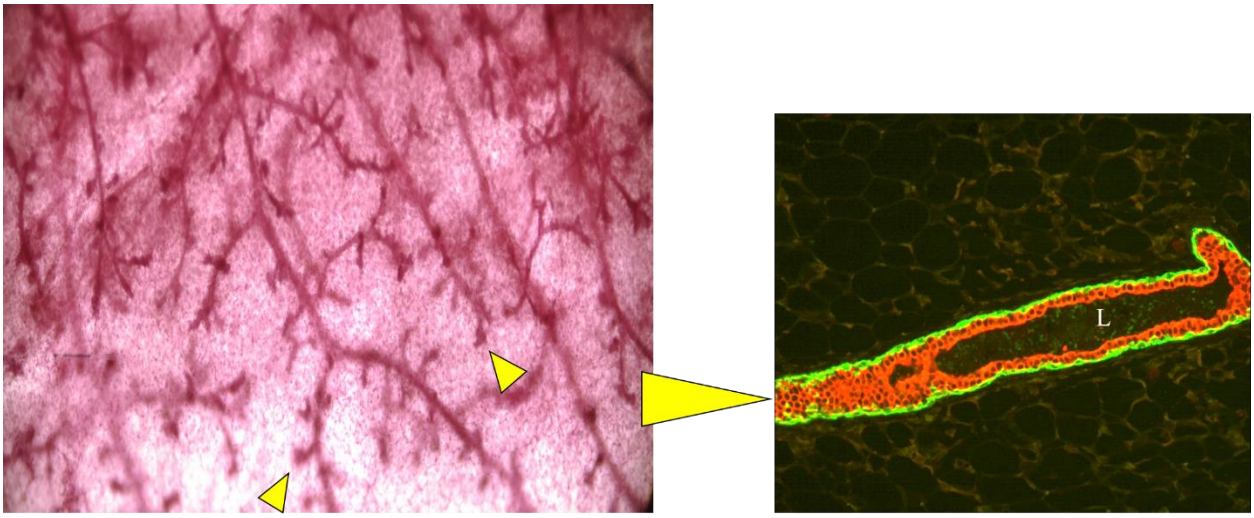
	Fibroblasts	**MaSC**	**Basal Progenitor**	**Luminal Progenitor**	**Luminal**
EpCAM	-	-	+	++	+++
CD49f	-	+++	++	++	-/+
ALDH-1	++	++	-	-/+	-
SCA-1	++	-	-	-/+	+
c-KIT	++	-	+	++	-
CD14	-	-	-/+	++	-
CD49b	++	-	?	-/+	-

Figure 1.1. Embryonic Mammary Gland Development



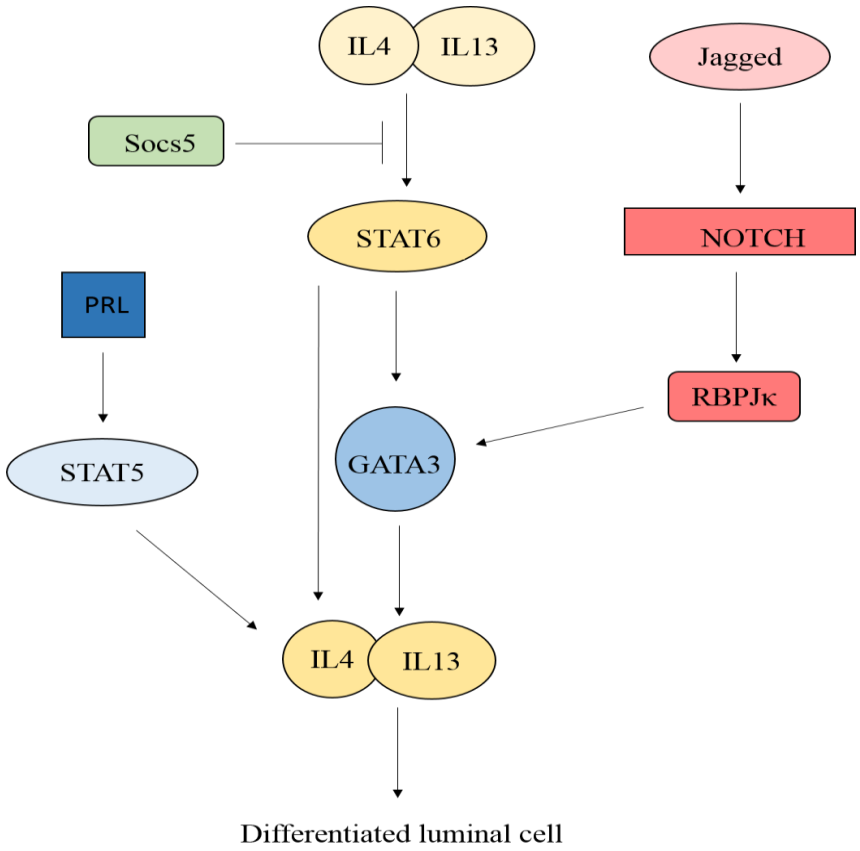
Adapted from Christine J. Watson and Walid T. Khaled, "Mammary Development in the Embryo and Adult: A Journey of Morphogenesis and Commitment," 2008, *Development* 135(6): 995-1003. Reproduced with permission.

Figure 1.2. Mammary Ductal Network from a 20 Week Old Mouse



Note. Close-up of terminal end bud (TEB) with lumen marked with L. Luminal cells are body cells, and stained red, while basal cells (cap cells) are stained green. The surrounding stroma is stained brown and consists of adipocytes. 40x magnification.

Figure 1.3. The IL4/13-STAT6-GATA-3 Pathway in Mammary Gland Development



Adapted from Christine J. Watson and Walid T. Khaled, "Mammary Development in the Embryo and Adult: A Journey of Morphogenesis and Commitment," 2008, *Development* 135(6): 995-1003.

Figure 1.4. Differentiation Hierarchy for Mouse Mammary Development

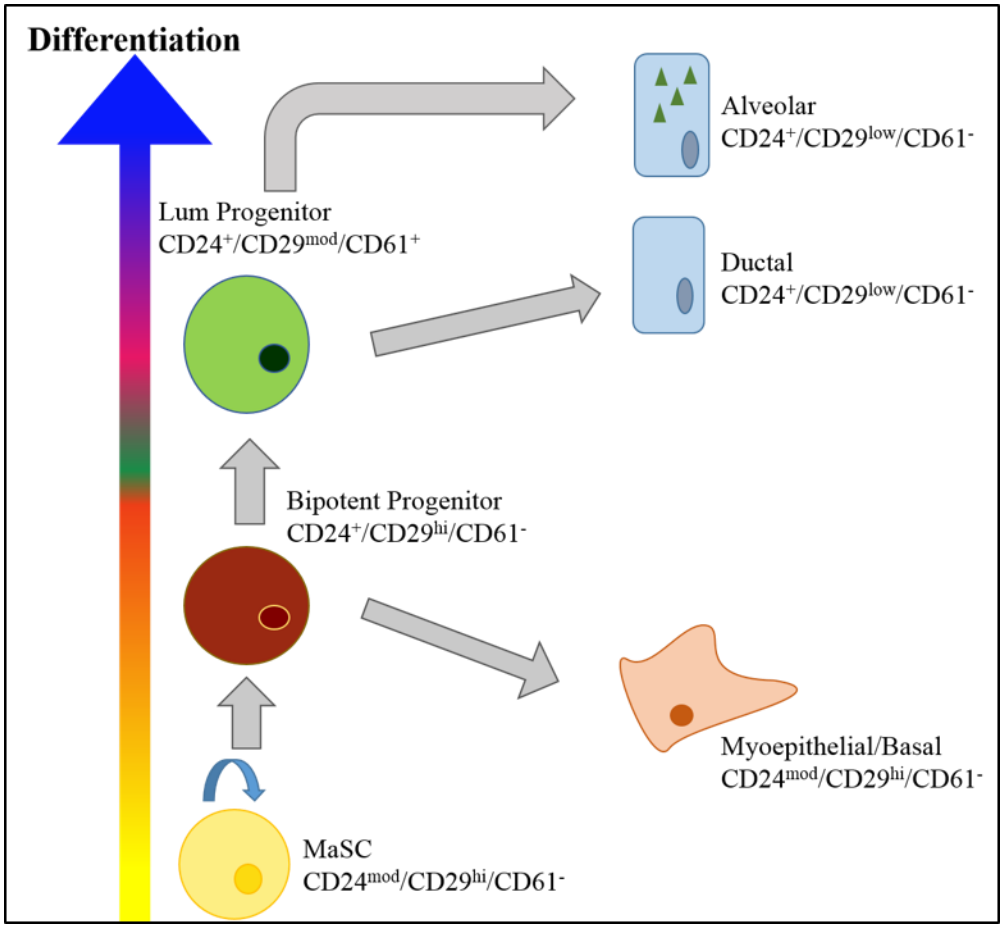
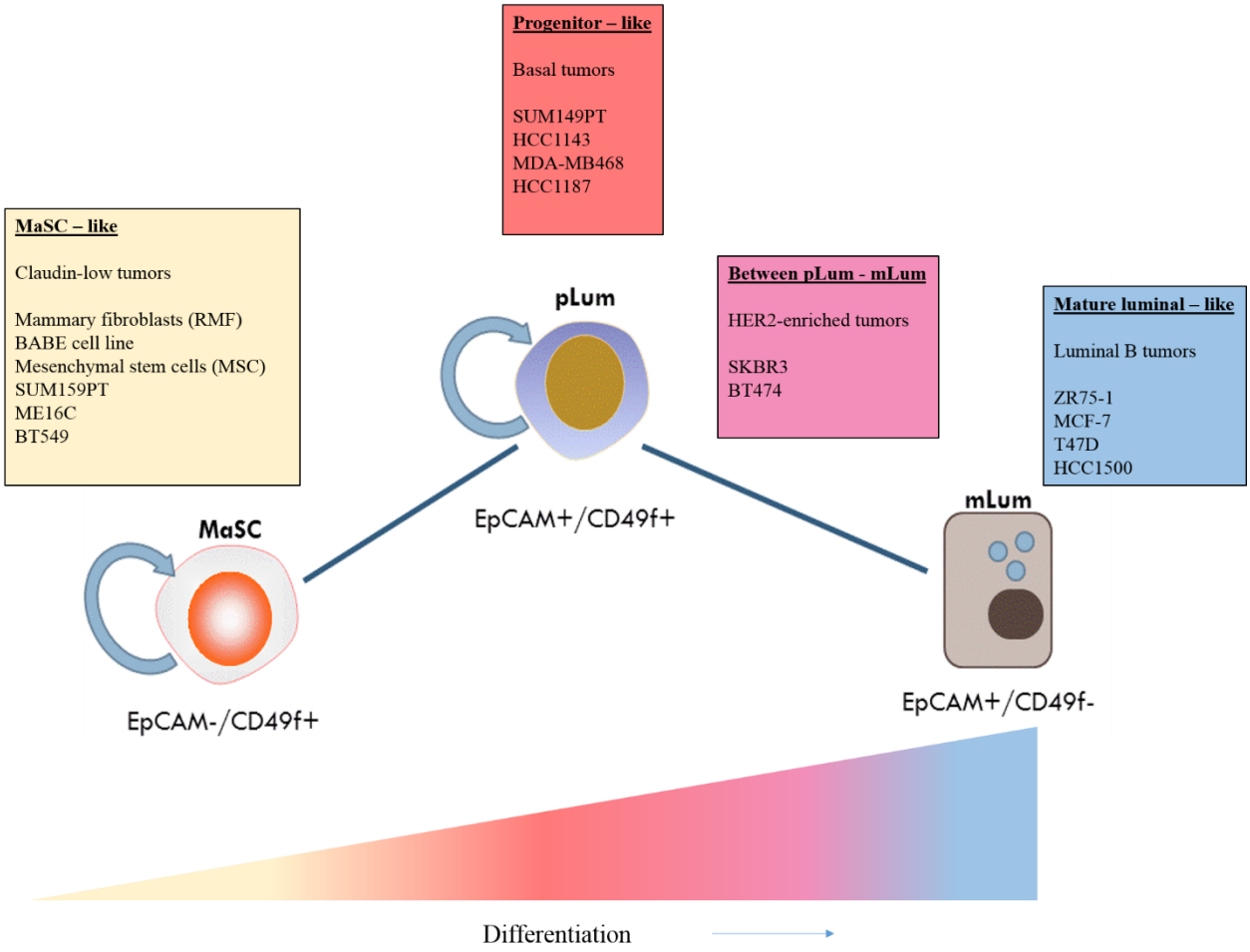


Figure 1.5. Human Breast Cancer/Normal Cell Line Alignment along a Differentiation Axis



CHAPTER 2: SPECIFIC COMBINATIONS OF TRANSCRIPTION FACTORS INDUCE TRANSDIFFERENTIATION TO A MAMMARY LUMINAL EPITHELIAL CELL PHENOTYPE[1]

Introduction

Luminal A tumors, one of 5 intrinsic subtypes originally defined by Perou et al., account for roughly 40% of all breast tumors diagnosed (Prat et al., 2010). Patients with luminal A tumors have a favorable prognosis compared to the other intrinsic subtypes, with a 10-year relapse free survival rate of 70% (Koboldt et al., 2012). However, the 30% relapse rates indicate that additional knowledge and better treatment are still needed. Typically, treatment consists primarily of surgical resection and estrogen-modulating therapy. Radiation and chemotherapy are also considered depending on location, lymph node involvement, tumor size, and/or OncoTypeDX intermediate/high recurrence score. While these patients experience superior overall and disease-free survival compared to patients with luminal B, basal, HER2 positive, claudin-low tumors, they do have a risk for relapse. Studies that directly evaluated the response to neo-adjuvant chemotherapy across the subtypes have shown that luminal A tumors achieved a pathological complete response (pCR) of 7%, versus a pCR of 43%, 36%, and 17% for basal, HER2-enriched, and luminal B tumors, respectively (Keller et al., 2010). These and other studies have highlighted the significant differences in chemotherapeutic sensitivities between luminal (predominantly ER+) and basal (predominantly ER-) tumors. There have been several attempts

[1] Francesca G. Bargiacchi, Joel Parker, Charles M. Perou, and H. Shelton Earp, "Specific Combinations of Transcription Factors Induce Transdifferentiation to a Mammary Luminal Epithelial Cell Phenotype," rest of journal info.

to explain this difference. Luminal A tumors are clinically classified as ER+, HER2-, and ki67-low, suggesting that slow cycling may cause these tumors to be relatively insensitive to chemotherapy (Inic et al., 2014). Because these patient tumors are slowly proliferating, in vitro propagation has been largely unsuccessful. In addition, there are no luminal A mouse models due to a lack of hormone dependency in murine mammary development.

Unlike the pivotal iPS reprogramming experiments reported 8 years ago, the discovery of approaches to directly convert one differentiated cell type into another has occurred incrementally, starting with the directed differentiation of fibroblasts into muscle cells by Myod (Davis et al., 1987). Later came the demonstration that committed and fully differentiated cells can be switched within the hematopoietic system and more recently the finding that cell types from different germ layers can be interconverted. For example, laboratories have recently published the successful transdifferentiation of fibroblasts into functional neurons and liver cells (Vierbuchen et al., 2010; Huang et al., 2011). While these studies showed that the phenotype of the new transdifferentiated cells might not be entirely identical to that of their "normal" counterparts, the fact that they acquire critical cell-type specific functions in vivo raises hopes that defined transcription factor perturbations can be harnessed to generate useful cells for both research and clinical purposes.

Currently, the understanding of genetic reprogramming in breast cancer has largely been limited to two main areas. Studies have demonstrated that pre-neoplastic/breast cancer cell lines can be reprogrammed in a paracrine fashion via the stromal environment (Bussard et al., 2010; Verbeke et al., 2014). Secondly, breast cancer cells can be partially reprogrammed to a more "plastic" state via direct introduction of defined transcription factors. A key study introduced the 4 defined transcription factors identified in Yamanaka's pivotal iPS work—Oct4, Sox2, Klf4,

and c-Myc (OSKM)—into MCF-7 cells, a human luminal breast cancer cell line (Corominas-Faja et al., 2013). While the MCF-7 cells did not adopt an iPS functional phenotype, they exhibited some stem cell characteristics, such as ES organoid formation in Matrigel and expression of stem cell markers ALDH and CD44 (Corominas-Faja et al., 2013). Similar studies have introduced TGF-β and/or downstream SMAD family TFs to induce cancer "stemness" from normal mammary epithelia (Piek et al., 1999). However, there are few, if any, studies attempting to transdifferentiate breast epithelial cells either from more stem-like cell types or from cells representing other lineages. It is likely that fully differentiated breast cells may represent a quiescent state or even be committed to senescence or cell death under some circumstances. This could create a serious roadblock to creating model systems. Other considerations for transdifferentiation strategies include host cell and/or genes used for direct reprogramming. Yamanaka's iPS experiments were successfully executed using somatic fibroblasts (Takahashi and Yamanaka, 2006). However, the transcription factors used (OSKM) are also involved in epigenetic events that influence chromatin modifications and alterations in the methylation of DNA, which contribute to the formation of iPS cells (Patel and Yang, 2010). Hence, methylation is a concern when starting with a lineage-committed cell line, and the selection of transcription factors with chromatin modification abilities and/or methyltransferase inhibitors may represent strategies to circumvent this.

In the present study, the selection of transcription factors for breast luminal cell reprogramming was predicated on their enrichment in luminal A human tumors as compared to all other intrinsic subtypes. For this analysis, we used the preexisting UNC337 data set, which has been extensively studied and subtyped. We chose the following nine candidate transcription

factors with known relevance to breast biology: AR, DACH1, ESR1, FOXA1, GATA3, LEF1, MYB, PGR, and XBP1.

Estrogen receptor alpha (ESR1) and progesterone receptor (PGR) are used clinically to define hormone-positive breast cancer (Parker et al., 2009). FOXA1, a pioneering factor, has been shown to been an indispensable partner for ESR1 and PGR signaling (Hurtado et al., 2011). FOXA1 can effectively bind chromatin domains during development and enable gene activity. The FOXA1-DNA binding domain structurally resembles linker histones and can bind nucleosomes, thus influencing the accessibility of certain transcription factors to their binding sites on particular promoters (Gong et al., 2014). Utilizing two different cell types, first BABES, an *hTERT*-immortalized human mammary epithelial cell line, and later mouse embryonic fibroblasts (MEFS), immortalized via the knockout of *cdkn2a,* we set out to determine if the creation of a luminal-like breast cell line would be feasible in either cell line. These two lines represent different lineages and begin with different methylation patterns (Bloushtain-Qimron et al., 2008; Doi et al., 2009). Experiments with the human BABE cell line led to a hypo proliferating and senescent population. Subsequently, we discovered that 3 of the 9 genes tested—ESR1, FOXA1, and PGR—can transdifferentiate MEFS into functional epithelial cells with breast luminal and basal characteristics. These findings shed light on the necessary molecular mechanisms for normal breast development and those that perhaps underlie the various intrinsic subtypes of breast cancers.

2.1 Materials and Methods

2.1.1 Plasmids

All gene ORFs were purchased in a pDONR223 entry vector (Harvard/Dana Farber Human ORFeome 5.1). Lentiviral destination vectors pLenti-PGK-hygro-DEST, pLenti-PGK-

blast-DEST, pLenti-PGK-neo-DEST, and pLenti-PGK-puro-DEST were used as the mammalian expression vectors (http://www.addgene.org). pLENTI-PGK-GFP-blast and the pLENTI-PGK-DEST backbones were used as positive/negative transduction controls. Final lentiviral expression plasmids were created using Invitrogen's recommended protocol for the Gateway system. Plasmids were then sequenced to verify ORF insertion and sequence.

2.1.2 Virus Production

293FT cells (Invitrogen, Life Technologies) were cultured in high-glucose DMEM (Gibco, Life Technologies) with 10% fetal bovine serum (FBS), 500 μg/ml Geneticin®, 1 mM MEM Sodium Pyruvate, 0.1 mM MEM Non-Essential Amino Acids (NEAA), 6 mM L-glutamine, and 1% Pen-Strep up until the day before transfection. Then they were switched to the same media without Geneticin. On the day of transfection, in one tube, 3ug of either one of four pLENTI-PGK backbone plasmids was added to 9ug (9ul) Virapower (Invitrogen, Life Technologies) in 1.5 mls of Optimem media (Gibco, Life Technologies). This was performed for each of the 9 genes, the GFP, and the negative control plasmids (backbones only). In a separate tube, 36ul of Lipofectamine2000 (Invitrogen, Life Technologies) was added to 1.5 mls Optimem media. Both tubes sat for 5 min at room temp and then were mixed together and incubated at room temp for an additional 20 min. Each 10 cm dish of 293FT cells received 5 mls of DMEM growth media plus 3 mls of the plasmid-lipofectamine2000 mixture. Cells were incubated at 37 $^{\circ}$C for 24 hours. Then the media were switched out for 10 mls fresh DMEM growth media for another 24 hours. The media were then collected, passed through a .2 um filter, and stored at –80 °C.

2.1.3 Cell Culture

The BABE cell line (*hTERT*-immortalized human mammary epithelial cells from the Counter lab, Durham, NC) was maintained at 37 °C in HuMEC medium supplemented with bovine pituitary extract, supplements, and 1% Pen-Strep (Troester, 2004). Embryonic fibroblasts from *cdkn2a*$^{-/-}$ C57BL/*6* mice were maintained at 37 °C in RPMI medium (Gibco, Life Technologies) with 10% FBS and 1% Pen-Strep (kindly donated to us by the Sharpless Lab, UNC-Chapel Hill). For the 4-day proliferation studies, cells were treated with either a range of tamoxifen (Sigma) or 17β-estradiol (Sigma). For the prolactin (Sigma) and combination 17β-estradiol and progesterone 4-day treatments, 100ng/ml, 10ng/ml, and 100nM of each were used, respectively.

2.1.4 Generation of Stable Cell Lines

BABE cells were plated into 6-well plates at a seeding density of 100,000 per well. Lentiviral infections were performed with 500 ul single virus plus 8 ug/ml polybrene in 2 mls HuMEC media with supplements (no antibiotics), as described above. For the combination viruses (1 ml of pooled virus) was added to 1 ml HuMEC media for a final volume of 2 mls. Cells were incubated at 37 °C for 24 hours. The media were changed to HuMEC with supplements for another 24 hours. Finally, selection media were added (HuMEC plus supplements) with one of the following antibiotics, depending on the pLenti-PGK- backbones. For puromycin, 1 ug/ml, 400 ug/ml Geneticin®, 200 ug/ml hygromycin, and 2 ug/ml blasticidin were in media for 2 weeks. *Cdkn2a*$^{-/-}$ MEFS were plated in 6-well plates at a seeding density of 100,000 per well. Combination viruses totaled 1 ml and were added to 1 ml RPMI plus 10% FBS and 8 ug/ml polybrene. For GFP and negative control viruses, 500 ul were used. Cells were incubated at 37 °C for 24 hours. The media were changed to RPMI plus 10% FBS for 48 hours.

Cells were then serum starved and placed in HuMEC media plus supplements. Both cell lines were tested and confirmed for gene expression by quantitative PCR over several time points (data not shown).

2.1.5 RNA Isolation and cDNA Synthesis

Cells were grown to about 80% confluency, trypsinized, and washed twice with PBS (Gibco, Life Technologies) before isolating the total RNA with the Qiagen RNeasy kit (Qiagen). The RNA was quantified using the Nanodrop (ThermoFisher), and 1ug of RNA was made into cDNA with the iScript kit (BioRad).

2.1.6 Quantitative PCR

For each of the following human gene targets (AR, DACH1, ESR1, FOXA1, GATA3, LEF1, MYB, PR, and XBP1), forward and reverse primer mixes were purchased from Origene. The primer sequences for murine genes KRT14, CDH1, PRLR, Vimentin, COL1A2 are listed here in order: **KRT14-F**: GAAGAACCGCAAGGATGCTGAG, **KRT14-R**:TGCAGCTCGATCTCCAGGTTCT. **CDH1-F**: GCTGTTGTGCTCAAGCCTTCAC, **CDH1-R**: CGGAAAGTGGAATCCTTGCAG. **PRLR-F**: GCATGATGACCTGCATCTTTCC, **PRLR-R**: CAAGGCACTCAGCAGTTCTTCT. **VIM-F**: AGCAGTGAGGTCAGGCTTGGAA, **VIM-R**: AGCAGTGAGGTCAGGCTTGGAA. **COL1A2-F**: TTCTGTGGGTCCTGCTGGGAAA, **COL1A2-R**: TTGTCACCTCGGATGCCTTGAG. TATA Box Binding Protein (TBP) was used as the endogenous control. **TBP-F**: CTACCGTGAATCTTGGCTGTAAAC, **TBP-R**: AATCAACGCAGTTGTCCGTGGC. Taqman SYBR Green Master Mix (ABI, LifeTechnologies), 20ng of cDNA, and a final primer concentration of 200nM were used and run on the ABI 7900 (ABI, Life Technologies). Relative concentrations were first normalized to

TBP. Then relative fold changes were calculated using either parental BABE or MEF as the control. Samples were run in triplicate.

2.1.7 Western Blotting

Cells were lysed in RIPA buffer containing 50 mM TRIS HCL, pH 8, 150 nM NaCl, 1 mM EDTA, 1% NP-40, 0.25% deoxycholate, and 30ug of protein were separated on SDS-PAGE. Primary antibodies were used at a dilution of 1:1000, and secondary antibodies were used for two-color fluorescent detection at a dilution of 1:5000 for rabbit (680) and 1:10,000 for mouse (800). Analysis was conducted with the LI-COR Odyssey system (LI-COR Biosciences).

2.1.8 Proliferation Assays/IC_{50} Curve Analysis

BABE cells and their respective single transductants were seeded at a density of 5×10^3 per well in a 96-well plate for 4 d. MEF and M-EFP cells were seeded at a density of 3×10^3 per well in a 96-well plate for 4 d. The MEF experiments included a treatment of either β-estradiol 10 uM–1 pM) or tamoxifen (10 uM–1 nM) in a dose response fashion. Next, 20 ul of CellTiter96® Aqueous One Solution cell proliferation assay solution (Promega) was added to each well and incubated at 37 ˚C for 2 hours. Then its absorbance was measured at 490 nm on a spectrophotometer (Molecular Probes). Four to 6 replicates for each cell line/condition were measured. This was repeated twice. We calculated the IC50 by calculating the nonlinear regression, sigmoidal dose-response, 3-parameter curve fit in GraphPad (GraphPad Software, Inc.). At least 5 replicates were used for each dose. Each experiment was repeated 2–3 times. The data shown include mean and SD values.

2.1.9 Flow Cytometry

Cells were grown to 80% confluency, trypsinized, counted, washed with HBSS (Gibco, Life Technologies) plus 2% FBS (HF), blocked in goat serum for 10 min at 4 °C, and stained for

30 min at 4 °C with the following antibodies at a 1:100 dilution: CD61-FITC (Biolegend), CD29-APC (Biolegend), CD24-PE (BD Biosciences), and EdU-FITC (Invitrogen, Life Technologies). Finally, cells were fixed in 2% PFA for 30 min, washed 2X with HF, and analyzed using Beckman-Coulter Cyan (Dako). The data were analyzed and compensated using FlowJo X (FlowJo, LLC). The FACS experiments were repeated twice.

2.1.10 Immunofluorescence with and without Matrigel

MEFS and M-EFP cells were plated on collagen-coated 8-well chamber slides at a density of 2 x 10^4 per well for 4 days at 37 °C. For cells on collagen-coated glass chamber slides, cells were fixed in 2% PFA, permeabilized with .5% triton X-100, and washed with PBST before blocking with 10% donkey serum in PBST. The secondary block used was the primary plus 13 ug/ml mouse $F(ab)^2$ (Jackson labs) in PBST. Primary antibodies were used at a 1/100 dilution and included CDH1 (Santa Cruz), Vimentin (Novus), PRLR (Santa Cruz), CSN2 (Santa Cruz), KRT18 (Abcam), KRT5 (BD Biosciences), and KI67 (Cell Signaling Technologies). Secondary antibodies used at a 1/200 dilution were Alexa dyes 488/594/647 (Molecular Probes, Life Technologies). DAPI stain was also purchased from Life Technologies and used at a concentration of 1 ug/ml. Cells were mounted with Prolong Anti-fade (Molecular Probes, Life Technologies). For Matrigel IF, 100 ul of low-growth Matrigel (BD Biosciences) was carefully placed in each well of an 8-well chamber slide and dried at 37 °C for 20 min. Next, 5 x 10^3 cells were plated for a total of 4 days in HuMEC media plus 2% low-growth Matrigel with the following supplements: 5% FBS, .5 mg/ml hydrocortisone, 10 ug/ml insulin, 100 ng/ml cholera toxin, 10 ng/ml prolactin, and 1% Pen-Strep. Cells in matrigel were processed as previously published (Debnath et al., 2003a), and antibodies, DAPI, and Prolong Anti-fade were used in the same manner as cells on glass slides (Debnath et al., 2003b). Images were taken on a Zeiss710

confocal microscope. Images were analyzed and median filtered on all samples to reduce minor background using ImageJ (freeware, http://imagej.nih.gov/ij/)

2.1.11 Gene Expression Microarrays

The human tumor and cell line microarray data set came from Prat el al. (UNC337, GSE18229) (Prat et al., 2010). The murine tumor data set is a subset of 385 gene expression microarrays from Pfefferle et al. (2013). Total RNA was purified from the BABE and MEF experimental samples using Qiagen's RNeasy minikit following the manufacturer's protocols. The RNA quantity was assessed using the Nanodrop spectrophotometer (ThermoFisher). Total RNA (1 ug starting input) was reverse transcribed and labeled with cyanine-5 (Cy5) dye for experimental samples and cyanine-3 (Cy3) dye for either human (BABES) or mouse (MEFS) reference samples using the Agilent Low RNA Input Fluorescent Linear Amplification Kit (Agilent). Equal quantities of either human or mouse reference RNA and experimental RNA were cohybridized overnight to either human 4 x 44k or mouse 4 x 180k Agilent microarrays, washed and scanned. Microarray intensity values were extracted as $\log_2$ Cy5/Cy3 ratios. Probes were filtered with Lowess normalized intensity values greater than 10 in both channels and averaged by mapping probes to the same gene (Entrez ID). Finally, only probes with data in more than 70% of the arrays were used in the analysis (Herschkowitz et al., 2007; Prat et al., 2010).

2.1.12 Informatics Analysis

GSA was performed using the R-script Arraytools (https://cran.r-project.org/). For the analysis, two biological and technical replicates were used for both the MEFS–ctrl and M–EFP microarray data. The set of transcriptional gene pathways used to define the enriched data set in M–EFP relative to MEFS–ctrl originated from the Molecular Signature Database (MSigDB),

which is a collection of gene sets maintained by the GSEA team (http://www.broadinstitute.org). The particular gene set we used to enrich for genes in the M–EFP cells was the C3 motif data set, transcription factor targets (TFT). The TFT is comprised of transcription factors and their downstream targets, as described previously (Xie et al., 2005). Differentially regulated TFTs (4) between MEFs and M–EFP were clustered using Treeview software.

For the differentiation score algorithm (d-score), human and mouse data sets were analyzed separately. Samples were scaled to mean zero and variance of 1. Features were then collapsed to the mean of each gene identifiers. Lim et al. (2009) generated gene expression profiles of three different epithelial cell-enriched subpopulations in breast: mammary stem cells (MaSC), luminal progenitors (pL), and mature luminal cells (mL). We used these as references to create a differentiation predictor for each sample as a measure of any sample's position along a MaSC pL mL axis as defined by gene expression. The axis was expressed as 2 vectors, MaSC pL and pL mL, where pL is transformed for 0, and each vector length is –1, 1, respectively. Data were then graphed either combining both axis or by values on each axis. ANOVA and unpaired *t* tests were used to determine significant differences between groups/samples.

2.2 Results

2.2.1 Transcription Factors Enriched in Luminal Tumor/Cell Lines

Nine transcription factors (see Table 2.1) were chosen based on several factors, including a literature review, previously published data from our lab, and SAM analysis (Significance Analysis Microarrays, Stanford University) performed on the UNC337 tumor data set, which comprised 320 breast tumors and 17 normal breast samples (GSE18229) (Prat et al., 2010). Tumors in the UNC337 tumor data set were first classified by their PAM50 status, as defined by Parker et al. (2009). SAM analysis was performed comparing the differential gene expression

patterns of luminal A primary tumors compared to all other tumor types (luminal B, HER2-enriched, basal, claudin low) using a false discovery rate (FDR) of < 1%. To identify genes found enriched in the luminal A primary tumors that are also expressed in luminal B cell lines, we performed SAM analysis at a FDR of < 1% on a breast cancer cell line data set published previously (Neve et al., 2006). Selecting the genes in common with both luminal A tumors and luminal B cell lines, we then filtered the gene set based on gene function. We devised a list of 9 transcription factors, which include (AR, DACH1, ESR1, FOXA1, GATA3, LEF1, MYB, PGR, and XBP1). The selected transcription factors have also been shown to have some role in luminal breast cellular pathophysiology (Perou et al., 2000; Prat et al., 2010, 2013).

2.2.2 Stable Cell Lines Expressing Candidate Transcription Factors

In order to determine the optimal combination of genes for luminal cell reprogramming, we transduced the BABE cell line with the individual TF genes and assessed their differentiation potential via a differentiation predictor model using data from Lim et al. (2009). The model assumes that mammary differentiation starts from a mammary stem cell (MaSC) and develops in progression to a progenitor luminal phenotype (pL), with terminal differentiation as a mature luminal cell (mL). In this model, higher scores represent a higher degree of differentiation along either of the two axes (see Figure 2.1A). We applied this model to both the tumor and the cancer cell line data sets as a comparison of where breast cancer intrinsic subtype samples fall along the differentiation axis. The data were fit onto both axis (MaSC → pL, and pL → mL), to determine whether the degree of differentiation is moving sequentially in our linear model or skipping phenotypes (pL for example). The luminal tumors had the greatest shift along the pL → mL axis, which was expected since these are the most differentiated of the tumor subtypes. Basal tumors, on the other hand, have more movement along the MaSC → pL axis than the pL

→ mL axis, though they are in the middle of the MaSC → pL axis largely due to heterogenous characteristics of basal tumors. The luminal B and HER2-enriched cell lines move much more closely along the MaSC → pL axis, likely due to the fact they are grown in culture and lose some mature luminal features such as a low proliferation rate, for example. The BABE cell line has been classified as basal by PAM50 and correlates with the basal cell lines on the MaSC → pL axis. BABES stably transduced with individual TFs from our panel also move more along the MaSC → pL axis, with BABE-GATA3 moving significantly in the pL direction. BABE-ESR1 moved the most along the pL → mL axis, making this a candidate gene for inclusion in the TF combinations to be tested in the next steps. Based on previous data suggesting that FOXA1 is critical for ESR1 function and response to 17β-estradiol (Hurtado et al., 2011), FOXA1 was also included in all of the combinations we tested. Two of the combinations shown, ESR1–FOXA1–PGR (EFP), and ESR1–FOXA1–GATA3 (EFG) show the greatest movement along the pL → mL axis, with EFG having a slightly higher mL score than EFP. Interestingly, BABE–PGR has a lower pL → mL score than the control BABE cell line BABE–GFP. Interestingly, BABE–ESR1 and BABE–EFP are indistinguishable by level of movement towards the mL phenotype, which may suggest that FOXA1 and PGR do not influence changes along the pL → mL axis like they do individually along the MaSC → pL axis.

2.2.3 ESR1 Expression in BABES Induces Cellular Senescence

We observed that stably transduced cells expressing different TFs displayed markedly different growth rates while being maintained in culture. This was quantified by plating monolayers of each TF-induced cell line, including both the parental and GFP-transduced BABES, and measuring viable cells by MTS assay after 4 days at 37 °C (Figure 2.1B). BABES expressing ESR1 ($p = .0001$, unpaired t test, 2-tailed) or AR ($p < .0001$, unpaired t test, 2-tailed)

grew more slowly than the other cell lines. To corroborate these data, we replated the same groups and stained them with EdU, a BrdU analog, to assess cellular proliferation via DNA replication. BABE–ESR1 cells display extremely low proliferative activity compared to negative control cells (Figure 2.1C) or to BABE–FOXA1 or –PGR cells (data not shown). Notably, a triple-transduction protocol including FOXA1 and PGR failed to rescue the growth suppression induced by ESR1 and these BABE–EFP cells died off after about 2 weeks after viral infection. In order to elucidate the mechanism underlying cellular arrest in the BABE–ESR1 cells, we performed a 2-class unpaired SAM analysis comparing the differential expression of genes between BABE–GFP and BABE–ESR1. We then overlaid these data onto a Cytoscape (http://www.cytoscape.org/) cell cycle pathway map developed in our lab. As shown, genes involved in halting cell cycle progression (p53, RB, p15) are upregulated relative to the BABE–GFP control group (Figure 2.1D). The alterations of cell cycle genes observed in BABE cells transduced with other TFs in our panel are also consistent with their cell count data (Figure 2.2). In addition, we performed a β-galactosidase assay to measure cellular senescence (data not shown), which confirmed the BABE–ESR1 cell lines were senescing relative to GFP control. Altogether these data showed that the BABE cell line was a difficult model with which to approach our objective of obtaining a transdifferentiated cell line with luminal A characteristics that could be propagated and studied in culture. Thus, it may not be an effective model for studying differentiation because of the inability to maintain the induced differentiated cells in culture.

2.2.4 MEFS Transduced with EFP (M–EFP) Exhibit Epithelial Phenotype

Cdkn2a$^{-/-}$ MEFS were transduced and serum starved for 1 month (Figure 2.3A). Based on our earlier findings, we transduced MEFS with several gene combinations, focusing on ESR1,

FOXA1, GATA3, DACH1, XBP1, and PGR. In order to determine which combination to pursue, ectopic expression was first confirmed by qPCR. Then RNA was amplified, labeled, and run on Agilent microarrays. Morphologic changes began to develop after maintaining the transduced cells for approximately 3 weeks in culture. Gene expression patterns were analyzed using a variation of the differentiation predictor method, in which the MaSC → pL and the pL mL axes were combined into a single MaSC mL axis as a way to measure overall differentiation. We then used this compilation score (d-score) value to screen different combinations of transcription factors. MEFS were transduced as described in methods, and after 1 month, stable expression of ectopically introduced TFs was confirmed by qPCR (data not shown).

Interestingly, while EFG produced the highest d-score shift in BABE cells (Figure 2.1A). This did not hold true when the combination was expressed in MEFS (Figure 2.3B). The EFP combination effect was more consistent between the two MEF cell lines, and so we initiated further studies with this combination. By 1 month, the M–EFP cells exhibited a uniform cobblestone pattern compared to the parental MEFS (Figure 2.3C). We analyzed protein levels of ESR1, FOXA1, and PGR on Day 30, as well as later passages (Figure 2.3B, data not shown), and the biological replicates maintain expression of the 3 proteins. In order to first assess whether the TFs produced a shift toward an epithelial phenotype, we performed qPCR on RNA prepared from MEFS and M–EFP after 1 month in culture to determine if there were expression losses in mesenchymal genes and/or gains in epithelial genes. As a screen, we chose 2 mesenchymal genes (Vimentin and COL1A2) and 2 epithelial genes (KRT14, a common basal breast marker, and CDH1). Both KRT14 and CDH1 were overexpressed relative to MEFS, while Vimentin and COL1A2 expression was decreased in M–EFP relative to MEFS (Figure 2.4A). This was confirmed by immunofluorescent detection of protein expression. M-EFP displayed positive

staining with CDH1 and KRT14, as well as pan-cytokeratin, while minimal or no staining for these proteins was observed in MEF control cells.

Interestingly, while MEFS stain positively for Vimentin as it is a classic mesenchymal marker, immunostaining of Vimentin in M-EFP did not correspond with the decrease in Vimentin RNA seen in the qPCR results. The immunofluorescence panel also included ERBB3, a tyrosine kinase expressed predominately in epithelial tissues, as well as a marker for luminal epithelia in the breast (Balko et al., 2012). ERBB3 expression is positive in the M-EFP cell line, with both membrane and nuclear staining present as compared to MEFS. Finally, we stained both cell lines with phalloidin, and under 40x-oil magnification, to examine cytoskeletal F-actin arrangements. As shown in Figure 2.4D, there is a clear morphological difference between the parental MEF and M–EFP, as the M–EFP cells take on a cuboidal epithelial phenotype with cortical actin circumferentially located around the membrane. On the other hand, the MEF cells retain their classic stress fiber structure.

2.2.5 M-EFP Cells Exhibit Expression of Breast Epithelial Markers

Both MEF and M–EFP cell lines were cultured in low-growth Matrigel and in it have distinct growth patterns at 10x magnification (Figure 2.5A). MEFs grew as dispersed single cells exhibiting little cell–cell cohesion, whereas M–EGP formed distinct spherical colonies. Western blot analysis for KRT5 and 18, markers for breast basal and luminal cells, respectively, demonstrated KRT5/18 expression in M–EFP cells, while MEF had no KRT5/18 expression (Figure 2.5B). To determine the localization pattern of KRT 5/18, we performed immunofluorescence staining of the cells in low-growth Matrigel for both keratins and found the M–EFP cells to be positive for both, with a concentration of KRT18 around the periphery of the spherical colonies (Figure 2.5C). Thus, M–EFP cells express keratins that are commonly

expressed in both basal and luminal mammary epithelial cell lines. Furthermore, we assessed three common markers for mammary stem cell plasticity and development (Lim et al., 2010). Both MEFS and M–EFPs were stained with CD24-PE, CD61-FITC, and CD29-APC antibodies. After performing fluorescent compensation, the samples were analyzed by flow cytometry as described earlier. MEFS have low CD24 positivity (35%) and high staining for CD29. As expected, MEFS were negative for CD61, a marker for the luminal progenitor population (Figure 2.5D). The M–EFP cell line was largely CD24 positive (84%) and was also CD29 positive, although somewhat lower than the MEF cell line. Approximately 15% of M–EFP cells were positive for CD61, suggesting there is a subpopulation of cells with a luminal progenitor marker (Figure 2.5D).

2.2.6 Genomic Analysis of M-EFP Suggests Shifts towards Mammary Development

GSA analysis comparing transcriptional factor networks (TFNs) demonstrated that the M–EFP cells were enriched in the transcription factors STAT5A and C/EBP. These are both involved in terminal mammary epithelial differentiation and development. The FREAC2 (FOXF2) TFN, which has been shown to play an important role in tissue homeostasis through regulating epithelium-mesenchyme interaction to maintain epithelium polarity, was also elevated in M–EFPs (Aitola et al., 2000). SRF (serum response factor), which is involved in epithelial–mesenchymal transition (EMT) and cellular proliferation, was significantly repressed in the M–EFP cell line (Busche et al., 2008). To better illustrate the differences of these pathways between the MEFS and M–EFP cell lines, density plots demonstrate the significant pathway differences, seen in Figure 2.6B. We separated the MaSC → pL and pL → mL axis again in order to assess how transdifferentiation occurred along each axis, as we observed that the M–EFP cells

→

exhibited both basal (pL) and luminal (mL) characteristics, plotting each along the x and y axes, respectively.

Our lab has previously demonstrated the utility of this algorithm across the human breast cancer subtypes. Comparisons in tumor types and mammary/breast development between human and mouse have been extensively studied (Visvader, 2009; Hollern and Andrechek, 2014) and the reported similarities provided the impetus to test this model with our mouse data. Initially, we examined mouse mammary tumor types to determine where they fall along the stem to luminal axes similar to their human counterparts. The C3(1)-Tag mouse model gives rise to basal-like tumor and falls in between the MaSC → pL axis in line with human basal tumors (Figures 2.1A and 2.6C) (Pfefferle et al., 2013). HC11, an immortalized mouse mammary epithelial cell line with basal-like features, exhibits a similar score along the axis. The mouse model MMTV-*neu*, a luminal tumor type, has the highest d-score among the mouse tumor/cell lines. However, due to the subtle differences in ESR1 and PGR regulation between human and mouse tumors, the d-score for human luminal tumors is higher relative to its mouse equivalent. MMTV-PyMT tumors are well characterized for their metastatic potential, which may explain why they are lower on the d-score than the MMTV-*neu* (Hollern and Andrechek, 2014). The PyMT cell line was derived from a C57BL/6 PyMT tumor and grown in vitro for several passages, which may also help explain why this derived cell line is slightly different from its FVB in vivo counterpart.

Our M–EFP cell line had a similar degree of movement along the MaSC → pL axis as the mammary norm tissue (p = .56) and significantly shifted from the parental MEFS (Figure 2.6C-ii). Interestingly, when M–EFP cells are treated with 17β-estradiol and progesterone for 4 days, the cells shift dramatically towards a pL phenotype relative to the cells without hormone

→

stimulation, while there is no significant movement along either axis when adding 17β-estradiol and progesterone to the parental MEFS (Figure 2.6C-ii). As a result, their score on the MaSC → pL axis is indistinguishable from the mammary lactating group (p = .44), suggesting that ligand-dependent activation of ESR1 and PGR are important for lactation (Brisken and O'Malley, 2010). The M–EFP cells were defined as being somewhat more differentiated on the pL → mL axis, as compared the M–EFP cells treated with hormones. ESR1 and PGR are known to activate bipotent/progenitor cell division when undergoing branching morphogenesis and/or alveologenesis (Brisken and O'Malley, 2010). A hormonally induced increase in proliferation would explain the decrease in d-score along the pL → mL axis. Mature luminal cells are generally nonproliferative, and increased proliferative gene expression profiles result in a lower d-score.

2.2.7 M–EFP Exhibit Estrogen-Dependent Growth

In order to determine ESR1 responsiveness in the M–EFP cell line, cells were treated with various concentrations of 17β-estradiol, ranging from 10pM to 10uM for 4 days, and assessed for proliferation (Figure 2.7A). While the parental MEFS lacked a response to 17β-estradiol after 4 days of treatment, the M–EFP cell line exhibits a concentration-dependent response with maximal effect at around 1 nM 17β-estradiol. The growth of M–EFP cells in standard culture medium containing phenol red was inhibited by the ESR1 antagonist tamoxifen with a calculated IC_{50} of 200 nM, whereas MEF cells were over 10-fold less sensitive to tamoxifen (Figure 2.7B). The growth responses to 17β-estradiol and tamoxifen were influenced by the presence of phenol red in the culture medium, which is well documented to have estrogenic activity (Welshons et al., 1988). Thus, the addition of 10 nM 17β-estradiol to phenol red–free media at 10 nM resulted in a loss of sensitivity in the M–EFP cells (Figure 2.8). Since

low proliferation rate as assessed by KI67 expression is a hallmark of the luminal A phenotype in human tumors (Inic et al., 2014), we examined KI67 expression by both qPCR and immunofluorescence (Figure 2.7C-D). M–EFP has significantly less KI67 expression than by MEFS (p = .0047, unpaired t test, 2-tailed), and less than HC11 and PyMT cell lines. This was confirmed by KI67 immunofluorescent staining; thus M–EFP cells cycle slower than their parental counterparts.

2.2.8 M-EFP Cell Line Is Responsive to Milk Producing Stimulus In Vitro

The primary specialized biological function of the normal breast or mammary gland is to produce milk when needed. Thus, the ability to synthesize milk components in response to prolactin is a key feature of differentiated breast luminal cells. After characterizing the morphological and genetic changes in the M–EFP cell line, we next sought to determine if these cells are responsive to prolactin. The MEFS and M–EFP cells were cultured either as monolayers in culture dishes or in Matrigel and stimulated with 100 ng/ml prolactin for 4 days. First, prolactin receptor (PRLR) protein levels were assessed since receptor activation is necessary for milk protein production. The 3 PRLR isoforms were expressed in M–EFP cells, whereas MEFS exhibited no expression of the long isoform and lower levels of the short isoforms than the M–EFP cell line (Figure 2.9A). PRLR expression in the M–EFP cells was confirmed by immunofluorescence (Figure 2.9B). We then assessed the ability of the cells to respond to prolactin by producing the milk proteins β-casein (CSN2) and whey-associated protein (WAP) (Figure 2.9A). M–EFP cells expressed both CSN2 and WAP in the prolactin-stimulated cells, with a lower amount of CSN2 detected in the M–EFP cells not treated with prolactin.

We further performed immunofluorescence for localization of PRLR and CSN2 in Matrigel cell cultures. Consistent with the Western blot results, the MEFS did not stain for PRLR

or CSN2 whereas M–EFP cells exhibited positive staining for both proteins. Not all PRLR-positive cells stained positively for CSN2, and CSN2 appeared to be present in the matrix around the cells, as well as within the cells, suggesting that it is secreted or released from the cells (Figure 2.9B). Consistent with the GSA STAT5A pathway enrichment in the M–EFP cells, we confirmed STAT5A and phosphorylated STAT5A expression in M–EFP by Western blot in both prolactin stimulated and naïve cells by Western blot (Figure 2.10). These results suggest that the M–EFP cells expressed pregnancy/lactation-specific genes not otherwise activated in nonbreast cells as with the MEFS.

2.3 Discussion

Efforts to create a breast luminal cell line in vitro have been futile thus far. Studies have been largely focused on the reprogramming of differentiated breast cancer cell lines into breast cancer stem cells containing bipotent capacities. Our scientific approach here was two-fold: first, determining the appropriate combination of TFs necessary to create a cell line with luminal characteristics, and secondly, achieving this in the appropriate host cell line.

We began with a simple hypothesis that the TFs required for luminal cell characterization and maintenance would be TFs enriched in luminal A tumors relative to all other tumor types (luminal B, HER2-enriched, basal, claudin-low). Our lab has shown significant overlaps of expression signatures between breast cancer subtypes and their corresponding normal counterparts (Prat et al., 2013); hence the TFs found in the tumor data set may be relevant for normal development. As mentioned earlier, several of these candidate TFs have been extensively studied in luminal breast cancers, and have roles in development (Kratochwil et al., 2003; Yeh, 2003; Bernardo et al., 2010; Brisken and O'Malley, 2010), milk production (Neville et al., 2002;

Brisken and O'Malley, 2010; Rajaram and Brisken, 2012), or growth (Lange et al., 1999; Wu et al., 2006; Thorner et al., 2010; Hasegawa et al., 2015).

We began transducing each of the 9 candidate TFs into BABE mammary epithelial cells to assess whether the TFs could induce luminal cell characteristics. Immediately, we noticed that the BABES transduced with either AR or ESR1 exhibited low proliferation rates, as determined by both the MTS proliferation assay, as well as EdU incorporation (Figure 2.1B-C). Functional pathway analysis of gene expression profiles (Figure 2.1D) revealed activation of p53/Rb/p15/p21 pathways in the BABE–ESR1 cells relative to GFP control, which explains why BABE–ESR1 cells grow so slowly. While some previous studies have reported opposing effects of estrogen and progesterone on breast cell proliferation (Mauvais-Jarvis et al., 1986), coexpression of PGR together with ESR1 and FOXA1 in our BABE model did not reverse the growth suppression (data not shown). However, the combination resulted in a higher d-score than for either gene alone, indicating that the combination may interact to promote luminal differentiation in our system. Because BABES are *hTERT* immortalized, major cell cycle control pathways are still intact, so growth suppression and senescence likely would continue to be barriers to creating a continuous cell line model going forward.

Several studies have suggested GATA3 is involved in luminal differentiation and hormone responsiveness (Koboldt et al., 2012; Cohen et al., 2014; Kong et al., 2011). Furthermore, GATA3 is part of the luminal A tumor signature and is significantly mutated in luminal A tumors relative to all other tumor types (Perou et al., 2000; Koboldt et al., 2012). In our BABE system, GATA3 produced the highest shift in d-score as a single transduced gene, as well as in combination with ESR1 (Figure 2.1A). Although for some TFs the MEF system proved to be more dynamic for revealing induced d-score shifts, in the mouse model, GATA3

failed to produce a meaningful increase in d-score or morphology change compared to our other candidate combo, M–EFP (Figure 2.3C). It is possible that no effect of GATA3 was observed in MEFS because in these cells it may induce genes that could obscure the d-score algorithm: for example, immune-regulated genes, as GATA3 is a known regulator of Th1-Th2 development (Fang et al., 2007; Prat et al., 2010). It is also possible that GATA3 is involved in the expression of endogenous ESR1, as this was the case in our GATA3-induced BABE cell line (Figure 2.11). This may explain in part why the addition of both genes ectopically does not confer a more differentiated phenotype. Finally, it is likely that GATA3 functions differently in different cell lineages or in the mouse, so that GATA3 may have more pronounced effects in a lineage-committed basal-like breast epithelial cell line than in a dedifferentiated embryonic cell line, as it is thought that GATA3 exerts its influence later in development, particularly during pregnancy (Asselin-Labat et al., 2007). Interestingly, although ESR1 and GATA3 are known to induce expression of each other in mammary epithelial cells, there was very little GATA3 expression in the M–EFP cell line, which may reflect the predominantly progenitor phenotype of these cells.

The creation of our MEF cell model that exhibited maximal transdifferentiation toward a breast luminal epithelial phenotype depended on two main conditions: first, using a combination of genes, with EFP acting together to yield the best d-score (Figure 2.3C), and second, the absence of serum during the transdifferentiation process. We serum-starved parental MEFS for 72 hours, ran them on microarrays and tested their d-score as well (data not shown). While the MEFS without serum displayed an increase in d-score relative to parental MEFS and M–GFP grown in serum, the genes involved in that increase were proliferation related and not breast-specific genes. Moreover, the MEFS in serum-free media did not exhibit a cobblestone morphology similar to the M–EFP, again showing serum starvation by itself does not promote

epithelial transformation and that the ectopic expression of EFP was required for that transformation to occur (Figure 2.3E).

The M–EFP cell line gained expression of key epithelial markers such as KRT14 and CDH1 and lost expression of one key mesenchymal marker, COL1A2. While M–EFP cells still maintained some level of Vimentin protein expression by immunofluorescence, CDH1 expression significantly increased relative to the parental MEFS, suggesting that epithelial transformation was occurring (Figure 2.4A-B). Increased cell–cell adhesion was evinced by the change in cell growth pattern from single discohesive cells to epithelioid colonies. To further confirm mesenchymal-to-epithelial transformation in these cells, phalloidin was used to demonstrate that the organization of F-actin stress fibers found in mesenchymal fibroblasts had changed to a cortical localization adjacent to the plasma membrane (Figure 2.4C).

Our M–EFP cell line furthermore expressed markers considered to be specific to both basal and luminal subclasses, including the positive expression of KRT 5 and 18, and breast lineage markers CD24 and CD29. KRT5 expression is associated with the basal-like subtype of breast cancer, MaSC, and progenitor normal cell types, while KRT18 expression is associated with luminal subtype of breast cancer and luminal normal cell types (Graham et al., 2011). By FACS analysis, M–EFP are $CD24^{high}/CD29^{high}$, which could be considered a progenitor cell (Figure 2.5B-C) (Sleeman et al., 2006). CD61 is normally expressed in the luminal progenitor population and lost in the MaSC and mature luminal fractions of normal mouse breast epithelial cells. We confirmed that a subpopulation of the M–EFP cells are CD61 positive (15%), which may be explained by one of several hypotheses (Lim et al., 2010). These cells may be more on the MaSC side of the MaSC pL axis with movement towards the pL phenotype, or they may be more on the mL side of the pL $\rightarrow$ mL axis. It is difficult to know for sure since they have high

CD29 but also express KRT18. Interestingly, we did not see any significant shift in EpCAM by FACS in either the single transduced BABES (data not shown) nor in M–EFP (Figure 2.12). This might be interpreted as showing a largely basal-like phenotype, despite their CD24 positivity, as EpCAM has been shown previously to better separate the luminal and basal populations more effectively than CD24 alone (Visvader and Stingl, 2014).

GSA afforded a glimpse into the regulatory networks that may be influencing the M–EFP phenotype. While future studies utilizing ChIP-seq would be more informative, the results from GSA suggested that transcriptional factor networks enriched in the M–EFP cell line relative to MEFS were involved with lumen formation in breast development (C/EBP β) and alveolar formation in pregnancy (STAT5A, FREAC2). Notably, both C/EBP β and FREAC2 (FOXF2) have been shown in previous studies to preferentially bind at PR promoter sites in breast and uterine cell lines (Yin et al., 2012), and C/EBP β is a known downstream target of PGR (Lange et al., 1999). In addition, several of the TF networks down in the M–EFP cells relative to MEFS are involved in EMT, which explains the change in morphology as seen in Figures 3E, 5C. Based on our data demonstrating that the M–EFP cells comprise a mixture of breast lineage and cancer subtype characteristics, we decided to examine the d-score as a two-dimensional spectrum rather than define it as a score determined by the addition of the two developmental axes.

It is worthwhile to note that there is not a mouse mammary cancer cell line that embodies the mature luminal characteristics of human luminal tumors. This is in large part due to the fact that mouse mammary tumors have lower ESR1 and PGR relative to human breast tumors, demonstrating why luminal-like mouse mammary tumors have more a progenitor luminal than a mature luminal phenotype. Treating M–EFP cells with 17β-estradiol and progesterone significantly increased their movement towards a progenitor phenotype, suggesting that ectopic

expression of the receptors may need ligand-dependent signaling to further differentiate. Further studies will need to demonstrate this is indeed the case.

While we have shown the activities of an agonist and an antagonist of ESR1 in the M–EFP cell line, curiously, MEFS have some endogenous ESR1 expression and yet do not respond to 17β-estradiol nor tamoxifen (Figure 2.7A), indicating their lack of estrogen-dependent growth. The IC_{50} of tamoxifen for M–EFP was roughly 10-fold less than the MEF cell line (Figure 2.7B); however, the sensitivity of M–EFP to tamoxifen is still less than that of the human luminal B breast cell line MCF-7, as reported by other labs (Lippman et al., 1976; Masamura et al., 1995; Hoffmann et al., 2004). However, M-EFP cells do respond to estradiol at similar concentrations to MCF-7 cells that have been grown in "normal" conditions (i.e., in phenol-red media) (Masamura et al., 1995). One explanation for these differences in sensitivity to tamoxifen and estradiol could be cell type, as MCF-7 cells are a cancer cell line, and the M–EFP cell line is immortalized but not transformed. While M–EFP cells exhibit responses to estrogen receptor signaling, they do not appear to be dependent on it for growth. In addition, they are not rapidly proliferating cells and thus may be less affected by down-regulation of the receptor. Finally, our M–EFP cell line is not grown in serum-containing media, and unlike the 17β-estradiol stimulation assay, the tamoxifen assay was performed in phenol-red media. After several attempts with phenol-red free media, with and without added 17β-estradiol, the cells were not only much more resistant to tamoxifen, they did not grow as well (data not shown). It is plausible that there are clear media differences beyond the dye component that affect their response to estrogen signaling, or more simply, phenol red may sensitize the cells to tamoxifen (Berthois et al., 1986). By contrast, the presence or absence of phenol red did not affect growth of the parental MEFS.

Studies have shown that PGR is necessary for alveologenesis and further side branching during pregnancy. Downstream signaling pathways are induced including PRLR and STAT5A, and upon stimulation by prolactin, milk proteins are ultimately produced. In mouse models lacking PRLR, STAT5A phosphorylation and production of the milk proteins WAP and β-casein is lost (Brisken et al., 1998). In this study, we demonstrated that M–EFP cells not only have STAT5A protein expression (Figure 2.10) but also enriched expression of STAT5A downstream targets as determined by GSA (Figure 2.6A-B). Along with STAT5A, there is PRLR co-expression (Figure 2.9A-B), which is required for milk production. While murine PRLR receptor isoforms are seen in a variety of cell types, the short and long isoforms (4 total, including 2 short isoforms) are responsible for reproductive development and lactogenic responses (Binart et al., 2003, 2010). We detected these 4 isoforms in our MEF model cell lines, and importantly, we found that that the long form is uniquely expressed in the M–EFP cell line (Figure 2.9A). We also considered whether PR/PRLR/STAT5A signaling alone is sufficient to induce milk production in any cellular context, or if the ability to produce milk depends on other "classical" mammary epithelial characteristics. Some of the established characteristics of mammary epithelial cells in culture include, but are not limited to, the formation of acini-like spheroids, polarization of the epithelial cell layers, and basal deposition of basement membranes (Debnath et al., 2003b; Tallkvist et al., 2015). Are external cell environmental cues leading to the establishment of cell polarity and acinic structures necessary for successful milk production? A mouse mammary epithelial cell line, HC11, is less differentiated than our M–EFP cells by d-score but when cultured in Matrigel produces acini-like spheroids and milk-producing features (Marte et al., 1995; Helguero et al., 2005). In our M–EFP model, the milk proteins WAP and β-casein were produced by prolactin stimulation; even the cells were grown as monolayers in

plastic culture dishes, as well as in Matrigel. Thus, expressing the EFP TF combination allowed MEFs to acquire the intracellular machinery to produce milk components without requiring a specific matrix interaction. However, the normal 3-D architecture may require the extracellular matrix environment to shape the ability of the cells to form polarized functional glands that ultimately deliver the milk product to the offspring.

To conclude, we demonstrated that transdifferentiation from one lineage to another via defined transcription factors could occur in a well-defined growth environment without a need for serum to provide additional factors. We also found that the selection of the starting cell line for transdifferentiation is paramount. Although the BABE cells are part of the same breast epithelial lineage as luminal-like cells, expressing the same set of 3 transcription factors in a mesenchymal cell line yields a much larger shift in d-score, morphology, and most importantly, function. We created a directly transdifferentiated cell line with epithelial and breast-specific characteristics, including proliferation dependence on estrogen receptor signaling and prolactin-stimulated milk production.

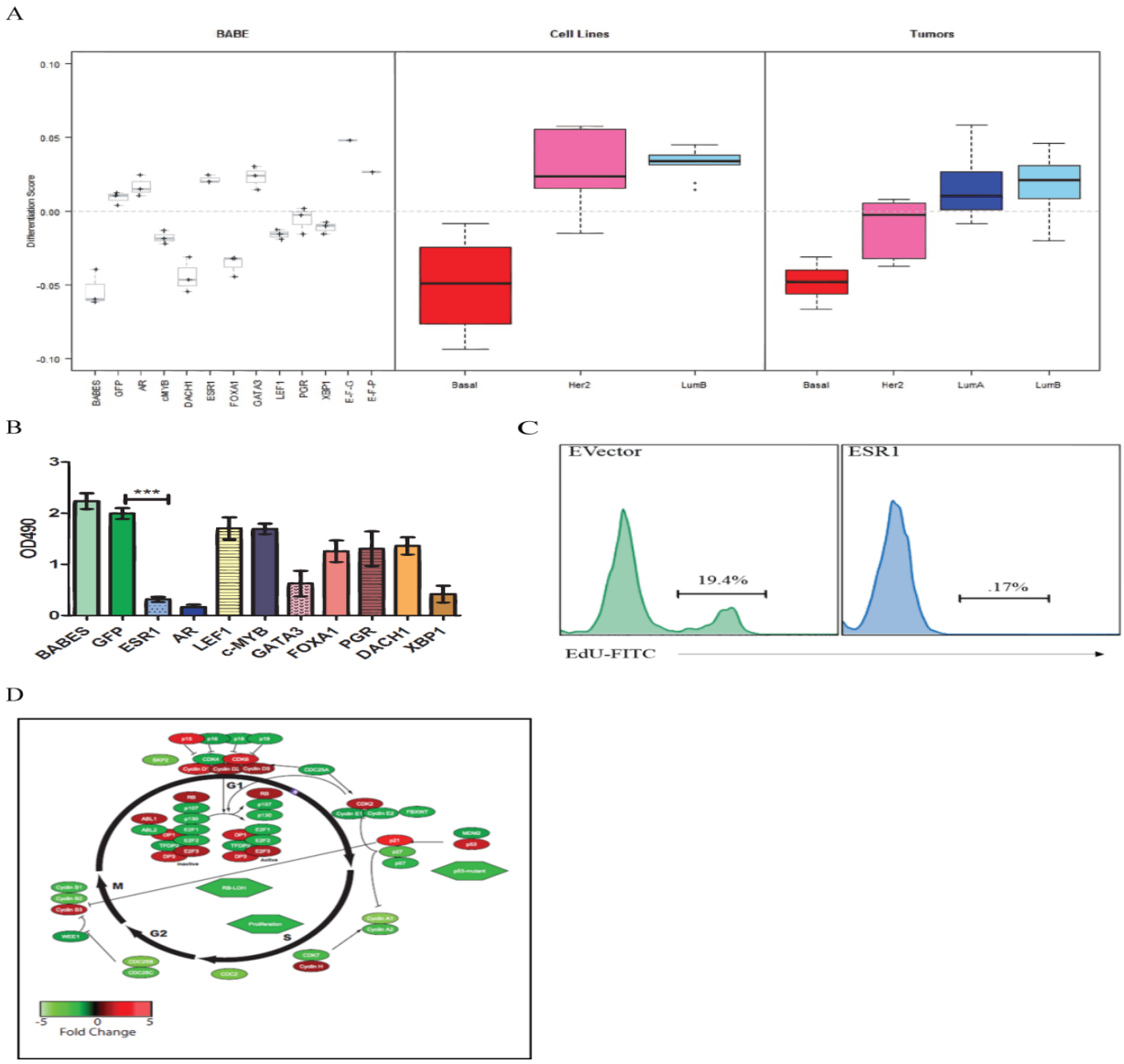
This initial work to create and characterize our luminal cell model opens the door to a variety of future directions to further characterize the model and to employ it to answer questions about luminal cell biology. The effects of various hormones on the breast are known to be quite complex. We have already found that the addition of such hormones to our M–EFP cells moves them significantly in d-score toward a more differentiated state, so more involved studies addressing hormone responses in our model would be enlightening. Secondly, can these cells be employed to create a model of mammary development and function in mice? Can they repopulate a mammary fat pad and create mammary gland architecture? Most importantly, are these cells truly reprogrammed? This question might be addressed by knocking out the

transduced genes to determine if the cells lose differentiation or function in the absence of one or more of the transcription factors? This work has produced a useful tool for understanding breast development, and the process can be used to offer greater insight toward the development of a mammary differentiated luminal A line, one whose growth in vitro has eluded scientists to date.

Table 2.1. Transcription Factor Candidate List

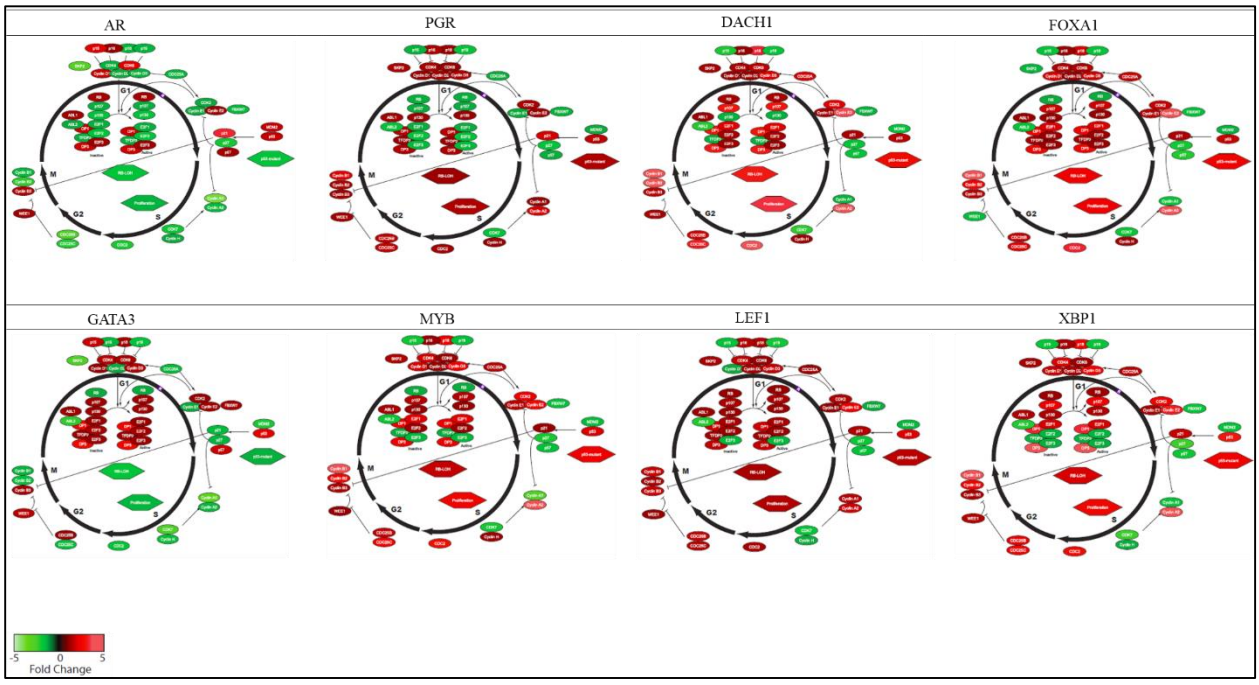
GENE	ACCESSION No.
AR	NM_000044
DACH1	NM_080759
ESR1	NM_000125.3
FOXA1	NM_004496.3
GATA3	NM_001002295.1
LEF1	NM_016269.4
MYB	NM_001130173.1
PGR	NM_000926.4
XBP1	NM_005080.3

Figure 2.1. TF Characterization in the BABE Cell Line



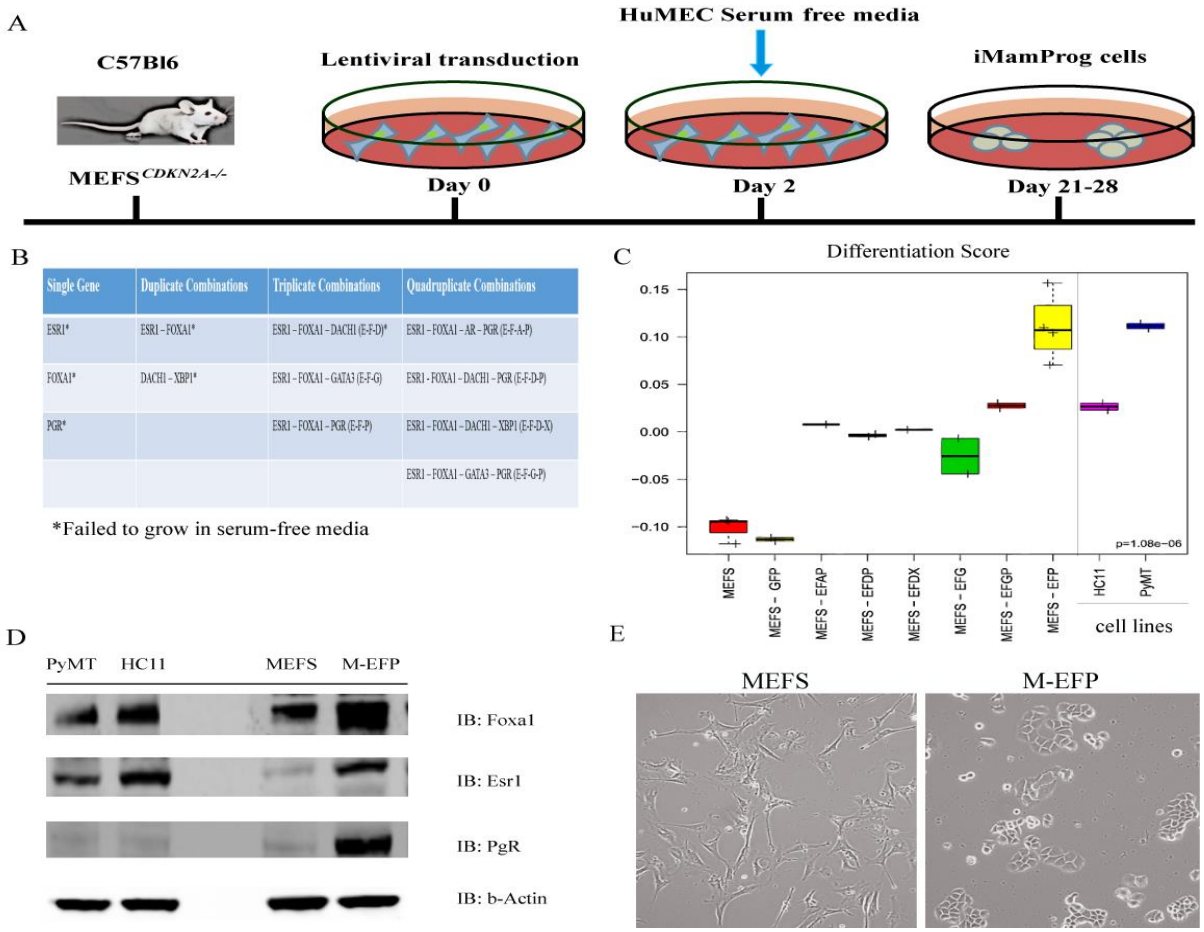
(A) D-score analysis and comparison across transduced BABEs, breast cancer cell lines, and tumors. **(B)** 4-day MTS assay comparing proliferation across the 9 different TF-induced BABE cell lines as compared to BABE–GFP control. *** p = .0001 for ESR1, error bars = SD. **(C)** FACS analysis of BABES with empty vector and BABES–ESR1 stained for EdU–FITC. Population of cells with EdU positivity are labeled by percentage. **(D)** Cytoscape illustration of cell cycle pathway. SAM analysis of cell cycle genes enriched in BABES–ESR1 relative to BABES–GFP control. Fold change is indicated by color, with red for higher expression in –ESR1 cells and green for higher expression in –GFP cells.

Figure 2.2. Cell Cycle Gene Expression Analysis of Single Transductants in the BABE Cell Line



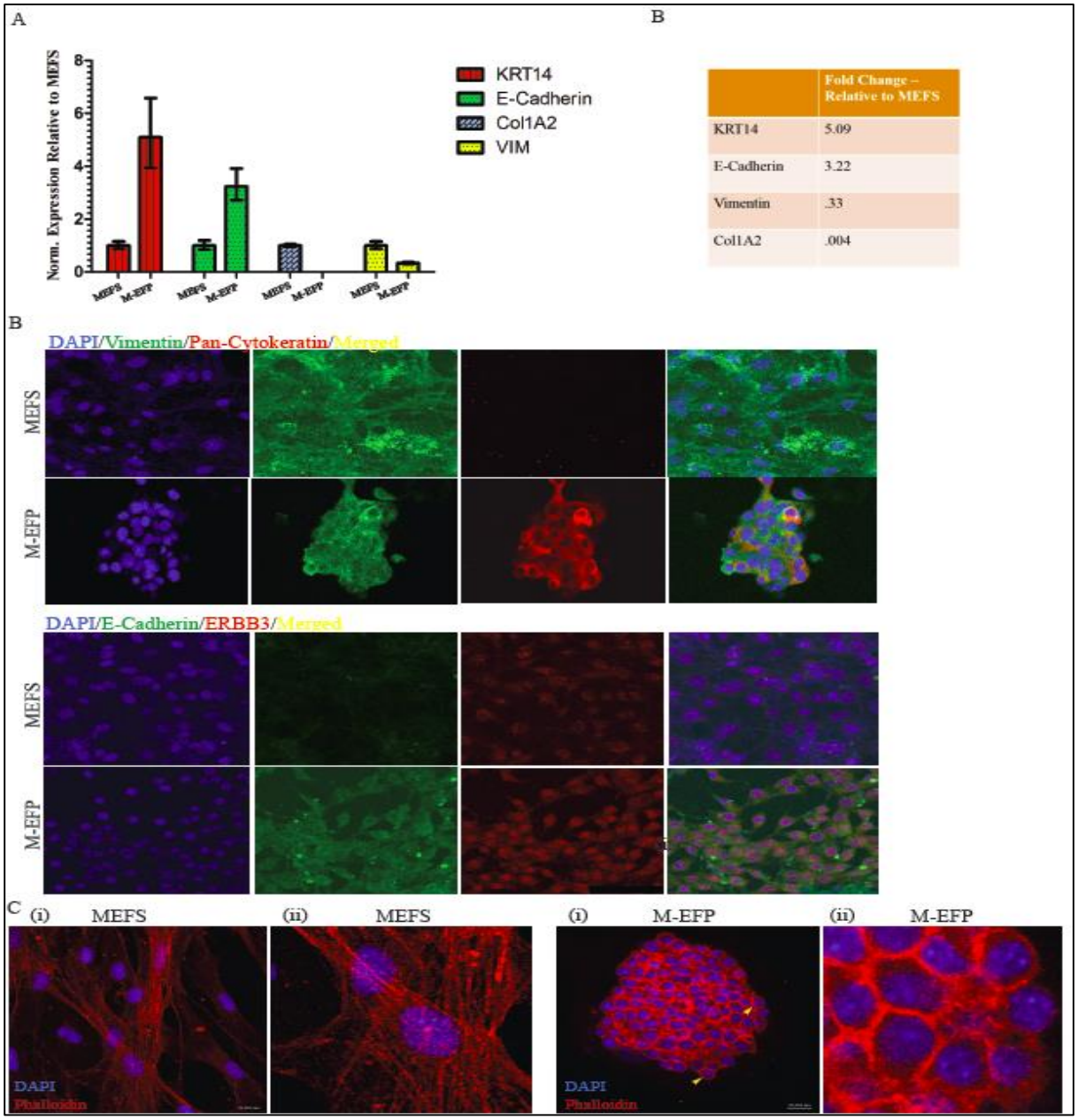
(A) TF combination screen setup in the MEF cell line. **(B)** Table of all the single and multiple TF combinations screened. **(C)** D-score analysis of the TF combinations screened in the MEF cell line, which include biological replicates of M–EFP. Cell lines HC11 and PyMT included as reference for basal and luminal-like, respectively. **(D)** Western analysis of proteins ESR1, FOXA1, PGR, and β-actin as a loading control in MEFS and M–EFP. As a reference, HC11 and PyMT cell lines are included. **(E)** Brightfield microscope image of MEFS and M–EFP at a magnification of 10X.

Figure 2.3. Initial Transduction Experiments in the MEFS



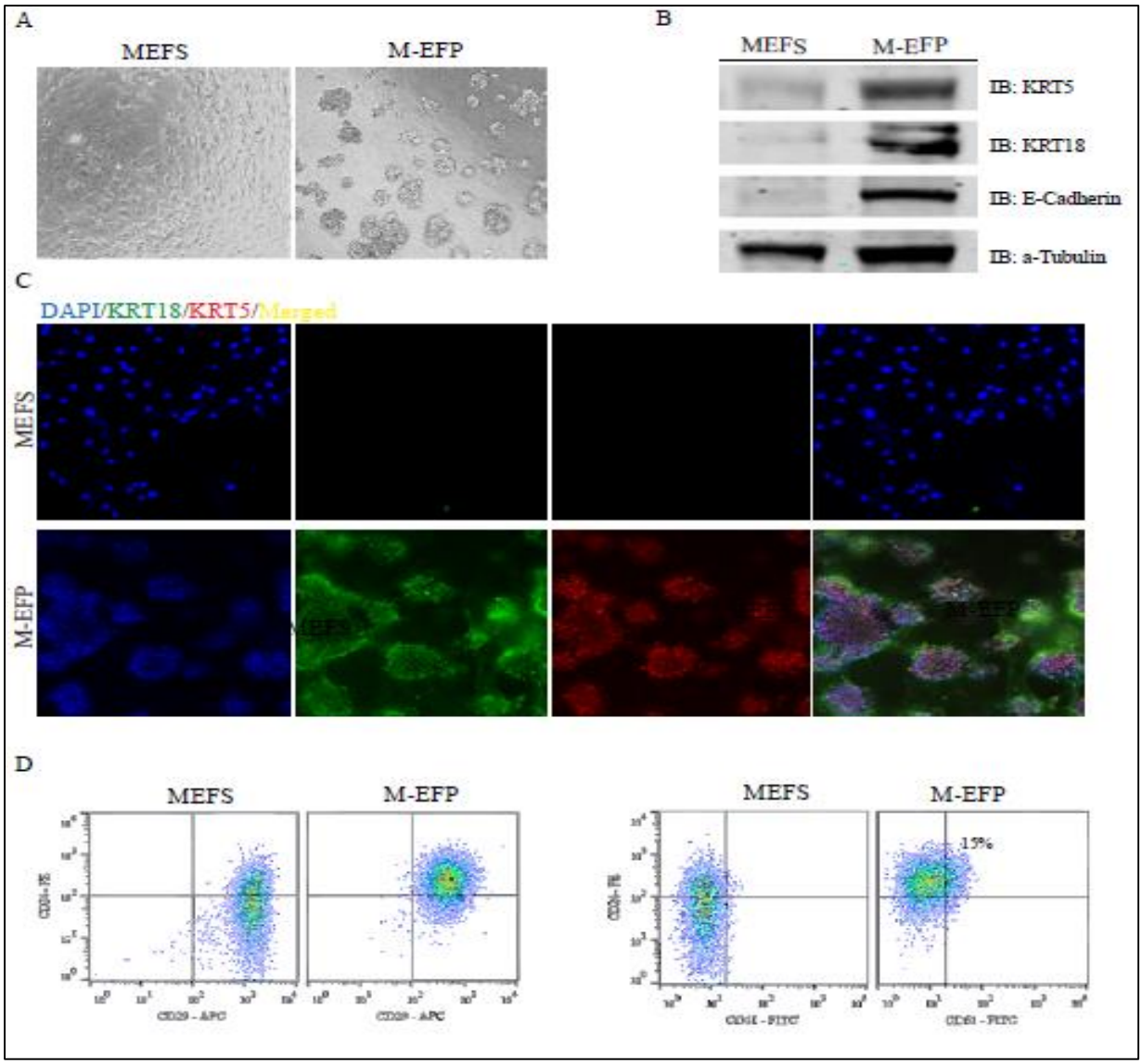
(A) TF combination screen setup in the MEF cell line. **(B)** Table of all of the single and multiple TF combinations screened. **(C)** D-score analysis of the TF combinations screened in the MEF cell line, which include biological replicates of M–EFP. Cell lines HC11 and PyMT included as reference for basal and luminal-like, respectively. **(D)** Western analysis of proteins' ESR1, FOXA1, PGR, and β-actin as a loading control in MEFS and M–EFP. As a reference included are HC11 and PyMT cell lines. **(E)** Brightfield microscope image of MEFS and M–EFP at a magnification of 10X.

Figure 2.4. Epithelial Characterization of M–EFP Cell Line



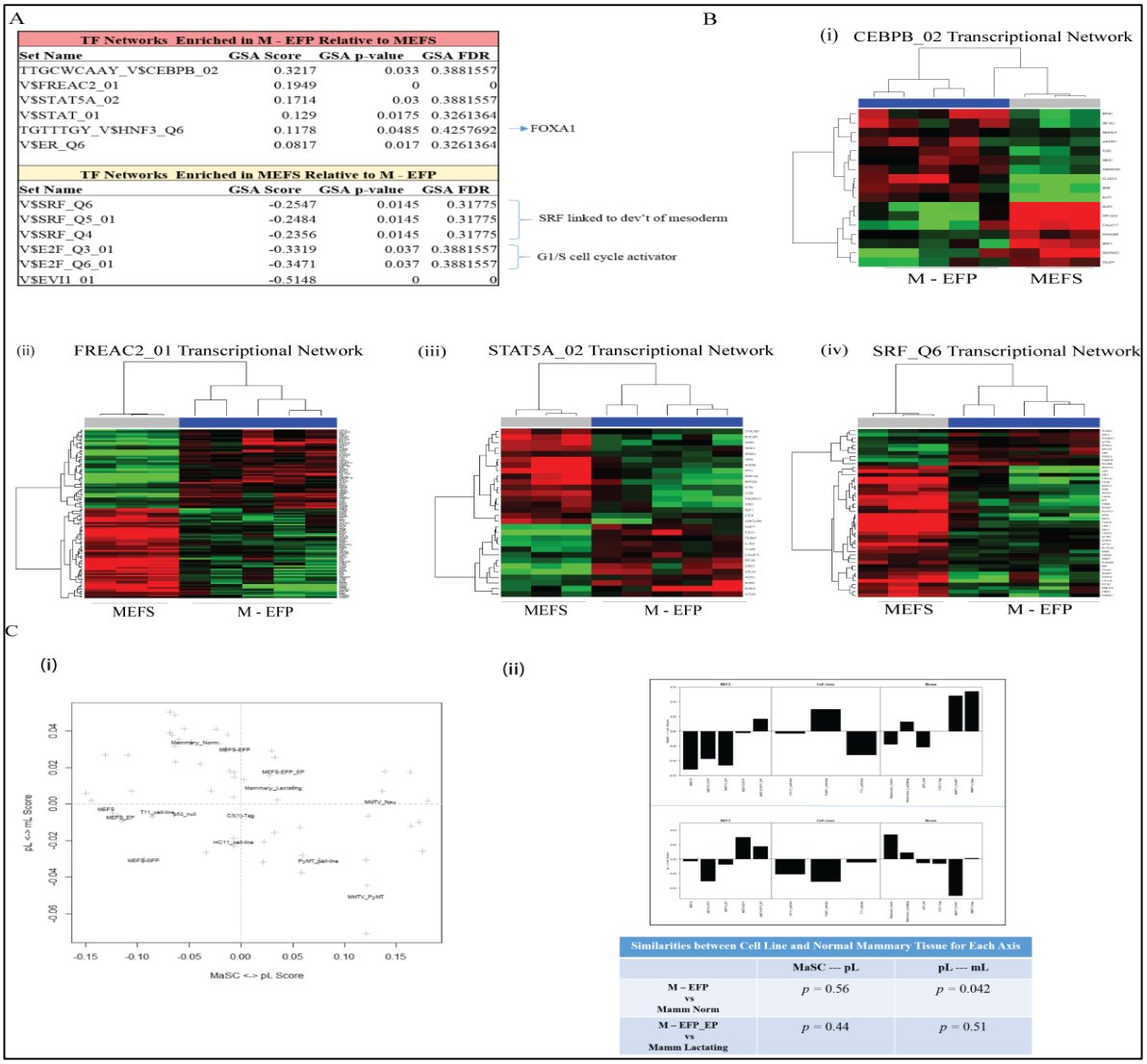
(A) Real-time PCR analysis of mesenchymal/epithelial gene expression in the parental MEF and M–EFP cell lines. **(i)** Data were normalized to housekeeping gene TBP; $\Delta\Delta Ct$ values of M–EFP are relative to MEFS. Error bars are SD. **(ii)** Table of relative fold-change values for genes KRT14, E-Cadherin, Vimentin, and Col1A2 from graph in **(i)**. **(B)** Immunofluorescence of the following, from left to right: DAPI, Vimentin, Pan-Cytokeratin, merged, in parental MEFS and M–EFP. Magnification at 20X. **(C)** Immunofluorescence of the following, from left to right: DAPI, E-Cadherin, ERBB3, merged, in parental MEFS and M–EFP. Magnification is 20X. **(D)** Immunofluorescence of phalloidin-red and DAPI staining in both the parental MEFS and M–EFP. **(i)** Magnification at 40X with oil. **(ii)** Zoomed-in view of F-actin staining.

Figure 2.5. Interrogation of Breast-Specific Phenotypic Markers in the M–EFP Cell Line



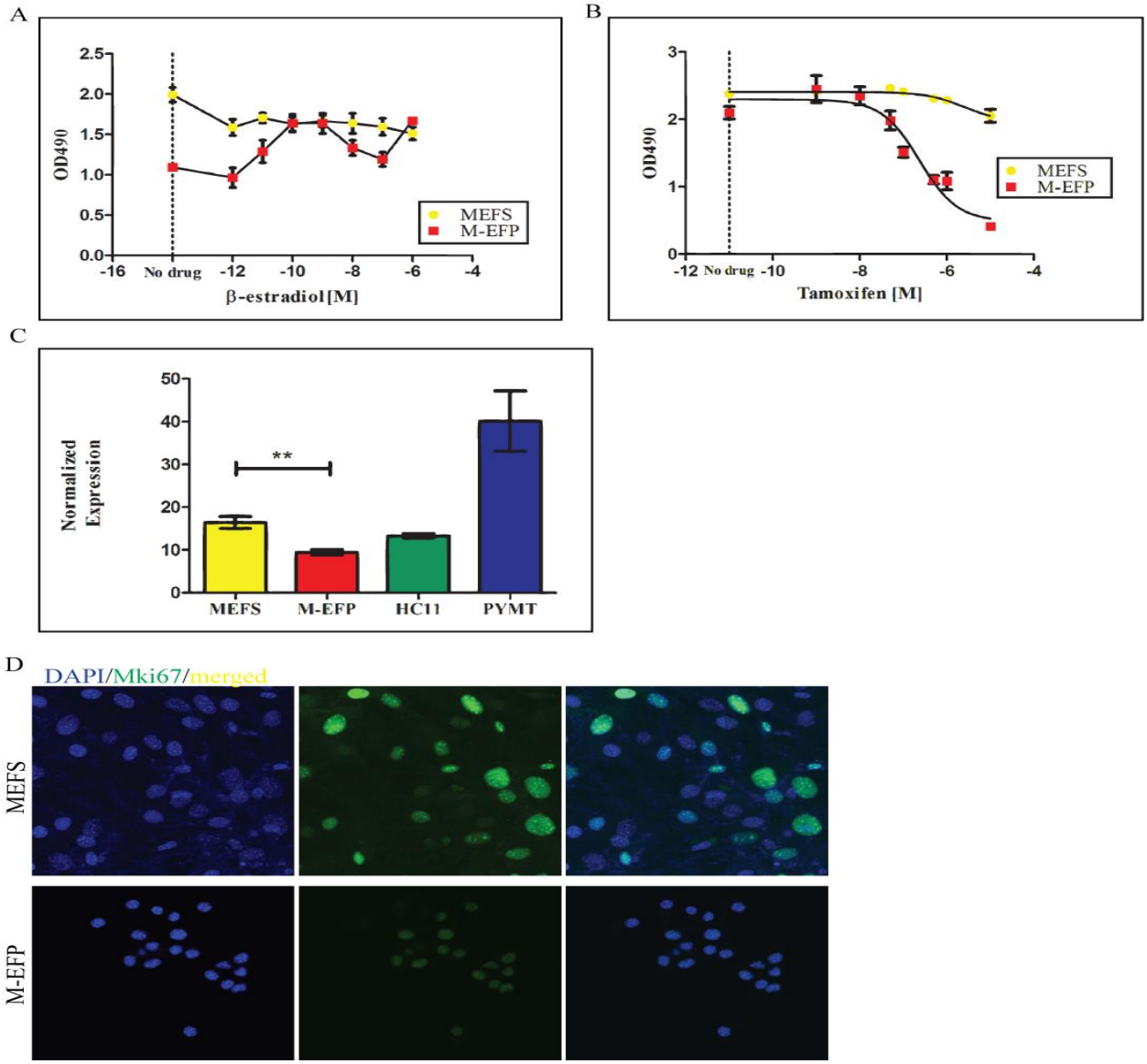
(A) Brightfield microscope image of MEFS and M–EFP grown in 3-D Matrigel. Magnification at 20X. **(B)** Western analysis of breast markers KRT5, KRT18, E-Cadherin, and α-Tubulin as a loading control in both MEFS and M–EFP cells. **(C)** Immunofluorescence of the following, from left to right: DAPI, KRT18, KRT5, merged, in parental MEFS and M–EFP grown in matrigel. Magnification at 20X. **(D)** FACS analysis of breast lineage markers in MEFS and M–EFP. **(i)** CD24 and CD29, and **(ii)** CD24 and CD61. A total of 15,000 events were collected for each sample.

Figure 2.6. Breast-Specific TFT Enrichment and Differentiation Analysis



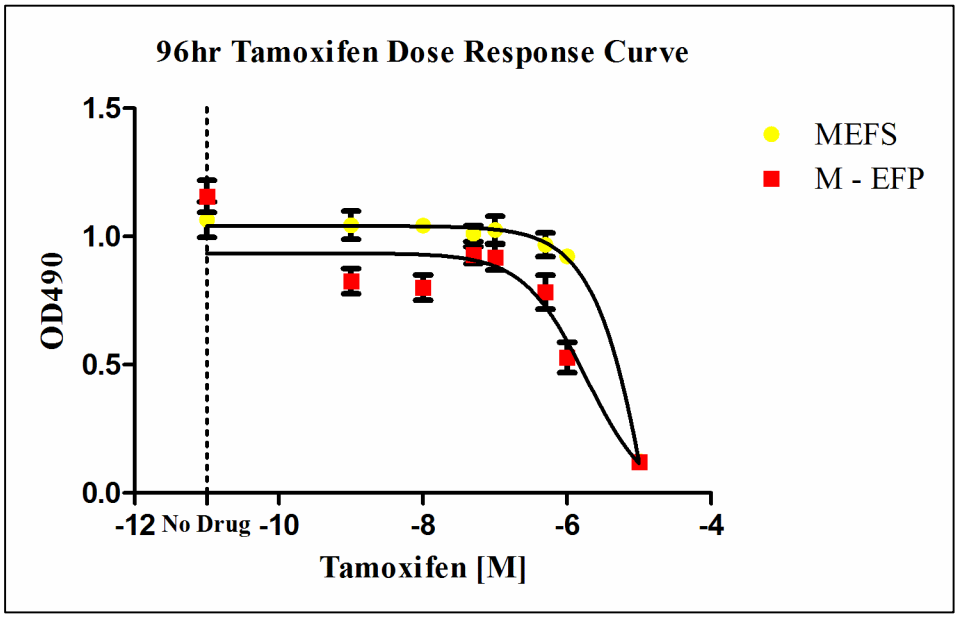
(A) Table of TF networks significantly up/down regulated in M–EFP relative to parental MEFS. GSA FDR < .5 for all listed TFT. **(B)** Hierarchical cluster of TF networks involved in breast development. TFT clusters' **(i-iv)** are comprised of transcription factors and their downstream targets, as described previously (Xie et al., 2005). **(i)** Cluster of CEBPB_02 (C/EBPβ) downstream target genes. **(ii)** Cluster of FREAC2_01 (FOXF2) downstream target genes. **(iii)** Cluster of STAT5A_02 (STAT5A) downstream target genes. **(iv)** Cluster of SRF_Q6 (SRF) downstream target genes. **(C)** D-score assessment via each of the two lineage vectors → MaSC pL, → and pL mL. **(i)** Scatterplot of d-score, by axis. The X-axis is the MaSC → pL vector, while the Y-axis is the pL → mL vector of the d-score. All mouse groups are labeled and individual samples are represented by a +. **(ii)** Bar graph depicting individual vector data across MEF, cell line and tumor/tissue groups. M–EFP treated with hormones' EP and mammary lactating groups not significantly different in either lineage vector, p =.44, p = .51. M–EFP and mammary normal group significantly different along pL → mL vector, p = .042, and similar along the MaSC → pL vector, p = 0.56.

Figure 2.7. ESR1 Responsiveness and KI67 Expression



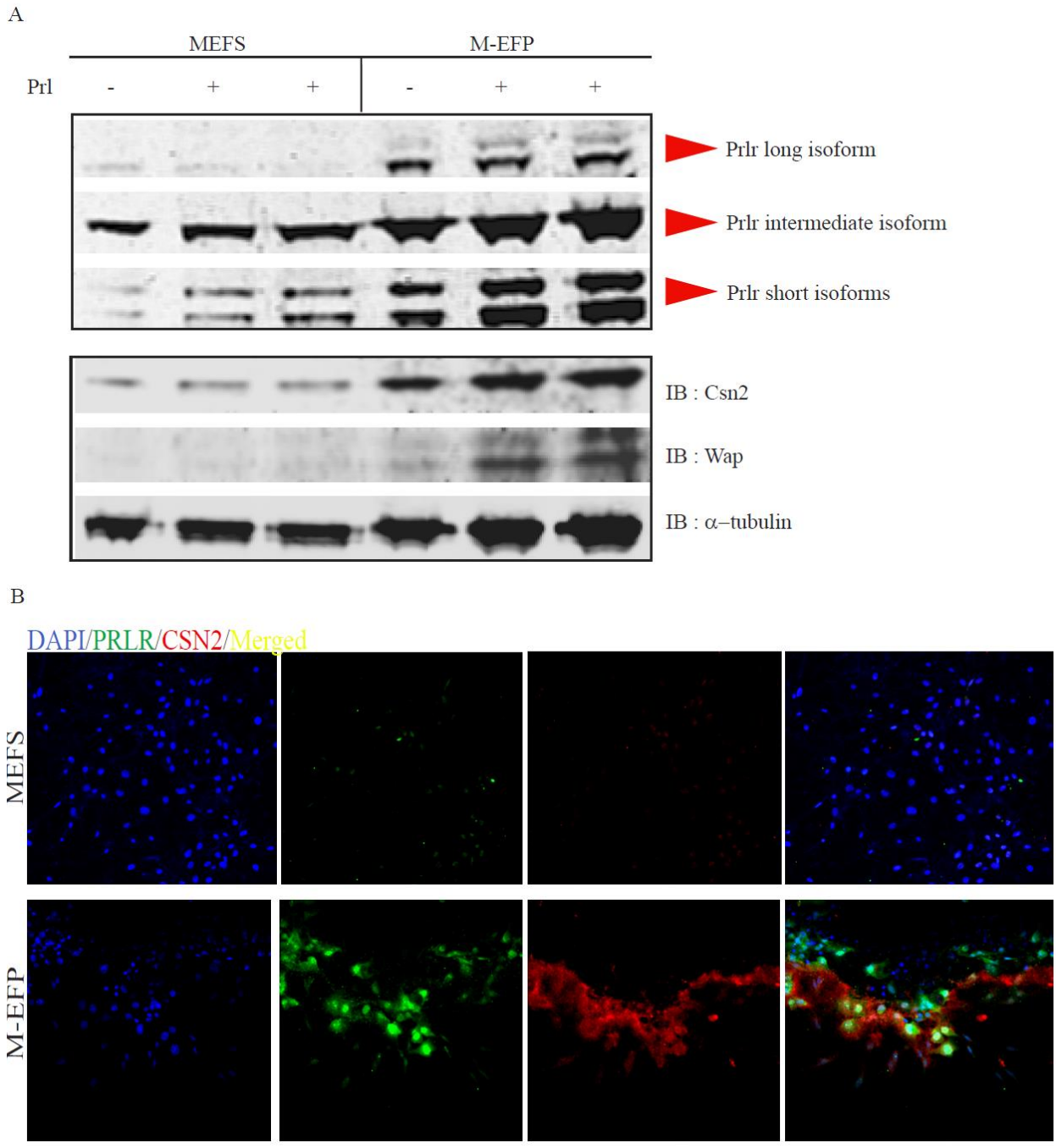
(A) MEFS and M–EFP cells were treated with varying concentrations of 17β-estradiol in phenol red-free media for 4 days then assessed for proliferation. **(B)** MEFS and M–EFP cells were treated with varying concentrations of tamoxifen in phenol red media for 4 days then assessed for proliferation. IC_{50} value for M–EFP is 200 nM. **(C)** KI67 expression in MEFS, M–EFP, HC11, and PyMT cells by real-time PCR. Samples were normalized with housekeeping gene TBP. **p* = .0047, error bars = SD. **(D)** Immunofluorescence staining from left to right: DAPI, KI67, merged. Magnification is 20X.

Figure 2.8. 96 hr Treatment with Tamoxifen in Phenol-Red Free Media



IC_{50} of M–EFP is 1.82 uM, while MEF IC_{50} is 23.92 uM.

Figure 2.9. M–EFP Exhibit Breast Functional Characteristics



(A) Cells were treated with +/- prolactin for 4 days. Then protein expression of PRLR, CSN2, WAP, and α-Tubulin (loading control) was determined via Western blot. **(B)** Immunofluorescence of MEFS and M–EFP in Matrigel, from left to right: DAPI, PRLR, CSN2, merged. Magnification is 20X.

Figure 2.10. STAT5A Activation in Prolactin Treated M–EFP Cells

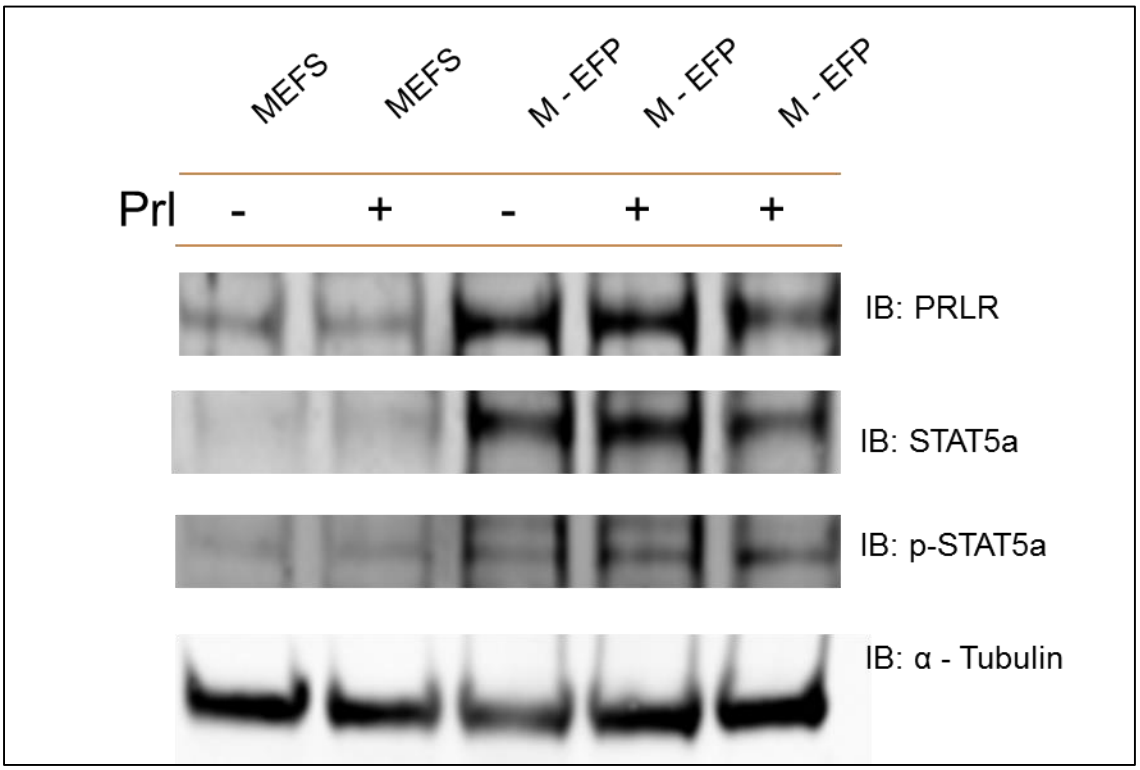


Figure 2.11. TF Regulation in the BABE Cell Line

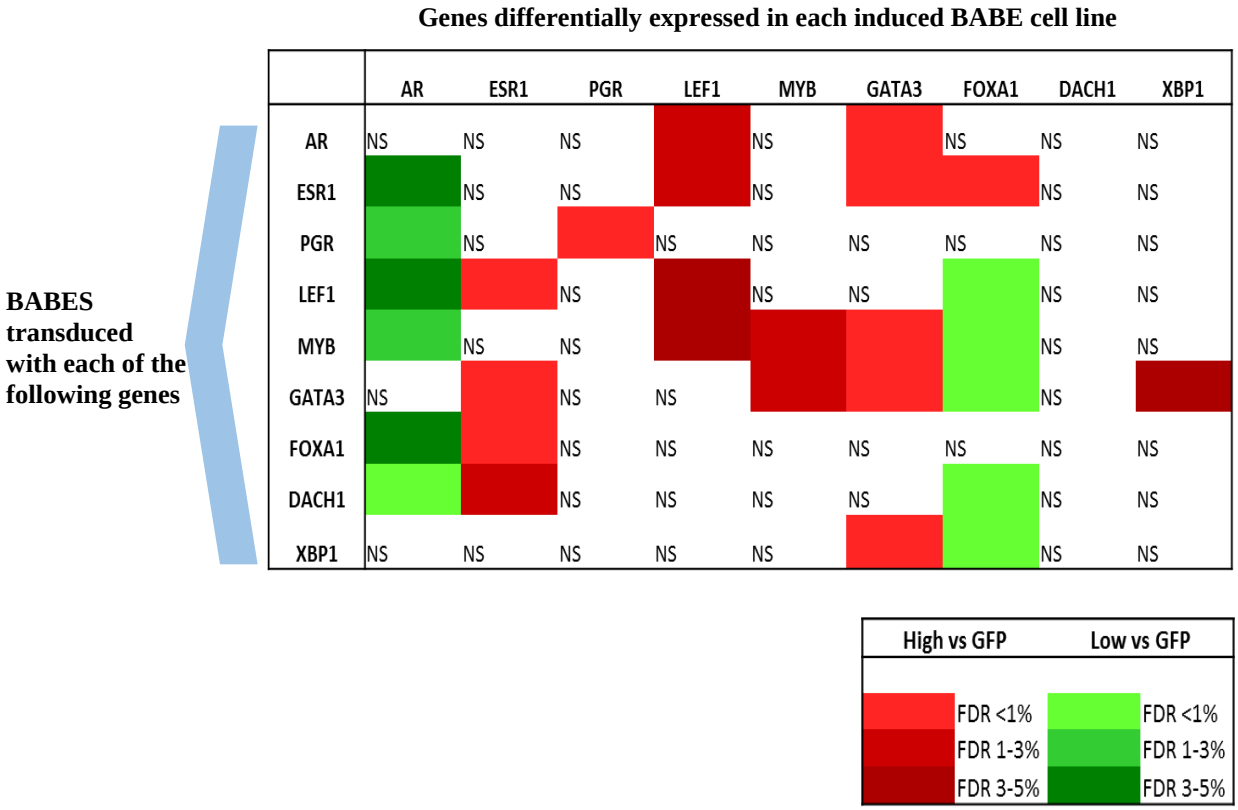
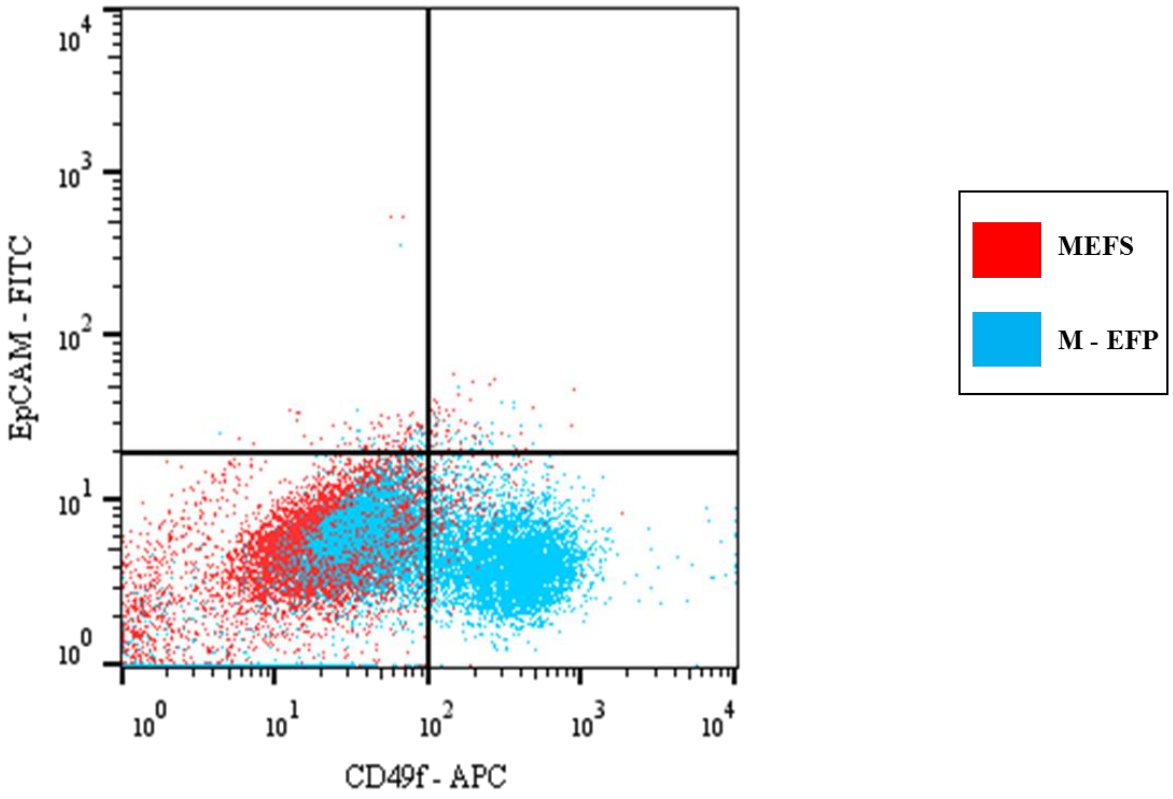


Figure 2.12. EpCAM/CD49f FACS Analysis in the Mouse Model



CHAPTER 3: ROLE OF GATA3 IN BREAST DEVELOPMENT AND LUMINAL TUMORS

Introduction

GATA binding protein 3 (GATA3) is a member of the GATA family of transcription factors. There are 6 members of the GATA family that are tissue specific. GATA3 has critical roles in the development of T cells and the kidney, mammary, and neural tissues (Grote et al., 2008; Hendriks et al., 1999; Kouros-Mehr et al., 2006; van Doorninck et al., 1999). In addition to its role in tissue differentiation, GATA3 is essential for proper embryonic development, as GATA3-mutant mice die at embryonic day 11 or 12 post coitum due to severe CNS abnormalities and deficiencies in liver hematopoiesis (Pandolfi et al., 1995), while in humans, GATA3 mutations result in hypoparathyroidism, renal dysplasia syndrome, and deafness (Nesbit et al., 2004; Van Esch et al., 2000).

GATA3 is a transcription factor that binds to either a consensus (A/T)GATA(A/G) sequence or palindromic site (ATCWGATA, W = A/T). GATA3 regulates target genes by directly binding to target promoter sequences or through long-range chromatin remodeling and DNA looping (Chen et al., 2012; Theodorou et al., 2012). Studies with X-ray crystallography were determined how GATA3 binds to DNA. The DNA binding domain (DBD) of the human GATA3 protein (amino acids 260-370) contains two zinc fingers, which are also known collectively as the N and C finger, for zinc finger 1 and zinc finger 2, respectively (Figure 3.1). Adjacent to the N and C fingers are critical basic regions that help facilitate the binding to GATA consensus sequences. The N and C fingers can bind to DNA independently of each other

and also have slight differences in sequence preference (Pedone et al., 1997; Visvader et al., 1995). Recently it has been discovered that GATA zinc fingers can bind to DNA via two different mechanisms: (a) both zinc fingers wrapping around a single palindromic GATA site and (b) a bridging method, where each N or C finger binds to two separate DNA molecules (Chen et al., 2012). Two more regions of the GATA3 protein important for DNA binding stability and regulation are the linker region and the C-terminus tail (C-tail). The linker region, located between the N and C fingers, is a highly basic region that allows for significant conformational flexibility, particularly when the N and C fingers are bound to different DNA strands. This allows for the simultaneous long-range regulation of multiple genes. While the amino acid sequence of the linker is similar to the C-tail, their functions are different. The C-tail, like the linker, contains a series of basic residues; however, the C-tail acts as a stabilizer, binding to the minor groove of DNA, wrapping itself around the DNA, to reach the N-finger (Chen et al., 2012). The dual role of the zinc finger domains combined with the flexibility of the linker region may result in effective lineage control during development.

GATA3 increases enhancer accessibility for ESR1-mediated transcription, in part explaining the significant correlation in expression to the luminal subtype of breast cancers (Theodorou et al., 2012). Evidence has suggested that GATA3 may act as a tumor suppressor, as ectopic expression of GATA3 results in reversal metastasis, induction of mesenchymal-epithelial transition, and loss of tumor dissemination (Dydensborg et al., 2009; Kouros-Mehr et al., 2008; Yan et al., 2010). However, in addition to directly regulating ESR1 expression, GATA3 upregulates genes involved in oncogenesis (Eeckhoute et al., 2006; Jiang et al., 2010). Promoter analysis of GATA3 target sites has demonstrated differential binding patterns between normal

and cancer cell lines, implying that the function of GATA3 is mediated in part by the oncogenic activity of the cell line (Cohen et al., 2014).

Recently, efforts have focused on genomic sequencing of tumors to better understand mutational drivers of disease. Several groups have published reports on GATA3 mutations in breast cancer patient samples following an initial report from our lab in 2012 (Arnold et al., 2010; Koboldt et al., 2012; Usary et al., 2004). Interestingly, the vast majority of GATA3 somatic mutations are found within luminal A tumors, further suggesting the existence of a relationship between GATA3 and ESR1 signaling in these tumors. Somatic mutations of GATA3 have been found enriched in a few key domains of the protein: Exons 4-6, which comprise the N and C fingers (DBD), as well as a portion of the C-tail, and the C-terminus (largely at the end of the protein, which does not have known motifs). DBD mutations would be expected to affect DNA binding, as the N and C fingers are responsible for the recognition of GATA-specific and palindromic motifs. To characterize the functional consequence of DBD mutations, the luminal B cell line, MCF-7, contains a GATA3 mutation, a nucleotide (G) insertion at position 1566, which is located in the DBD. Comparing the mutant MCF-7 with MCF-7 overexpressing wild-type GATA3 resulted in a differential loss of key GATA3 downstream target genes in the mutant cells (Cohen et al., 2014). This mutation confers a truncated GATA3 protein, which, in addition to DBD malfunction, may also affect the stability properties of the C-tail. Mutations in the C-tail may influence GATA3 stability on the DNA molecule, as binding the minor groove provides this. Strikingly, the TCGA breast cancer data set published in 2012 not only confirmed that GATA3 mutations were present significantly in luminal A tumors compared to all other intrinsic subtypes, but there were a significant number of mutations at the C-terminus between amino

acids 400–440 (Koboldt et al., 2012). The significance of this is yet to be understood due to the lack of known motifs or function within the last 40 amino acids of the protein.

Based on the locations of the GATA3 mutations found in our lab and others, I want to further understand the functional consequences of these mutations, particularly at the C-terminus, and how this region impacts the binding and regulation of ESR1-responsive genes. In the larger context, discerning the function of these mutations may help define the role of GATA3 in these cells. Does it act as a tumor promoter, or accentuate the ESR1 phenotype, giving these luminal A tumors a good prognosis relative to the other tumor types? I began by developing a transfection system and followed with structured bioinformatics analysis. Continuing experiments include the creation of the mutant GATA3 constructs from the distinctive regions of high mutational activity in the C-terminal tail as reported in the TCGA data (Koboldt et al., 2012). The work performed in this chapter describes that analysis, which sets the stage for others to pursue experimental validation of several hypotheses.

3.1 Methods

3.1.1 GATA3 Expression and Reporter Plasmid Construction

The GATA3 wild-type sequence was obtained from the Harvard Human ORFeome, Version 5.1 (http://horfdb.dfci.harvard.edu/hv5/). The open reading frame (ORF) sequence corresponded to NCBI accession # BC006793. The GATA3 ORF was in a pDONR223 plasmid, making it compatible with GATEWAY technology (LifeTechnologies, Inc). A lentiviral mammalian expression plasmid that was GATEWAY compatible was chosen as the destination vector for wild-type GATA3 and had the following features: PGK promoter with *Neo* resistance (pLENTI-PGK-DEST vector, Addgene.org). Recombination of GATA3 ORF into the pLENTI-PGK-DEST vector was performed as recommended by Invitrogen's GATEWAY protocol. In

addition, GATA3 cDNA inserted into a pBABE-puro retroviral vector (http://www.addgene.org) was also tested. Two forms of DNA were used to amplify the FOXA1 region of interest. Genomic DNA from normal human females (a kind gift from the D. Eberhard lab, UNC) and a BAC clone that spans 2.4kb of FOXA1 promoter–5' RNA sequence were both tested as templates to create a GATA3 binding sequence for the reporter construct (BAC clone RP11-356O9, Empire Genomics). A 1.5kb sequence flanking FOXA1 promoter and 5' RNA sequence were generated with the following primers: **FOXA1-F** : CCGCAGTGCAGATGCGTTCCC and **FOXA1-R**: TCACCTCCTGCGTGTCTGCGTAGT. Subsequently, the amplified FOXA1 sequence was cloned into a pGL4.17-luc reporter construct (Promega) after adding restriction sites Sac1 and HindIII to the FOXA1 F and R primers. The FOXA1 insert was validated by restriction digest and sequencing.

3.1.2 Cell Culture and Transfection

The 293T cells used for the luciferase assay were grown in DMEM-high glucose (Gibco, Life Technologies) with 10% FBS with 100X pen–strep when expanding for experiments. During the selection process, 293T cells were placed in expansion media with the addition of either 200 ug/ml G418 (Gibco, Life Technologies), 1 ug/ml puromycin (Invitrogen, Life Technologies), or both. Prior to the transfections, 293T cells were plated at a density of 4×10^4 in 24-well plates. Next, 1 ug of pGL4.17-FOXA1-luc, pGL4.17-luc, and pBABE-GATA3-puro each was added with 3 ul Fugene transfection reagent (Roche) in 200 ul of media without antibiotics. This sat at room temperature for 20 min. For the plasmid combinations, 1ug each plasmid was used. There were 3 conditions total and each was run in triplicate: pGL4.17-luc only (neg ctrl), pGL4.17-FOXA1-luc, and pGL4.17-FOXA1-luc + pBABE-GATA3-puro.

3.1.3 Luciferase Assay

After 10 days, cells were washed in 1X PBS. 100 ul of reporter lysis buffer from the Luciferase Assay System (Promega), which was added to each well. Then they were immediately placed on dry ice. Once frozen, the cells were placed in 37 °C until warmed. Next, 5 or 20 ul of each replicate was added to white luminometer plate, and 100 ul of luciferase assay reagent was added by injector to each well using a luminometer (Molecular Devices) with a time delay of 2 s and read time of 10 s. Both 5 and 20 ul of cell lysate was quantitated for protein. The luciferase signal was normalized to protein levels for each group.

3.1.4 Structural Analysis

The ExPASy (http://www.ExPASy.org) online tool was used to align multiple GATA family protein sequences within metazoans. There was a total of 1,000 sequences with 250 alignments. Looking only at vertebrates, we accepted all sequences with an e-value of 10^{-6} or lower. In all, 250 sequences within vertebrates were used in the analysis. ClustalX 2.1 (http://www.clustal.org) was used to visualize sequence similarities and conserved regions across the six GATA genes (1–6) aligned in ExPASy. In predicting protein-folding patterns and regions of disorder in mutated GATA3 sequences, the Regional Oral Neural Network (RONN) was used (https://www.strubi.ox.ac.uk/RONN).

3.2 Results and Discussion

Based on recent data, GATA3 mutations tend to occur in luminal A and B tumors, consistent with the expression patterns of GATA3 in both normal and cancerous luminal cells (Chou et al., 2010; Koboldt et al., 2012; Yoon et al., 2010). In this study we set out to determine the functional role of GATA3 mutations, focusing on the C-terminus, where mutations are common but the function is currently unknown.

Initially, we created a both a lentiviral and retroviral expression plasmid containing the wild-type GATA3 sequence (Figure 3.2). Due to the size of the vector and sequence, we went forward with the retroviral plasmid (pBABE-GATA3-puro). Others in our lab had success transfecting pBABE-GATA3-puro, so there was no immediate need for packaging the virus for infection. Previous microarray experiments using overexpressed pBABE-GATA3-puro in breast cancer cell lines SUM149 and SUM159 confirmed the expression of GATA3 and its downstream effector genes (data not shown). The pBABE-puro backbone was also used for expressing the GATA3 mutants.

Since GATA3 is a transcription factor that recognizes specific GATA and palindromic motifs, we interrogated the promoter regions of genes known to be regulated by GATA3 in mammary development. Interestingly, we found several GATA3 binding sites in the FOXA1 promoter region with the online tool, rVista 2.0 (http://rvista.decode.org). Other studies have also reported GATA3 acting upstream of FOXA1 signaling (Kouros-Mehr et al., 2006; Theodorou et al., 2012).

The FOXA1 sequence encompasses some of 5' RNA coding region, but the site lies predominantly upstream within the CpG island-heavy promoter regions (Figure 3.3). This 1.5kb region contains numerous CpG islands and has at least one published GATA3 binding site (Kouros-Mehr et al., 2006). We attempted to amplify this region with various primer pairs; however the CG-rich sequence made it difficult to amplify. Hence we shortened the sequence from 2 to 1.5 kb upstream from the ATG start site. A recently published paper had success amplifying this region, so we modified our primers using published primer sequences (Naderi et al., 2012). Initially, our DNA template came from a human female DNA sample used as a normal control for DNA sequencing. After several attempts to reduce nonspecific amplified

products, we ordered a BAC clone that contained the FOXA1 gene and its upstream sequence (Chr14 : 37,940,666-38,144,263, based on alignment from Feb. 2009-GRC37/hg19). Subsequent PCR amplifications suggested that the addition of DMSO helped resolve the sticking of DNA strands, allowing for more efficient amplification using both the genomic DNA and the BAC clone (Figure 3.4A). Finally, we cloned this fragment into the pGL4.17-luc reporter construct (Figure 3.4B), which was ready to be tested along with the ectopic expression of wild-type GATA3.

There was a significant luciferase signal with the transfection of the pGL41.7-GATA3-luc as compared to the pGL41.7-luc empty vector control; however, when we transfected both reporter and pBABE-GATA3 vector together, the luciferase signal decreased (Figure 3.5). The loss of the luciferase signal occurred in both cell quantities and was consistent regardless of whether normalization with protein concentration was performed. This was not the result we were expecting, though there are several possibilities as to why this may have occurred. First, the loss in signal with the pBABE-GATA3 may be due to less promoter construct entering the cell. Further experiments examining transfection times and sequences will need to be carried out. Because lipophilic-based transfections are much more inefficient in cellular incorporation than viral methods, one thought is to create virus for the wild-type GATA3 construct, create a stable cell line, and then transfect those GATA3-expressing 293T cells with the reporter plasmid. Another explanation lies within the biology of the FOXA1 promoter sequence. Our motif analysis showed that in addition to GATA3, several genes bind to the FOXA1 promoter region (data not shown). Therefore, it is possible that the addition of ectopically expressed GATA3 may indirectly affect the binding of TFs that bind to FOXA1, causing a decrease in signal compared to the reporter-only cells. This was hard to resolve as we did not determine which TFs (including

GATA3) were binding the reporter. Alternatively, repeating this experiment with a GATA3-positive cell line such as MCF-7s can help objectively measure GATA3 activity without the technical obscurity of multiple transfection efficiencies. Likewise, with a positive control cell line, a luciferase reporter for a housekeeping gene would also be an effective indicator for transfection efficiency of both constructs. With the housekeeping gene reporter, luciferase signal is expected to be robust; however, transfection with an additional plasmid may affect its transfection and, thus, the signal. As illustrated, there is a high likelihood luciferase signal losses with dual transfections are due to technical reasons. Future experiments in addition to those mentioned will include the testing of sequential transfection and transduction methods with both reporter and expression plasmids.

Recent studies in breast cancer have explored the frequency of mutations enriched within the intrinsic subtypes. GATA3 was shown to be mutated at a significantly higher rate within the luminal tumors (Koboldt et al., 2012). Further analysis determined an enrichment of mutations in the C-terminus of the GATA3 protein (Figure 3.6). In order to elucidate the role and consequences of GATA3 mutations, we set out to determine which protein domains within GATA3 are conserved across both the family of GATA proteins and across vertebrates. Two key observations were seen after the analysis. First, across all vertebrates, GATA3 was closest in sequence to GATA2, with a homology of 55% (Figure 3.7) (Chou et al., 2010). Likewise GATA3 is extremely homologous across mammals, with a 97% homology between human and mouse. Interestingly, GATA2 has been shown to facilitate androgen receptor (AR) signaling in prostate cells, suggesting a role for GATA2 in hormonal signaling (Wu et al., 2014). Finally, it is worth noting that the last several residues in GATA3 are not only conserved across mammals but also well conserved across some avian species (Figure 3.7).

Given the frequency of mutations in that region, coupled with the degree of conservation across avian and mammalian species, this region has an extremely important function in ESR1-positive luminal breast tumors. The C-tail is a region rich with basic residues and is instrumental in establishing stability with DNA (Chen et al., 2012). Our lab sought out to determine the functional consequence of disrupting the C-tail region by site-directed mutagenesis at residue 367, from a charged, basic Arginine to a hydrophobic Leucine (Usary et al., 2004). The ectopic expression of GATA3-R367L mutant resulted in RNA expression levels similar to the GATA3 wild type in 293T cells. However, in contrast to the GATA3 wild type–expressing 293T cells, there was no nuclear GATA3 protein detected in the GATA3-R367L (Usary et al., 2004). It is possible that the Arginine to Leucine change in the C-tail rendered the protein unstable and unable to effectively bind to DNA within the nucleus.

Armed with this information, we focused our current effort in three keys areas (Figure 3.6). Due to the consistent reporting of the importance of the zinc finger domains, we planned to create a GATA3 mutant construct with the addition of an A at the Arginine residue 330. This would shift the amino acid sequence, resulting in an early stop codon, thus truncating the protein to 351 total amino acids. A previous study had similarly knocked out the C-finger function, which resulted in abrogated IL-4 and IFN-γ expression, demonstrating that loss of the C finger negatively affected binding to DNA (Ranganath and Murphy, 2001).

The second region of interest is a highly mutated area in the more distal C-terminus beyond the basic C-tail (amino acids 385–440). In the TCGA breast cancer mutational analysis, there were 6 confirmed mutations in the luminal B cohort and 5 in the luminal A cohort at residue 409 (Koboldt et al., 2012) (Figure 3.7). Upon further inspection, sequencing results showed an insertion of a G at this location, resulting in an in-frame shift with a larger than

normal protein product of 509 amino acids (Figure 3.8). Likewise, our third proposed mutagen site is found at the extreme C-terminus at residue 431. Mutations were confirmed at residue 431 in both luminal A and B patients and were either a TT insertion or a T deletion. Regardless of the type of mutation, the protein result was unchanged. As with the mutations seen at residue 409, the protein size was predicted to be 509 amino acids, significantly larger than its wild-type form (Figure 3.8).

Recently, a published study examined the functional consequence of an elongated GATA3 protein based on the frequency of this mutation in the familiar disorder hypoparathyroid-deafness-renal (HDR) syndrome (Ali et al., 2007). Demonstrating that the elongation of the protein resulted in loss of DNA binding, Ali et al. also predicted that the excess amino acids interrupted the binding stability due to steric hindrance. It is not known, however, how these structural and functional defects contribute/correlate with the luminal phenotype as seen in breast cancer.

Due to the growing evidence that C-terminus mutations in GATA3 may impact luminal breast cancer biology, determining protein-folding dynamics and stability utilizing protein prediction programs might offer insight as to the function of the C-terminus. First, we analyzed previously published work on X-ray crystallography of GATA3 to further our understanding of how the DBD might interact with other GATA3 motifs. To that end, protein motifs in their native state that fold into 3-D structures can be crystallized. The X-ray structure of these motifs are characterized by their folding patterns, leading to models that predict overall rigidity, movement, and structure. This is particularly useful for discovering known domains and also areas of the protein that cannot be crystallized. The GATA3 C-finger domain (amino acids 307–370) has been successfully crystallized (Bates et al., 2008). These data illustrate the engagement

of the zinc finger domains along with other transcription factors, leading to self-association. In addition to DNA binding and interactions with other proteins, the ability to self-associate has been implicated in the assembly of complex protein/DNA complexes within areas of gene activity in other GATA family members as well (Crossley et al., 1995). In contrast, the C-terminus has not been successfully crystalized, likely due to its lack of electron density. In these instances, X-ray crystallography and NMR are not effective tools in visualizing structure. Structures not likely to be detected by either method are usually either disordered or extremely soluble, or contain several charged residues (charge imbalance) (Yang et al., 2005).

Disordered regions are regions without a tertiary structure, whose backbones tend to resemble either an expanded random coil or collapsed molten globular configuration. Often, proteins with disordered regions undergo a disorder-to-order transition upon binding to protein or nucleic acids (Dyson and Wright, 2002).

To help with structural predictions, we used an online program, regional order neural network (RONN, http://www.strubi.ox.ac.uk/RONN), to help analyze the C-terminus since it is a region lacking 3-D structure. Analysis of GATA3 with RONN suggests that the extreme C-terminal region is highly disordered, similar to the linker between the two zinc fingers (Figure 3.9). Likewise, there are other predicted disordered regions in GATA3 between the zinc fingers and within the N-tail. These regions are likely disordered in order to ensure flexible DNA recognition either in *cis* and/or *trans*, and they may undergo a disorder-to-order transition once contact is made with either neighboring TFs or DNA. Closer analysis of the C-terminal mutations in GATA3, particularly at residue 431, where a frame-shift mutation results in an elongated GATA3 (509 aa), leads to a partial loss of the disordered C-terminus region (Figure 3.10). As a result, DNA binding stability may not occur, interrupting GATA3 binding and

downstream signaling. Plans for developing GATA3 mutant expression constructs are currently ongoing. Future approaches that will be accomplished by others will confirm protein expression using 293T cells, a GATA3-negative cell line. In addition, GATA3 binding will need to be assessed with ChIP-seq at motifs important for GATA3 signaling in breast cancer. Finally, RNA should be isolated from cells overexpressing these mutants in a GATA3-deficient immortalized breast cell line (BABE) and hybridize to expression microarrays to determine changes GATA3 downstream signaling, as well as d-score values.

These studies will help shed light on the impact of GATA3 mutations in the most commonly diagnosed breast cancer subtype, the luminal subtype. It has been understood that GATA3 plays a critical role in the normal development of the breast, and as a biomarker, this TF is significantly associated with ESR1-positive luminal breast tumors, which generally have a better prognosis than the other subtypes. Although the frequency of mutations in GATA3 seem to be enriched in luminal tumors, the consequences of such mutations are not well understood. These studies were initiated to elucidate the downstream effects of mutations in 3 key areas. This work will lay the groundwork for others in the lab to not only determine the role of the C-terminal tail when its structure is altered by a mutational frame shift but also to better explain the role of GATA3 aberrations in luminal tumors.

Figure 3.1. GATA3 Protein Domains and Functions

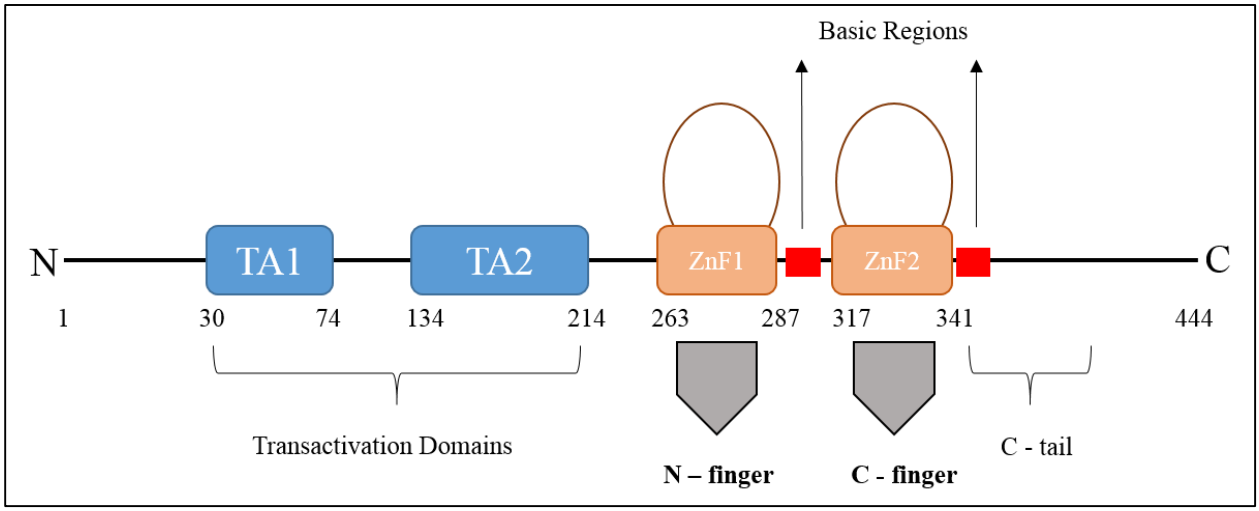


Figure 3.2. GATA3 Open Reading Frame and Lentiviral Plasmid Development

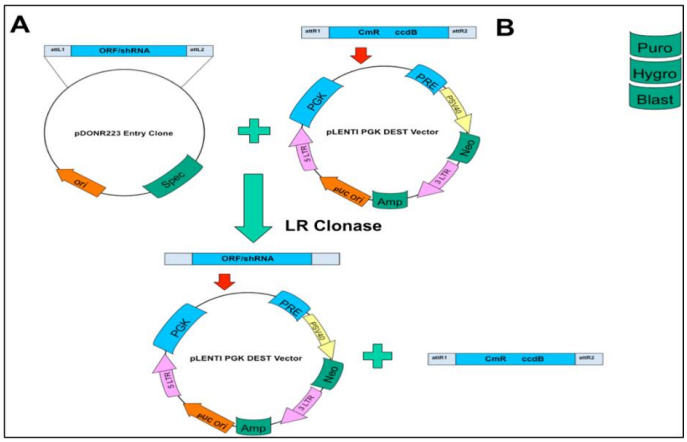
GATA3 open reading frame

NCBI accession # BC006793

GenBank Coding Sequence:

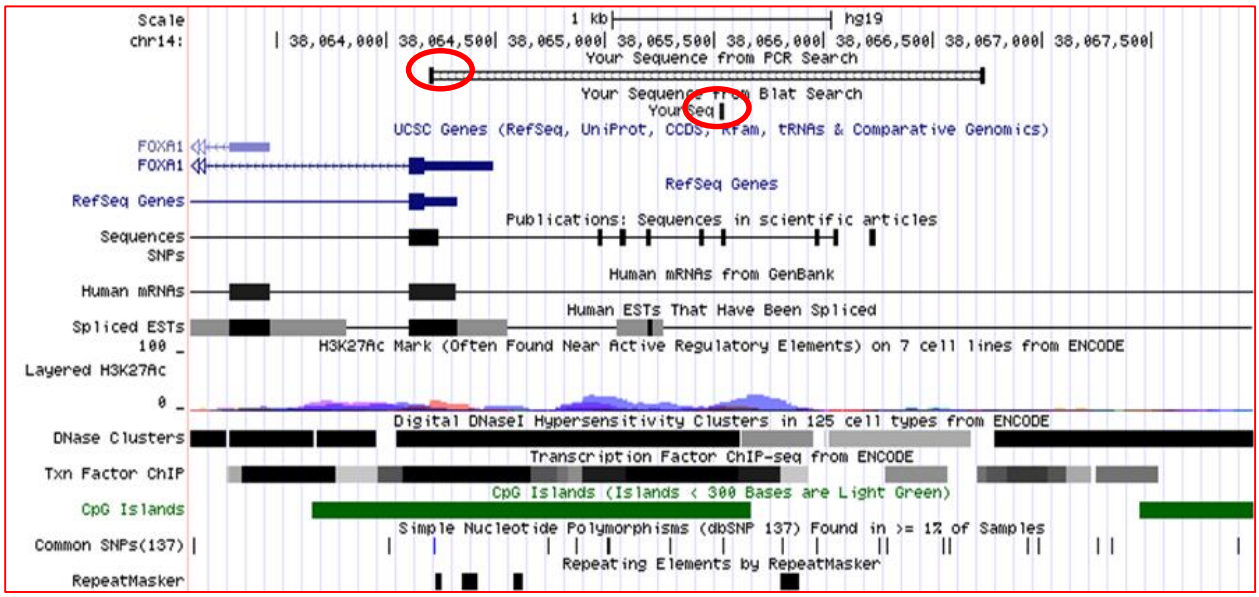
```
atggaggtgacggcggaccagccgccgctgggtgagccaccaccacccccgccgtgctcaacgggcagcacccggacacgcaccaccccgggcctcagccact
cctacatggacgcggcgcagtacccgctgccggaggaggtggatgtgtcttttaacatcgacggtcaaggcaaccacgtcccgccctactacggaaactc
ggtcagggccacggtgcagaggtaccctccgacccaccacgggagccaggtgtgccgcccgcctctgcttcatggatccctcaccctggctggacggcggc
aaagccctgggcagccaccacaccgcctccccctgggaatctcagccccttctccaagacgtccatccaccacgggctccccgggggccccctctccgtctacc
ccccggcctcgtcctcctccttgtcgggggggccacgccagcccgcacctcttcaccttcaccgtcccgcaccccgccgaaggacgtctccccggaccccatcgct
gtccaccccaggctcggccggctcggccccggcaggacgagaaagagtgcctcaagtaccaggtgcccctgcccgacagcatgaagctggagtcgtcccac
tccgtggcagcatgaccgccctgggtggagcctcctcgtcgaccccaccaccccatcaccacctaccgccctacgtgccccgagtacagctccggcactct
tcccccccagcagcctgctgggcggctccccccaccggcttcggatgcaagtccaggcccaaggcccggtccagcacagaaggcagggagtgtgtgaactg
tgggggcaacctcgaccccactgtggcggcgagatggcacgggacactacctgtgcaacgcctgcgggctctatcacaaaatgaacggacagaaccggccc
ctcattaagcccaagcgaaggctgtctgcagccaggagagcagggacgtcctgtgcgaactgtcagaccacacacaccacactctggaggaggaatgcca
atggggaccctgtctgcaatgcctgtgggctctactacaagcttcacaatattaacagacccctgactatgaagaaggaaggcatccagaccagaaaccg
aaaaatgtctagcaaatccaaaaagtgcaaaaaagtgcatgactcactggaggacttccccaagaacagctcgtttaaccccggccgcccctctccagacac
atgtcctcccctgagccacatctcgccctttcagccactccagccacatgctgaccacgcccacgccgatgcaccccgccatccagcctgtcctttggaccac
accacccctccagcatggtcaccgccatgggttag
```

GATA3 Lentiviral Plasmid Development



(A) pDONR223 plasmid with GATA3 ORF. **(B)** pLENTI PGK DEST vector where directional recombination with LR clonase will transfer GATA3 ORF from one vector to the other.

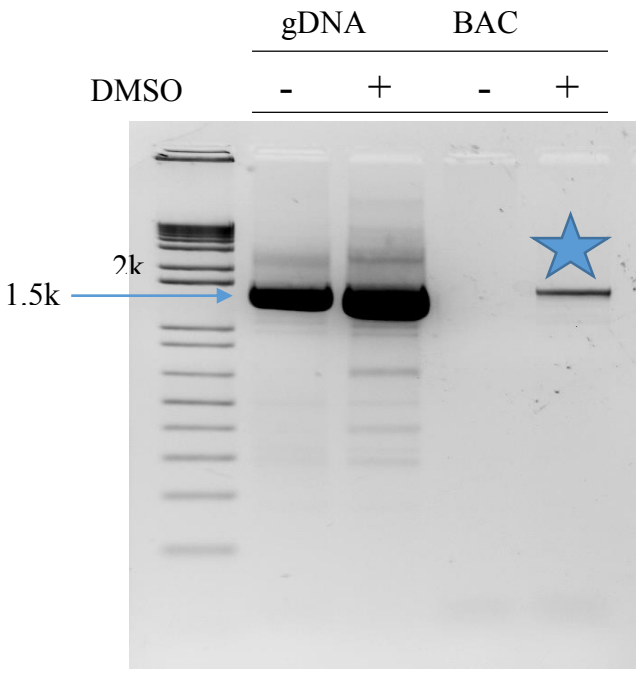
Figure 3.3. Genomic Location of FOXA1-Reporter Sequence



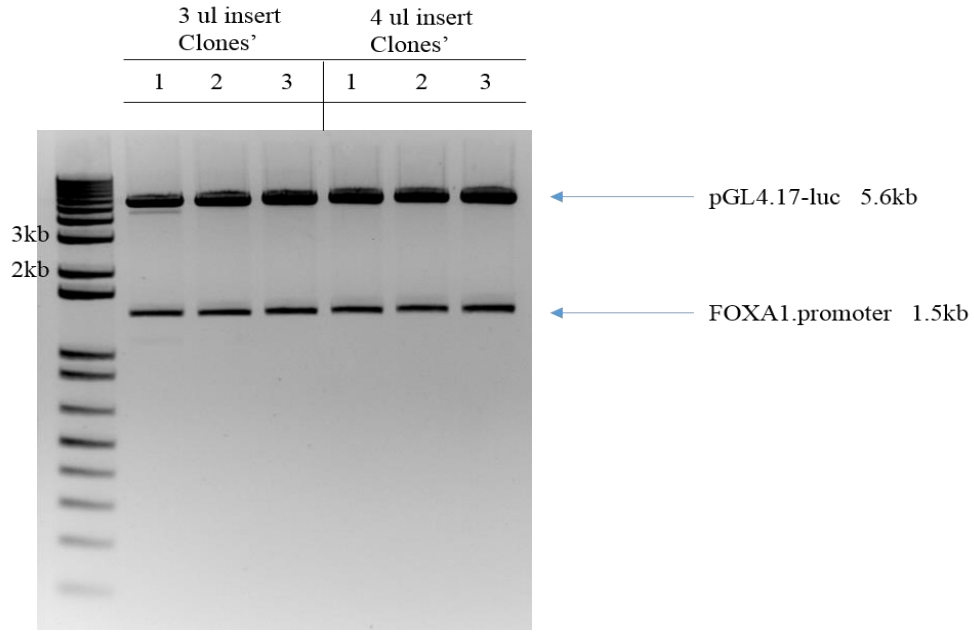
Note. Circles in red denote the forward and reverse primer locations. Sequence analysis and graphic from UCSC genome browser (https://genome.ucsc.edu/).

Figure 3.4. Cloning Result of FOXA1 Genomic DNA Sequence and FOXA1-luc Promoter Double Digest

A



B



Note. Blue star denotes PCR product used for pGL4-luc cloning. A total of 6 clones contain the appropriate direction and size of the FOXA1 promoter sequence.

Figure 3.5. Normalized FOXA1 – Reporter Activity in 293T Cells

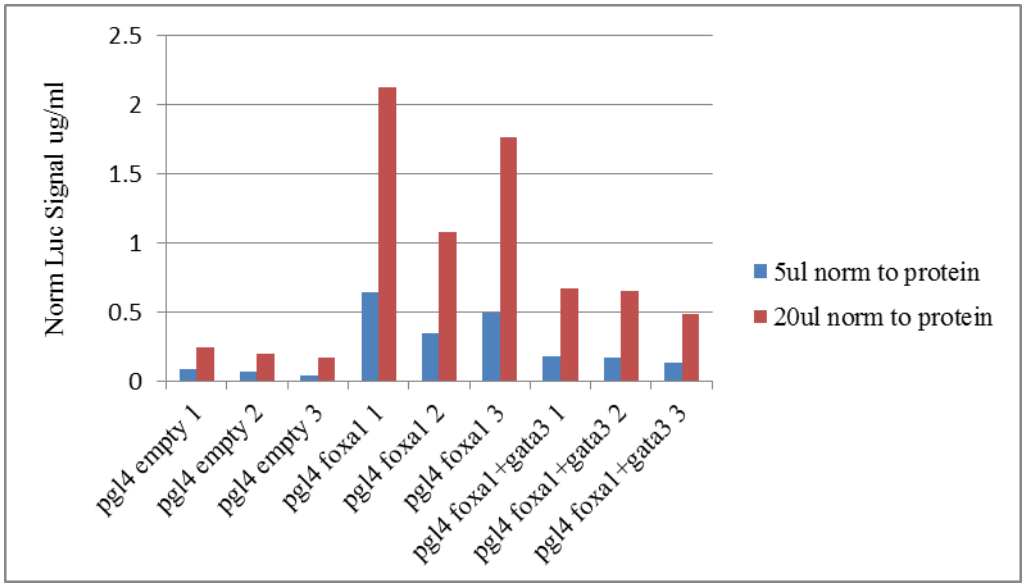
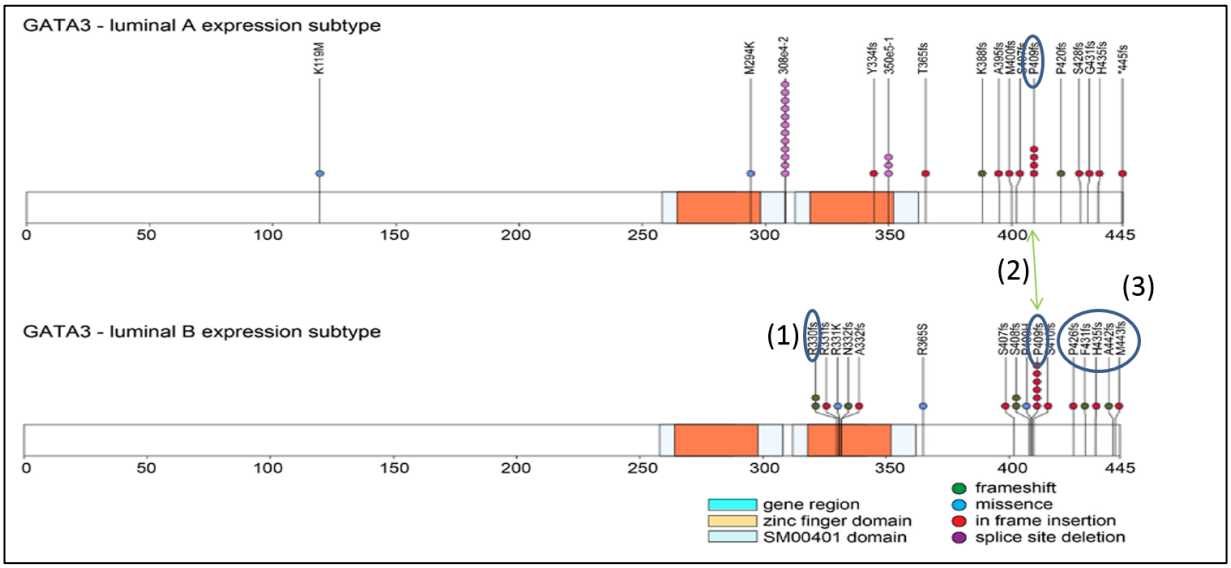


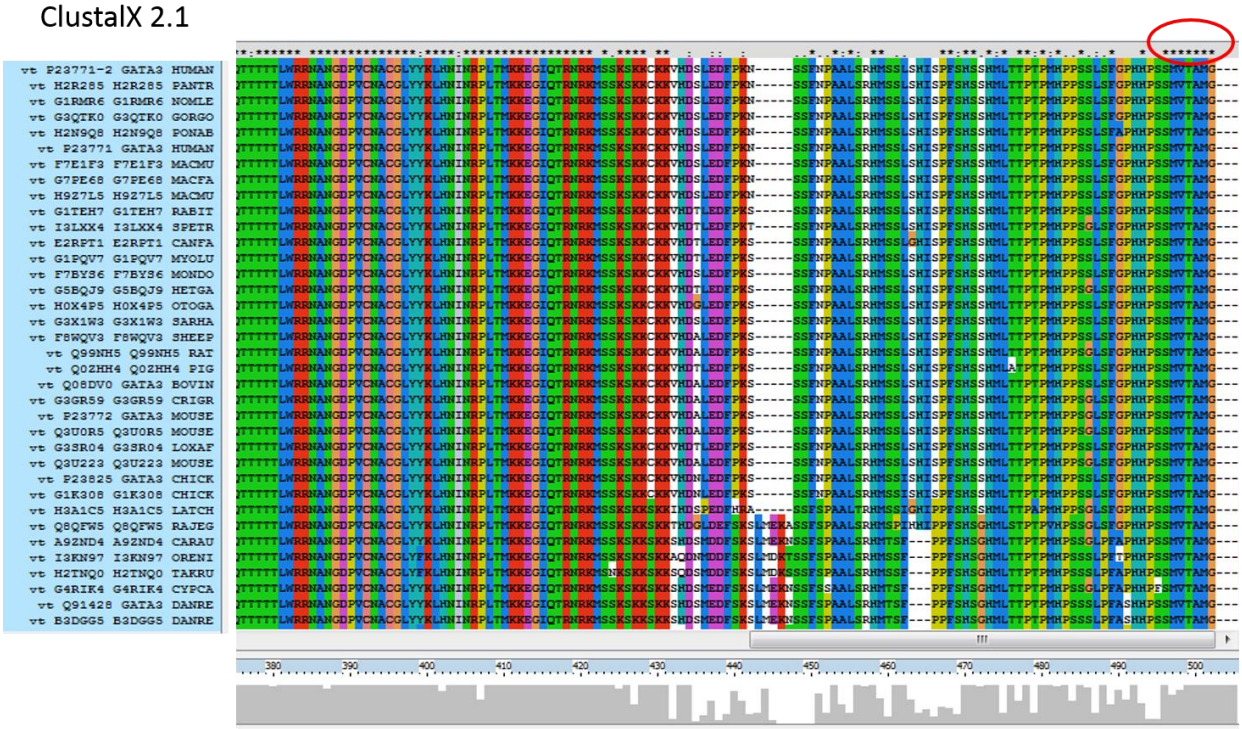
Figure 3.6. C–Terminus GATA3 Mutations in TCGA Data Set



Adapted from D. C. Koboldt et al., "Comprehensive Molecular Portraits of Human Breast Tumors," 2012, *Nature* 490(7418): 61–70.

Note. 1–3 are mutations that will be incorporated into mutant GATA3 constructs that cover the C – terminus. Note that some mutations are found in both luminal A and B subtypes.

Figure 3.7. GATA3 C–Terminus: A Highly Conserved Region across Several Vertebrates



AA 261-315 Proteasome degradation
AA 345-354 YxKxHxxxRP
AA 396-422 Lysine ubiquitination sites

Figure 3.8. GATA3 Mutation Consequences

1) **R330fs – insertion A**

MEVTADQPRWVSHHHPAVLNGQHPDTHHPGLSHSYMDAAQYPLPEEVDVLFNIDGQGNHVPPYYGNSVRATVQRYPPTHHGSQVCRPPLLHGSLPWL
DGGKALGSHHTASPWNLSPFSKTSIHHGSPGPLSVYPPASSSSLSGGHASPHLFTFPPTPPKDVSPDPSLSTPGSAGSARQDEKECLKYQVPLPDSMKLESS
HSRGSMTALGGASSSTHHPITTYPPYVPEYSSGLFPPSSLLGGSPTGFGCKSRPKARSSTEGRECVNCGATSTPLWRRDGTGHYLCNACGLYHKMNGQNR
PLIKPKRRLSAARRAGTSCANCQTTTTTLWRKECQWGPCLQCLWALLQASQY* 351AA

- **R330fs – deletion GA 349AA**
- **R331fs – insertion CTGGAGG 353AA**
- **R332fs – deletion AT**

MEVTADQPRWVSHHHPAVLNGQHPDTHHPGLSHSYMDAAQYPLPEEVDVLFNIDGQGNHVPPYYGNSVRATVQRYPPTHHGSQVCRPPLLHGSLPWL
DGGKALGSHHTASPWNLSPFSKTSIHHGSPGPLSVYPPASSSSLSGGHASPHLFTFPPTPPKDVSPDPSLSTPGSAGSARQDEKECLKYQVPLPDSMKLESS
HSRGSMTALGGASSSTHHPITTYPPYVPEYSSGLFPPSSLLGGSPTGFGCKSRPKARSSTEGRECVNCGATSTPLWRRDGTGHYLCNACGLYHKMNGQNR
PLIKPKRRLSAARRAGTSCANCQTTTTTLWRRSQWGPCLQCLWALLQASQY*QTPDYEEGRHPDQKPKNV* **369AA**

2) **P409fs – insertion G**

MEVTADQPRWVSHHHPAVLNGQHPDTHHPGLSHSYMDAAQYPLPEEVDVLFNIDGQGNHVPPYYGNSVRATVQRYPPTHHGSQVCRPPLLHGSLPWL
DGGKALGSHHTASPWNLSPFSKTSIHHGSPGPLSVYPPASSSSLSGGHASPHLFTFPPTPPKDVSPDPSLSTPGSAGSARQDEKECLKYQVPLPDSMKLESS
HSRGSMTALGGASSSTHHPITTYPPYVPEYSSGLFPPSSLLGGSPTGFGCKSRPKARSSTEGRECVNCGATSTPLWRRDGTGHYLCNACGLYHKMNGQNR
PLIKPKRRLSAARRAGTSCANCQTTTTTLWRRNANGDPVCNACGLYYKLHNINRPLTMKKEGIQTRNRKMSSKSKKCKKVHDSLEDFPKNSSFNPAALS
RHMSSLSHISALQPLQPHADHAHADAPAIQPVLWTTPPLQHGHRHGLEPCSMLTGPPARVPAVPFDLHFCRSSIMKPKRDGYMFLKAESKIMFATLQRSS
LWCLCSNH* **506AA**

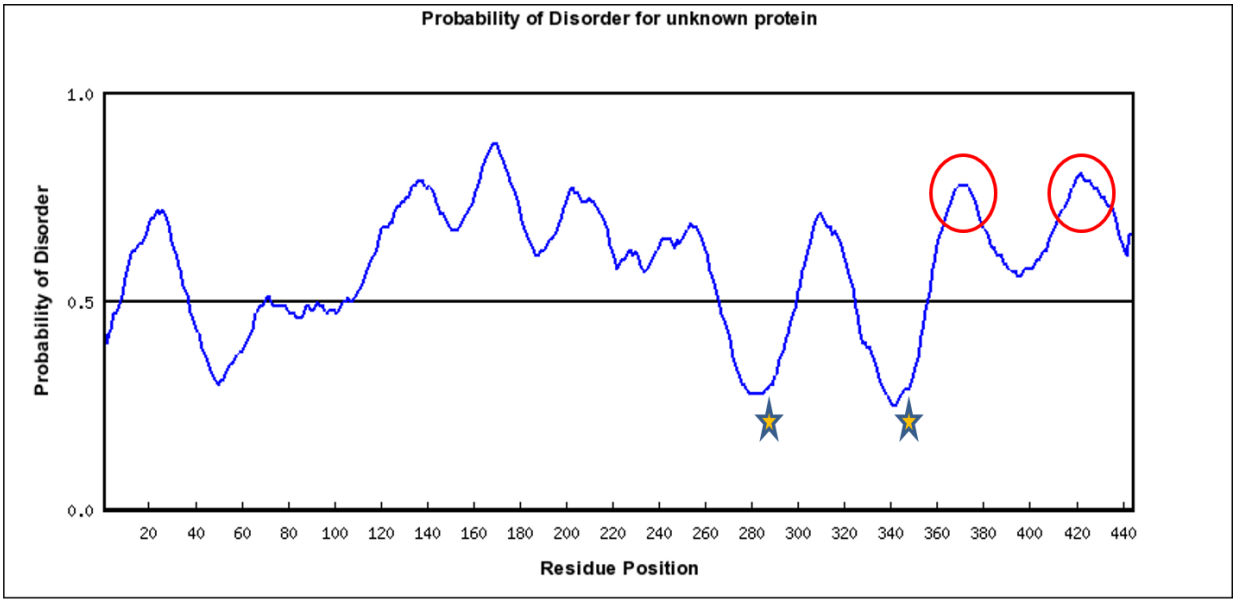
3) **F431fs – TT deletion**

MEVTADQPRWVSHHHPAVLNGQHPDTHHPGLSHSYMDAAQYPLPEEVDVLFNIDGQGNHVPPYYGNSVRATVQRYPPTHHGSQVCRPPLLHGSLPWL
DGGKALGSHHTASPWNLSPFSKTSIHHGSPGPLSVYPPASSSSLSGGHASPHLFTFPPTPPKDVSPDPSLSTPGSAGSARQDEKECLKYQVPLPDSMKLESS
HSRGSMTALGGASSSTHHPITTYPPYVPEYSSGLFPPSSLLGGSPTGFGCKSRPKARSSTEGRECVNCGATSTPLWRRDGTGHYLCNACGLYHKMNGQNR
PLIKPKRRLSAARRAGTSCANCQTTTTTLWRRNANGDPVCNACGLYYKLHNINRPLTMKKEGIQTRNRKMSSKSKKCKKVHDSLEDFPKNSSFNPAALS
RHMSSLSHISPFSHSSHMLTTPTPMHPPSSLSWTTPPLQHGHRHGLEPCSMLTGPPARVPAVPFDLHFCRSSIMKPKRDGYMFLKAESKIMFATLQRSSLW
CLCSNH* **506AA**

- **G431fs – insertion T**

MEVTADQPRWVSHHHPAVLNGQHPDTHHPGLSHSYMDAAQYPLPEEVDVLFNIDGQGNHVPPYYGNSVRATVQRYPPTHHGSQVCRPPLLHGSLPW
LDGGKALGSHHTASPWNLSPFSKTSIHHGSPGPLSVYPPASSSSLSGGHASPHLFTFPPTPPKDVSPDPSLSTPGSAGSARQDEKECLKYQVPLPDSMKLES
SHSRGSMTALGGASSSTHHPITTYPPYVPEYSSGLFPPSSLLGGSPTGFGCKSRPKARSSTEGRECVNCGATSTPLWRRDGTGHYLCNACGLYHKMNGQ
NRPLIKPKRRLSAARRAGTSCANCQTTTTTLWRRNANGDPVCNACGLYYKLHNINRPLTMKKEGIQTRNRKMSSKSKKCKKVHDSLEDFPKNSSFNPA
ALSRHMSSLSHISPFSHSSHMLTTPTPMHPPSSLSFWTTPPLQHGHRHGLEPCSMLTGPPARVPAVPFDLHFCRSSIMKPKRDGYMFLKAESKIMFATLQR
SSLWCLCSNH* **506AA**

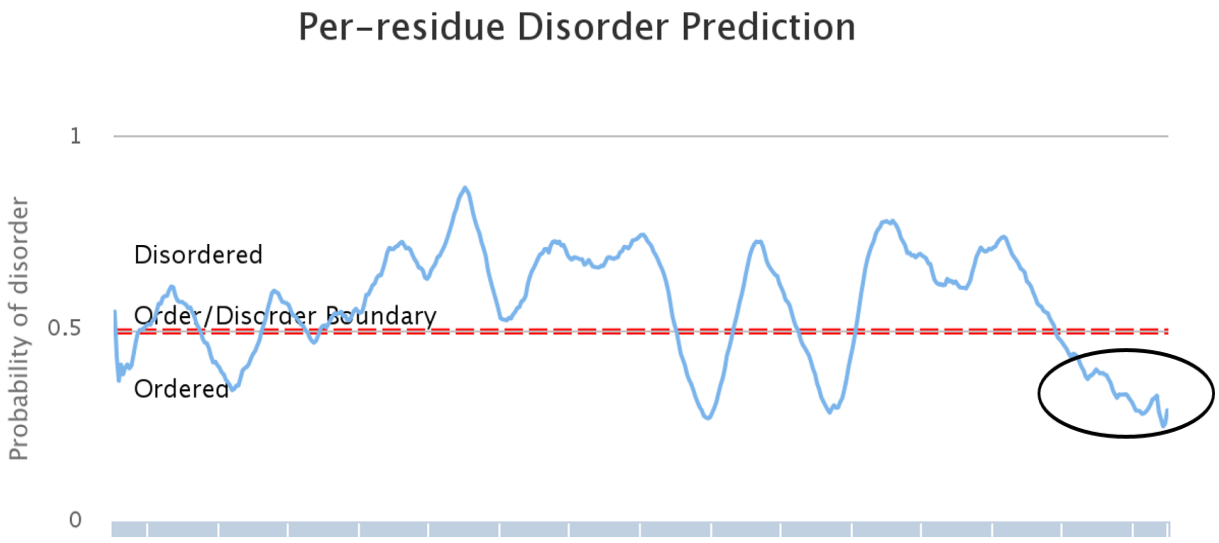
Figure 3.9. Predicted Disorder Regions for Wild-Type GATA3



Disordered regions:
9 - 37, 71 - 72, 106 - 106, 109 - 265, 300 - 324, 357 - 444

Zinc finger domains' 1 and 2

Figure 3.10. Predicted Disorder Regions for Mut GATA3 (F431—Frame Shift)



CHAPTER 4: DISCUSSION

Online media mogul man Tim O'Reilly once said, "An invention has to make sense in the world it finishes in, not in the world it started." For Tim, being a catalyst means creating leading-edge Internet tools to further drive innovation. A key example of this is the advent of the open source movement, through which Tim spurred the adoption of publicly available tools on the Internet. Internet companies that survived the dot-com bust in the early 2000s were limited financially and felt disenfranchised by the downfall of the Internet renaissance. During this time, creative leaders in the industry came together and advocated for the acceptance of open source tools that would ultimately shift the Internet business model from one of software to one of data. As a result, Web 2.0 was developed, which brought about inventions such as Google, mobile applications, and coding platforms, just to name a few. The goal for Tim and others was to create tools available to all, to be used in the creation of new technologies, and to breathe life into existing technologies that would otherwise become irrelevant. Many academic scientists will point out that open source platforms for data analysis have made scientific publishing of big data an affordable reality, allowing for the exchange and independent reproduction of analysis methods. Such exchanges have created an environment of innovation by affording open source tools in solving complex biological questions.

In the mid 2000s, stem cell scientists successfully reprogrammed somatic cells into fully functional induced pluripotent cells (iPSC). The creation of this tool was a seminal leap in cell biology for several reasons. Basic questions about cell fate determination and differentiation programming had persisted within the community for decades despite various technical advances

that afforded better methods for answering these questions. From a practical standpoint, by the turn of the century a hostile environment had developed toward the scientific usage of embryonic stem cells. On August 9, 2001, then-president George Bush Jr. signed into law a ban on federal funds for embryonic stem cell research (Sandel, 2004). Scientists were also examining the research utility of adult stem cells, a much less contentious stem cell type, during this time. Due to their scarcity in tissues and lack of pluripotency, adult stem cells were deemed poor substitutes for embryonic stem cell research. Hence, when the Yamanaka and Thompson labs directly reprogrammed mouse and human somatic cells into functional iPSCs via the induction of four defined transcription factors, they not only circumvented the political issues surrounding embryonic research but also demonstrated the molecular feasibility of transforming one cell type into another. Reprogramming techniques employed by both labs are used today by researchers, including ones who have successfully published on the creation of a variety of tissues from unrelated cell types.

Our goal in the present study was different from those in the pivotal iPSC studies. Instead of developing a dedifferentiated cell line from a somatic cell, we chose to start with a differentiated somatic cell line and by introducing specific transcription factors known to be important in breast development, to reprogram these cells directly into functional luminal-like breast cells. Normal luminal breast cells are differentiated cells with a low proliferative capacity and, like their luminal A tumor counterpart, do not grow in culture. To isolate these luminal epithelial cells for research studies, primary cell populations derived from normal breast tissue must be labeled for luminal markers and must be sorted using FACS; otherwise, the highly proliferative basal cells will rapidly outgrow the luminal population. The isolated luminal cell population does not propagate under culture conditions, so immediate use of these cells for

downstream applications is necessary. Working with primary cells can be a serious drawback for this reason.

Others in the reprogramming field have chosen to work with primary cells that have undergone immortalization in order to overcome growth limitations in vitro (Huang et al., 2011; Ieda et al., 2010; Vierbuchen et al., 2010). We went forward with both a mouse embryonic fibroblast (MEF) and a normal human breast cell line (BABE) immortalized with inactivated *Cdkn2a* and by overexpression of *hTERT*, respectively. Both immortalization methods have demonstrated a lack of tumorigenicity, stable genome, and a functioning TP53/RB pathway, attributes that are critical for maintaining the fidelity of the primary cell line (Dove, 2015; Huang et al., 2011; Toouli et al., 2002).

Initially, we hypothesized that the BABE cell line would be easier to transdifferentiate, in part because the BABE cell line is already a basal breast epithelial cell line. We expected that transdifferentiation to a luminal line would require relatively few changes, as opposed to starting with a cell line from a different lineage. While transductions with ESR1, FOXA1, GATA3, and/or PGR resulted in an increase in differentiation score (d-score) in the BABE cell line, the combinations with ESR1 triggered senescence, inherently limiting their use for further studies. TP53 and Rb are intact and functional in *hTERT*-immortalized cells, providing a plausible explanation for their growth restrictions. There are other available immortalized breast epithelial cell lines; one that is used more frequently is the HMLE cell line, which is immortalized via SV40 viral transduction of the large T antigen. HMLEs have a less differentiated phenotype and a lower d-score than the BABEs. However, viral inactivation of TP53 and Rb function in HMLEs confers tumorigenic potential (Dimri et al., 2005; Schmeichel and Bissell, 2003) making them an impractical cell line to work with.

Furthermore, our laboratory has shown that TFs can have varying effects on differentiation parameters between different mammary epithelial cell lines. In these experiments, either SNAI1 or TWIST (TFs involved in induction of EMT) were transduced into BABEs. Array analysis showed a higher degree of similarity to the parent BABE cell line than to less-differentiated mammary cell lines (ME16C for example, data not shown). A similar experiment was also performed in HMLE cells; however, after array analysis, SNAI1 and TWIST-induced cells expression profiles were most similar with the claudin-low tumors (Mani et al., 2008). These data underscore the serious considerations around the selection of a starting cell line.

Contrary to our initial expectations, we found that TF-induced mammary differentiation could be more robustly demonstrated in a cell line from a different species and lineage, mouse embryonic fibroblasts (MEFs). Indeed, these cells have a healthy publication record when it comes to the field of reprogramming. Initially used as feeder cells for the maintenance of stem cells, MEFs were also used in pioneering demonstrations of direct transdifferention (Huang et al., 2011; Vierbuchen et al., 2010). In the present study, we utilized *cdkn2a*$^{-/-}$ MEFs from a C57BL/6J mouse strain. These immortalized MEFs are replication competent in appropriate growth media. Our studies of the MEF cells, together with those of the BABEs, indicated that the direction and degree of differentiation changes depend on the cell environment (e.g., culture medium, growth factors, and hormones), interaction of cell line (cell context), and transcription factor combinations. These interactions are discussed in detail below.

When devising the protocol for viral infection and media selection, there were considerations to be addressed. Earlier viral experiments with the BABE cell line were straightforward, as HuMEC media used for infection were swapped out 48 hours later for virus-free media. In essence, we anticipated transdifferentiation would occur in media that are

normally used for differentiated luminal cells. However, there are no known normal mammary luminal cell lines, as BABE, HMLE, and ME16C cell lines are basal-like and are grown in serum-free HuMEC media. Keeping with the idea of HuMEC media as the most appropriate for the transdifferentiated MEFs, MEFs were infected with a virus in RPMI media with 10% FBS and then switched to serum-free HuMEC 48 hours post infection. The relative contribution to differentiation, as defined by the d-score, was assessed in serum-starved MEFs. Remarkably, serum-starved MEFs become more differentiated by d-score calculations than their parental counterpart (data not shown). The alternative question to address was determining the effect of serum undergoing transdifferentiation. To answer this, we compared the d-score between MEFs induced with –EFP in the presence and absence of serum (data not shown). Strikingly, the M–EFP cell line in serum had a significantly lower d-score than the serum-free M–EFP cell line, suggesting that sera can negatively affect differentiation. This led us to the hypothesis that transitioning the induced MEFs in HuMEC media will facilitate differentiation in addition to the contributions of the transcription factors.

Limitations with ectopic expression of multiple genes have been reported. While the selection of tissue-specific TF combinations are extremely important for successful transdifferentiation, the occurrence of such an event is extremely low (Brambrink et al., 2008; Desponts and Ding, 2010). The induction of each cell with each of the three viruses is not 100%. Moreover, the addition of serum-free conditions create a stressful environment, especially for cells lacking one or more of the TFs. In our study, serum starvation killed about 80% of the induced MEFs due to the reasons mentioned. For the 20% remaining, morphological changes became evident after 3–4 weeks, which is in contrast to the 1–2 week epithelial colony confirmation seen in other studies (Huang et al., 2011; Vierbuchen et al., 2010). With the lack of

a reporter or antibiotics for selection, we confirmed transcript and protein presence for each TF. After testing multiple combinations of TFs, we selected M–EFP based on morphology, d-score, and growth characteristics.

The addition of hormones further enhanced the differentiation potential of the M–EFP cells. Although we found the treatment of MEFs with hormones did not result in any d-score change as compared to no treatment, the use of ligands, drugs, or other chemicals can be used synergistically in combination with the induction of breast-specific TFs to maximize differentiation potential (Brisken and O'Malley, 2010; Bussard et al., 2010; Hou et al., 2013).

We revealed in Chapter 2 that the M–EFP cells exhibit classic epithelial structure and expression of markers. The structural rearrangement of actin into a cortical formation around the membrane is a hallmark feature of epithelial cells and also associated with CDH1 expression (Yap et al., 2014). We demonstrated the rearrangement of F-actin and confirmed the expression of CDH1 in the M–EFP cell line. The presence of breast markers supported the selection of TFs as ones that are involved in regulating a breast-specific phenotype. Furthermore, we demonstrated their gene expression profile is more like that of mouse mammary cell lines HC11 and PyMT than of MEFs. Finally, we showed that the M–EFPs behave like a luminal breast cell, exhibiting sensitivity to an estrogen agonist/antagonist and secreting the milk protein beta-casein when stimulated by the hormone prolactin.

While we confirmed an enrichment of several transcription factor networks involved in breast epithelial functional signaling, such as STAT5A and C/EBPβ, much work is still needed in understanding how ESR1, FOXA1, and PGR have reprogrammed MEFs. ChIP-seq experiments would help determine the specific promoter binding sites –EFP binds to. FOXA1 has been shown to largely regulate where ESR1 binds to in luminal breast cancer cell lines.

Without the expression of FOXA1, ESR1 binds to different promoter regions (Carroll et al., 2005). Moreover, how will silencing these genes individually impact the overall phenotype, if at all?

The most difficult aspect in the creation of the artificially engineered M–EFP is its developmental classification compared to natural cells. It is the general consensus that in the mouse, a mouse mammary repopulating breast cell has low to moderate CD24 and high CD29 levels (Mark et al., 2006; Shackleton, 2006; Stingl et al., 2006). Likewise, progenitor cells, which some call the bipotent population, are high for CD24 and moderate for CD29. They also have expression of CD61, a marker unique to the progenitor and absent in the stem cell and luminal subpopulations. The progenitor population gives rise to the basal and luminal lineages, thus expressing basal cytokeratin 5 and/or luminal cytokeratin 8/18 (Prat and Perou, 2011; Rios et al., 2014). Our cells exhibited positive staining for both, though only about 15% of the M–EFP was CD61 positive. On the d-score axis MaSC → pL, M–EFP cells are the same as normal mammary cells, implying that they exhibit both basal and luminal features, and so they resemble a developmental progenitor cell line. However, this would need to be determined by syngeneically transplanting GFP-labeled M–EFP cells in the fat pad of a mouse, likely at different developmental time points, where hormones are present/absent.

There is lacking a luminal A cell line, which makes studying the most diagnosed breast cancer subtype in the lab arduous. While the M–EFP cell line is not considered a luminal A cell line, it has several important features in common with normal luminal cells. As an example, ESR1 positive breast tumors frequently become resistant to antiestrogen therapy. In the mouse, mammary development and carcinogenesis are less dependent on ESR1 signaling and thus are more difficult to study in mouse luminal tumor models. The PyMT cell line has been derived

from a PyMT-*neu* mouse tumor and expresses low levels of ESR1. Our cell line would be extremely useful in understanding ESR1-mediated events during mammary normal development and tumorigenesis. Understanding which genes these TFs regulate may also shed light on what is missing in terms of terminal differentiation. While the addition of hormones increased the differentiation of the M–EFP cells, there may be other TFs necessary to complete the process.

In human breast development, GATA3 has been shown to be indispensable for alveologenesis and strongly correlates with ESR1 expression (Chou et al., 2010; Wilson and Giguere, 2008). Results from our initial TF screen demonstrated expressing GATA3 has different effects in the BABE and MEF cell lines. We observed a decrease in d-score with GATA3 in the MEFS and therefore were interested in determining how GATA3 function is affected by the cluster of mutations reported in luminal A and B tumors. GATA3 binding to DNA is altered by mutations in the N and C finger domains and tails, while the effects of mutations in the C-terminus have yet to be studied. These are often frame-shift mutations that would produce a C-terminal truncated protein. GATA3 mutations might be predicted to affect transcriptional regulation by other genes as well: GATA3 functions as a coregulator with ESR1, influencing the expression of ESR1-specific genes. Deciphering the relationship between ESR1 signaling and GATA3 mutations will be critical in determining the role that GATA3 mutations play in ESR1-positive tumors.

We anticipate that the creation of the M–EFP cell line will be a valuable tool in studying the progression of basal-to-luminal differentiation in the breast. Since the present study represents the first published work in this area, we hope that our technical approaches will assist others wishing to perform future transdifferentiation studies in breast and other tissue types. The

present work identified three TFs that together were able to induce critical aspects of a breast epithelial phenotype. Other TFs such as GATA3 may have additional roles in the differentiation process that have yet to be defined.

REFERENCES

Aitola M, Carlsson P, Mahlapuu M, Enerback S and Pelto-Huikko M (2000) Forkhead transcription factorFoxF2 is expressed in mesodermal tissues involved in epithelio-mesenchymal interactions. *Dev Dyn* **218**: 136-149.

Akslen WDF, Ingunn MS, Pierre OC, Louis RB, John RG, Nora W, Michel T and Lars A (2003) Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst* **95**(19): 1482-1485.

Ali A, Christie PT, Grigorieva IV, Harding B, Van Esch H, Ahmed SF, Bitner-Glindzicz M, Blind E, Bloch C, Christin P, Clayton P, Gecz J, Gilbert-Dussardier B, Guillen-Navarro E, Hackett A, Halac I, Hendy GN, Lalloo F, Mache CJ, Mughal Z, Ong ACM, Rinat C, Shaw N, Smithson SF, Tolmie J, Weill J, Nesbit MA and Thakker RV (2007) Functional characterization of GATA3 mutations causing the hypoparathyroidism-deafness-renal (HDR) dysplasia syndrome: insight into mechanisms of DNA binding by the GATA3 transcription factor. *Hum Mol Genet* **16**(3): 265-275.

Arnold JM, Choong DY, Thompson ER, Waddell N, Lindeman GJ, Visvader JE, Campbell IG and Chenevix-Trench G (2010) Frequent somatic mutations of GATA3 in non-BRCA1/BRCA2 familial breast tumors, but not in BRCA1-, BRCA2- or sporadic breast tumors. *Breast Cancer Res Treat* **119**(2): 491-496.

Asselin-Labat M-L, Sutherland KD, Barker H, Thomas R, Shackleton M, Forrest NC, Hartley L, Robb L, Grosveld FG, van der Wees J, Lindeman GJ and Visvader JE (2007) Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* **9**(2): 201-209.

Asselin-Labat M-L, Vaillant F, Sheridan JM, Pal B, Wu D, Simpson ER, Yasuda H, Smyth GK, Martin TJ, Lindeman GJ and Visvader JE (2010) Control of mammary stem cell function by steroid hormone signalling. *Nature* **465**(7299): 798-802.

Bachelot A and Binart N (2007) Reproductive role of prolactin. *Reproduction* **133**(2): 361-369.

Balko JM, Miller TW, Morrison MM, Hutchinson K, Young C, Rinehart C, Sanchez V, Jee D, Polyak K, Prat A, Perou CM, Arteaga CL and Cook RS (2012) The receptor tyrosine kinase ErbB3 maintains the balance between luminal and basal breast epithelium. *Proc Natl Acad Sci* **109**: 221-226.

Bates DL, Chen Y, Kim G, Guo L and Chen L (2008) Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *J Mol Biol* **381**(5): 1292-1306.

Beleut M, Rajaram RD, Caikovski M, Ayyanan A, Germano D, Choi Y, Schneider P, and Brisken C (2010) Two distinct mechanisms underlie progesterone-induced proliferation in the mammary gland. *Proc Natl Acad Sci USA* **107**(7):2989-2994.

Bernardo GM, Lozada KL, Miedler JD, Harburg G, Hewitt SC, Mosley JD, Godwin AK, Korach KS, Visvader JE, Kaestner KH, Abdul-Karim FW, Montano MM and Keri RA (2010) FOXA1 is an essential determinant of ER expression and mammary ductal morphogenesis. *Development* **137**: 2045-2054.

Berthois Y, Katzenellenbogen JA and Katzenellenbogen BS (1986) Phenol red in tissue culture media is a weak estrogen: implications concerning the study of estrogen-responsive cells in culture. *Proc Natl Acad Sci* **83**: 2496-2500.

Bertucci F, Finetti P and Birnbaum D (2012) Basal breast cancer: a complex and deadly molecular subtype. *Curr Mol Med* **12**(1): 96-110.

Binart N, Bachelot A and Bouilly J (2010) Impact of prolactin receptor isoforms on reproduction. *Trends Endocrinol Metab* **21**: 362-368.

Binart N, Imbert-Bolloré P, Baran N, Viglietta C and Kelly PA (2003) A short form of the prolactin (PRL) receptor is able to rescue mammopoiesis in heterozygous PRL receptor mice. *Mol Endocrinol* **17**: 1066-1074.

Bloushtain-Qimron N, Yao J, Snyder EL, Shipitsin M, Campbell LL, Mani SA, Hu M, Chen H, Ustyansky V, Antosiewicz JE, Argani P, Halushka MK, Thomson JA, Pharoah P, Porgador A, Sukumar S, Parsons R, Richardson AL, Stampfer MR, Gelman RS, Nikolskaya T, Nikolsky Y and Polyak K (2008) Cell type-specific DNA methylation patterns in the human breast. *Proc Natl Acad Sci* **105**: 14076-14081.

Bole-Feysot C, Goffin V, Edery M, Binart N and Kelly PA (1998) Prolactin (PRL) and its receptor: actions, signal transduction pathways and phenotypes observed in PRL receptor knockout mice. *Endocr Rev* **19**(3): 225-268.

Booth BW and Smith GH (2006) Estrogen receptor-alpha and progesterone receptor are expressed in label-retaining mammary epithelial cells that divide asymmetrically and retain their template DNA strands. *Breast Cancer Res* **8**(4): R49.

Brambrink T, Foreman R, Welstead GG, Lengner CJ, Wernig M, Suh H and Jaenisch R (2008) Sequential expression of pluripotency markers during direct reprogramming of mouse somatic cells. *Cell Stem Cell* **2**(2): 151-159.

Brisken C and Duss S (2007) Stem cells and the stem cell niche in the breast: an integrated hormonal and developmental perspective. *Stem Cell Rev* **3**(2): 147-156.

Brisken C and O'Malley B (2010) Hormone action in the mammary gland. *Cold Spring Harb Perspect Biol* **2**(12): a003178.

Brisken C, Park S, Vass T, Lydon JP, O'Malley BW and Weinberg RA (1998) A paracrine role for the epithelial progesterone receptor in mammary gland development. *Proc Natl Acad Sci USA* **95**(9): 5076-5081.

Brisken C and Rajaram R (2006) Alveolar and lactogenic differentiation. *J Mammary Gland Biol Neoplasia* **11**(3-4): 239-248.

Busche S, Descot A, Julien S, Genth H and Posern G (2008) Epithelial cell-cell contacts regulate SRF-mediated transcription via Rac-actin-MAL signalling. *J Cell Sci* **121**: 1025-1035.

Buser AC, Gass-Handel EK, Wyszomierski SL, Doppler W, Leonhardt SA, Schaack J, Rosen JM, Watkin H, Anderson SM and Edwards DP (2013) Progesterone receptor repression of prolactin/signal transducer and activator of transcription 5-mediated transcription of the β-casein gene in mammary epithelial cells. *Mol Endocrinol* **21**(1): 106-125.

Bussard KM, Boulanger CA, Booth BW, Bruno RD and Smith GH (2010) Reprogramming human cancer cells in the mouse mammary gland. *Cancer Res* **70**(15): 6336-6343.

Cairns JM and Saunders JW (1954) The influence of embryonic mesoderm on the regional specification of epidermal derivatives in the chick. *J Exp Zool* **127**(2): 221-248.

Carroll JS, Liu XS, Brodsky AS, Li W, Meyer CA, Szary AJ, Eeckhoute J, Shao W, Hestermann EV, Geistlinger TR, Fox EA, Silver PA and Brown M (2005) Chromosome-wide mapping of estrogen receptor binding reveals long-range regulation requiring the forkhead protein FoxA1. *Cell* **122**(1): 33-43.

Chen Y, Bates DL, Dey R, Chen P-H, Machado ACD, Laird-Offringa IA, Rohs R and Chen L (2012) DNA binding by GATA transcription factor suggests mechanisms of DNA looping and long-range gene regulation. *Cell Rep* **2**(5): 1197-1206.

Cho HJ, Lee CS, Kwon YW, Paek JS, Lee SH, Hur J, Lee EJ, Roh TY, Chu IS, Leem SH, Kim Y, Kang HJ, Park YB and Kim HS (2010) Induction of pluripotent stem cells from adult somatic cells by protein-based reprogramming without genetic manipulation. *Blood* **116**(3): 386-395.

Chou J, Provot S and Werb Z (2010) GATA3 in development and cancer differentiation: Cells GATA have it! *J Cell Physiol* **222**(1): 42-49.

Clarke RB, Howell A, Potten CS and Anderson E (1997) Dissociation between steroid receptor expression and cell proliferation in the human breast. *Cancer Res* **57**(22): 4987-4991.

Cohen H, Ben-Hamo R, Gidoni M, Yitzhaki I, Kozol R, Zilberberg A and Efroni S (2014) Shift in GATA3 functions, and GATA3 mutations, control progression and clinical presentation in breast cancer. *Breast Cancer Res* **16**(6): 464.

Coleman-Krnacik S and Rosen JM (1994) Differential temporal and spatial gene expression of fibroblast growth factor family members during mouse mammary gland development. *Mol Endocrinol* **8**: 218-229.

Conneely OM, Mulac-Jericevic B and Lydon JP (2003) Progesterone-dependent regulation of female reproductive activity by two distinct progesterone receptor isoforms. *Steroids* **68**(10-13): 771-778.

Conneely OM, Mulac-Jericevic B, Lydon JP and De Mayo FJ (2001) Reproductive functions of the progesterone receptor isoforms: lessons from knock-out mice. *Mol Cell Endocrinol* **179**(1): 97-103.

Cooke LS, Rex AH and Paul S (2010) Spermatogonial stem cells, in vivo transdifferentiation and human regenerative medicine. *Expert Opin Biol Ther* **10**(4): 519-530.

Corominas-Faja B, Cufí S, Oliveras-Ferraros C, Cuyàs E, López-Bonet E, Lupu R, Alarcón T, Vellon L, Manuel Iglesias J, Leis O, Martin A, Vazquez-Martin A and Menendez JA (2013) Nuclear reprogramming of luminal-like breast cancer cells generates Sox2-overexpressing cancer stem-like cellular states harboring transcriptional activation of the mTOR pathway. *Cell Cycle* **12**: 3109-3124.

Couse JF and Korach KS (1999) Estrogen receptor null mice: what have we learned and where will they lead us? *Endocr Rev* **20**(3): 358-417.

Cowin P and Wysolmerski J (2010) Molecular mechanisms guiding embryonic mammary gland development. *Cold Spring Harb Perspect Biol* **2**(6): a003251.

Crossley M, Merika M and Orkin SH (1995) Self-association of the erythroid transcription factor GATA-1 mediated by its zinc finger domains. *Mol Cell Biol* **15**(5): 2448-2456.

Davis RL, Weintraub H and Lassar AB (1987) Expression of a single transfected cDNA converts fibroblasts to myoblasts. *Cell* **51**(6): 987-1000.

Debnath J, Mills KR, Collins NL, Reginato MJ, Muthuswamy SK and Brugge JS (2002) The role of apoptosis in creating and maintaining luminal space within normal and oncogene-expressing mammary acini. *Cell* **111**(1): 29-40.

Debnath J, Muthuswamy SK and Brugge JS (2003) Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**: 256-268.

Desponts C and Ding S (2010) Using small molecules to improve generation of induced pluripotent stem cells from somatic cells. *Methods Mol Biol* **636**: 207-218.

Dimri G, Band H and Band V (2005) Mammary epithelial cell transformation: insights from cell culture and mouse models. *Breast Cancer Res* **7**(4): 171.

Doi A, Park I-H, Wen B, Murakami P, Aryee MJ, Irizarry R, Herb B, Ladd-Acosta C, Rho J, Loewer S, Miller J, Schlaeger T, Daley GQ and Feinberg AP (2009) Differential methylation of tissue- and cancer-specific CpG island shores distinguishes human induced pluripotent stem cells, embryonic stem cells and fibroblasts. *Nat Genet* **41**: 1350-1353.

Doppler W, Hock W, Hofer P, Groner B and Ball RK (1990) Prolactin and glucocorticoid hormones control transcription of the beta-casein gene by kinetically distinct mechanisms. *Mol Endocrinol* **4**(6): 912-919.

Dove A (2015) The art of culture: developing cell lines. *Science* **346**: 1013-1015.

Dutertre M and Smith CL (2003) Ligand-independent interactions of p160/steroid receptor coactivators and CREB-binding protein (CBP) with estrogen receptor-alpha: regulation by phosphorylation sites in the A/B region depends on other receptor domains. *Mol Endocrinol* **17**(7): 1296-1314.

Dydensborg AB, Rose AA, Wilson BJ, Grote D, Paquet M, Giguere V, Siegel PM and Bouchard M (2009) GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. *Oncogene* **28**(29): 2634-2642.

Dyson HJ and Wright PE (2002) Coupling of folding and binding for unstructured proteins. *Curr Opin Struct Biol* **12**(1): 54-60.

Edwards J (2010) Why Dolly matters: Kinship, culture and cloning. *Ethnos* **64**: 301-324.

Eeckhoute J, Carroll JS, Geistlinger TR, Torres-Arzayus MI and Brown M (2006) A cell-type-specific transcriptional network required for estrogen regulation of cyclin D1 and cell cycle progression in breast cancer. *Genes Dev* **20**(18): 2513-2526.

Eeckhoute J, Keeton EK, Lupien M, Krum SA, Carroll JS and Brown M (2007) Positive cross-regulatory loop ties GATA-3 to Estrogen receptor α expression in breast cancer. *Cancer Res* **67**(13): 6477-6483.

Eggan K, Jacqueline R and Garrett B (2007) Developmental reprogramming after chromosome transfer into mitotic mouse zygotes. *Nature* **447**(7145): 679-685.

Fang TC, Yashiro-Ohtani Y, Del Bianco C, Knoblock DM, Blacklow SC and Pear WS (2007) Notch directly regulates gata3 expression during T helper 2 cell differentiation. *Immunity* **27**: 100-110.

Fernandez-Valdivia R, Mukherjee A, Ying Y, Li J, Paquet M, DeMayo F and Lydon J (2009) The RANKL signaling axis is sufficient to elicit ductal side-branching and alveologenesis in the mammary gland of the virgin mouse. *Dev Biol* **328**: 127-139.

Fillmore CM and Kuperwasser C (2008) Human breast cancer cell lines contain stem-like cells that self-renew, give rise to phenotypically diverse progeny and survive chemotherapy. *Breast Cancer Research* **10**(2): R25.

Frasor J and Gibori G (2003) Prolactin regulation of estrogen receptor expression. *Trends Endocrinol Metab* **14**(3): 118-123.

Gomm JJ, Smith J, Ryall GK, Baillie R, Turnbull L and Coombes RC (1991) Localization of basic fibroblast growth factor and transforming growth factor beta 1 in the human mammary gland. *Cancer Res* **51**(17): 4685-4692.

Gong C, Fujino K, Monteiro LJ, Gomes AR, Drost R, Davidson-Smith H, Takeda S, Khoo US, Jonkers J, Sproul D and Lam EW-F (2014) FOXA1 repression is associated with loss of BRCA1 and increased promoter methylation and chromatin silencing in breast cancer. *Oncogene* **34**: 5012-5024. doi:10.1038/onc.2014.421.

Gordon KE, Binas B, Chapman RS, Kurian KM, Clarkson RWE, Clark JA, Lane BE and Watson CJ (2000) A novel cell culture model for studying differentiation and apoptosis in the mouse mammary gland. *Breast Cancer Res* **2**(3): 222.

Graf T (2011) Historical origins of transdifferentiation and reprogramming. *Cell Stem Cell* **9**(6): 504-516.

Graham JD and Clarke CL (1997) Physiological action of progesterone in target tissues. *Endocr Rev* **18**(4): 502-519.

Graham K, Ge X, de las Morenas A, Tripathi A and Rosenberg CL (2011) Gene expression profiles of estrogen receptor-positive and estrogen receptor-negative breast cancers are detectable in histologically normal breast epithelium. *Clin Cancer Res* **17**: 236-246.

Grober O, Mutarelli M, Giurato G, Ravo M, Cicatiello L, De Filippo M, Ferraro L, Nassa G, Papa M, Paris O, Tarallo R, Luo S, Schroth G, Benes V and Weisz A (2011) Global analysis of estrogen receptor beta binding to breast cancer cell genome reveals an extensive interplay with estrogen receptor alpha for target gene regulation. *BMC Genomics* **12**: 36.

Grote D, Boualia SK, Souabni A, Merkel C, Chi X, Costantini F, Carroll T and Bouchard M (2008) GATA3 acts downstream of β-catenin signaling to prevent ectopic metanephric kidney induction. *PLoS Genet* **4**(12): e1000316.

Hamilton KJ, Arao Y and Korach KS (2014) Estrogen hormone physiology: reproductive findings from estrogen receptor mutant mice. *Reprod Biol* **14**(1): 3-8.

Hasegawa D, Calvo V, Avivar-Valderas A, Lade A, Chou H-I, Lee YA, Farias EF, Aguirre-Ghiso JA and Friedman SL (2015) Epithelial xbp1 is required for cellular proliferation and differentiation during mammary gland development. *Mol Cell Biol* **35**: 1543-1556.

Helguero LA, Faulds MH, Gustafsson J-A and Haldosén L-A (2005) Estrogen receptors alfa (ERα) and beta (ERβ) differentially regulate proliferation and apoptosis of the normal murine mammary epithelial cell line HC11. *Oncogene* **24**: 6605-6616.

Hendriks RW, Nawijn MC, Engel JD, van Doorninck H, Grosveld F and Karis A (1999) Expression of the transcription factor GATA-3 is required for the development of the earliest T cell progenitors and correlates with stages of cellular proliferation in the thymus. *Eur J Immunol* **29**(6): 1912-1918.

Hennighausen L and Robinson GW (2001) Signaling pathways in mammary gland development. *Dev Cell* **1**(4): 467-475.

Herschkowitz JI (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8**:R76.

Herschkowitz JI, Simin K, Weigman VJ, Mikaelian I, Usary J, Hu Z, Rasmussen KE, Jones LP, Assefnia S, Chandrasekharan S, Backlund MG, Yin Y, Khramtsov AI, Bastein R, Quackenbush J, Glazer RI, Brown PH, Green JE, Kopelovich L, Furth PA, Palazzo JP, Olopade OI, Bernard PS, Churchill GA, Van Dyke T and Perou CM (2007) Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol* **8**(5):R76.

Heuberger B, Fitzka I, Wasner G and Kratochwil K (1982) Induction of androgen receptor formation by epithelium-mesenchyme interaction in embryonic mouse mammary gland. *Proceedings of the National Academy of Sciences* **79**(9): 2957-2961.

Hoffmann J, Bohlmann R, Heinrich N, Hofmeister H, Kroll J, Kunzer H, Lichtner RB, Nishino Y, Parczyk K, Sauer G, Gieschen H, Ulbrich H-F and Schneider MR (2004) Characterization of new estrogen receptor destabilizing compounds: effects on estrogen-sensitive and tamoxifen-resistant breast cancer. *JNCI J Natl Cancer Inst* **96**: 210-218.

Hollern DP and Andrechek ER (2014) A genomic analysis of mouse models of breast cancer reveals molecular features of mouse models and relationships to human breast cancer. *Breast Cancer Res* **16**:R59.

Hou P, Li Y, Zhang X, Liu C, Guan J, Li H, Zhao T, Ye J, Yang W, Liu K, Ge J, Xu J, Zhang Q, Zhao Y and Deng H (2013) Pluripotent stem cells induced from mouse somatic cells by small-molecule compounds. *Science* **341**(6146): 651-654.

Huang P, He Z, Ji S, Sun H, Xiang D, Liu C, Hu Y, Wang X and Hui L (2011) Induction of functional hepatocyte-like cells from mouse fibroblasts by defined factors. *Nature* **475**(7356): 386-389.

Hurtado A, Holmes KA, Ross-Innes CS, Schmidt D and Carroll JS (2011) FOXA1 is a key determinant of estrogen receptor function and endocrine response. *Nat Genet* **43**: 27-33.

Ieda M, Fu JD, Delgado-Olguin P, Vedantham V, Hayashi Y, Bruneau BG and Srivastava D (2010) Direct reprogramming of fibroblasts into functional cardiomyocytes by defined factors. *Cell* **142**(3): 375-386.

Inic Z, Zegarac M, Inic M, Kozomara Z, Djurisic I, Inic I, Pupic G, Jancic S and Markovic I (2014) Difference between Luminal A and Luminal B subtypes according to ki-67, tumor size, and progesterone receptor negativity providing prognostic information. *Clin Med Insights Oncol* **8**: 107-111.

Ismail PM, Amato P, Soyal SM, DeMayo FJ, Conneely OM, O'Malley BW and Lydon JP (2003) Progesterone involvement in breast development and tumorigenesis—as revealed by progesterone receptor "knockout" and "knockin" mouse models. *Steroids* **68**(10-13): 779-787.

Jeselsohn R, Brown NE, Arendt L, Klebba I, Hu MG, Kuperwasser C and Hinds PW (2010) Cyclin D1 kinase activity is required for the self-renewal of mammary stem and progenitor cells that are targets of MMTV-ErbB2 tumorigenesis. *Cancer Cell* **17**(1): 65-76.

Jiang S, Katayama H, Wang J, Li SA, Hong Y, Radvanyi L, Li JJ and Sen S (2010) Estrogen-induced aurora kinase-A (AURKA) gene expression is activated by GATA-3 in estrogen receptor-positive breast cancer cells. *Horm Cancer* **1**(1): 11-20.

Jin V, Leu Y, Liyanarachchi S, Sun H, Fan M, Nephew K, Huang T and Davuluri R (2004) Identifying estrogen receptor alpha target genes using integrated computational genomics and chromatin immunoprecipitation microarray. *Nucl Acids Res* **32**: 6627-6635.

Kazansky AV, Kabotyanski EB, Wyszomierski SL, Mancini MA and Rosen JM (1999) Differential effects of prolactin and src/abl kinases on the nuclear translocation of STAT5B and STAT5A. *J Biol Chem* **274**(32): 22484-22492.

Keller PJ, Lin A, Arendt LM, Klebba I, Jones AD, Rudnick JA, DiMeo TA, Gilmore H, Jefferson DM, Graham RA, Naber SP, Schnitt S and Kuperwasser C (2010) Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Res* **12**:R87.

Kenny PA, Lee GY, Myers CA, Neve RM, Semeiks JR, Spellman PT, Lorenz K, Lee EH, Barcellos-Hoff MH, Petersen OW, Gray JW and Bissell MJ (2007) The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. *Mol Oncol* **1**(1): 84-96.

Keymeulen AV, Rocha AS, Ousset M, Beck B, Bouvencourt G, Rock J, Sharma N, Dekoninck S and Blanpain C (2011) Distinct stem cells contribute to mammary gland development and maintenance. *Nature* **479**: 189-193.

Khaled WT, Read EK, Nicholson SE, Baxter FO, Brennan AJ, Came PJ, Sprigg N, McKenzie AN and Watson CJ (2007) The IL-4/IL-13/Stat6 signalling pathway promotes luminal mammary epithelial cell development. *Development* **134**(15): 2739-2750.

Koboldt DC, Fulton RS, McLellan MD, Schmidt H, Kalicki-Veizer J, McMichael JF, Fulton LL, Dooling DJ, Ding L, Mardis ER, Wilson RK, Ally A, Balasundaram M, Butterfield YSN, Carlsen R, Carter C, Chu A, Chuah E, Chun H-JE, Coope RJN, Dhalla N, Guin R, Hirst C, Hirst M, Holt RA, Lee D, Li HI, Mayo M, Moore RA, Mungall AJ, Pleasance E, Gordon Robertson A, Schein JE, Shafiei A, Sipahimalani P, Slobodan JR, Stoll D, Tam A, Thiessen N, Varhol RJ, Wye N, Zeng T, Zhao Y, Birol I, Jones SJM, Marra MA, Cherniack AD, Saksena G, Onofrio RC, Pho NH, Carter SL, Schumacher SE, Tabak B, Hernandez B, Gentry J, Nguyen H, Crenshaw A, Ardlie K, Beroukhim R, Winckler W, Getz G, Gabriel SB, Meyerson M, Chin L, Park PJ, Kucherlapati R, Hoadley KA, Todd Auman J, Fan C, Turman YJ, Shi Y, Li L, Topal MD, He X, Chao H-H, Prat A, Silva GO, Iglesia MD, Zhao W, Usary J, Berg JS, Adams M, Booker J, Wu J, Gulabani A, Bodenheimer T, Hoyle AP, Simons JV, Soloway MG, Mose LE, Jefferys SR, Balu S, Parker JS, Neil Hayes D, Perou CM, Malik S, Mahurkar S, Shen H, Weisenberger DJ,

Triche Jr T, Lai PH, Bootwalla MS, Maglinte DT, Berman BP, Van Den Berg DJ, Baylin SB, Laird PW, Creighton CJ, Donehower LA, Getz G, Noble M, Voet D, Saksena G, Gehlenborg N, DiCara D, Zhang J, Zhang H, Wu C-J, Yingchun Liu S, Lawrence MS, Zou L, Sivachenko A, Lin P, Stojanov P, Jing R, Cho J, Sinha R, Park RW, Nazaire M-D, Robinson J, Thorvaldsdottir H, Mesirov J, Park PJ, Chin L, Reynolds S, Kreisberg RB, Bernard B, Bressler R, Erkkila T, Lin J, Thorsson V, Zhang W, Shmulevich I, Ciriello G, Weinhold N, Schultz N, Gao J, Cerami E, Gross B, Jacobsen A, Sinha R, Arman Aksoy B, Antipin Y, Reva B, Shen R, Taylor BS, Ladanyi M, Sander C, Anur P, Spellman PT, Lu Y, Liu W, Verhaak RRG, Mills GB, Akbani R, Zhang N, Broom BM, Casasent TD, Wakefield C, Unruh AK, Baggerly K, Coombes K, Weinstein JN, Haussler D, Benz CC, Stuart JM, Benz SC, Zhu J, Szeto CC, Scott GK, Yau C, Paull EO, Carlin D, Wong C, Sokolov A, Thusberg J, Mooney S, Ng S, Goldstein TC, Ellrott K, Grifford M, Wilks C, Ma S, Craft B, Yan C, Hu Y, Meerzaman D, Gastier-Foster JM, Bowen J, Ramirez NC, Black AD, Xpath Error: unknown variable "tname RE, White P, Zmuda EJ, Frick J, Lichtenberg TM, Brookens R, George MM, Gerken MA, Harper HA, Leraas KM, Wise LJ, Tabler TR, McAllister C, Barr T, Hart-Kothari M, Tarvin K, Saller C, Sandusky G, Mitchell C, Iacocca MV, Brown J, Rabeno B, Czerwinski C, Petrelli N, Dolzhansky O, Abramov M, Voronina O, Potapova O, Marks JR, Suchorska WM, Murawa D, Kycler W, Ibbs M, Korski K, Spychała A, Murawa P, Brzeziński JJ, Perz H, Łaźniak R, Teresiak M, Tatka H, Leporowska E, Bogusz-Czerniewicz M, Malicki J, Mackiewicz A, Wiznerowicz M, Van Le X, Kohl B, Viet Tien N, Thorp R, Van Bang N, Sussman H, Duc Phu B, Hajek R, Phi Hung N, Viet The Phuong T, Quyet Thang H, Zaki Khan K, Penny R, Mallery D, Curley E, Shelton C, Yena P, Ingle JN, Couch FJ, Lingle WL, King TA, Maria Gonzalez-Angulo A, Mills GB, Dyer MD, Liu S, Meng X, Patangan M, Waldman F, Stöppler H, Kimryn Rathmell W, Thorne L, Huang M, Boice L, Hill A, Morrison C, Gaudioso C, Bshara W, Daily K, Egea SC, Pegram MD, Gomez-Fernandez C, Dhir R, Bhargava R, Brufsky A, Shriver CD, Hooke JA, Leigh Campbell J, Mural RJ, Hu H, Somiari S, Larson C, Deyarmin B, Kvecher L, Kovatich AJ, Ellis MJ, King TA, Hu H, Couch FJ, Mural RJ, Stricker T, White K, Olopade O, Ingle JN, Luo C, Chen Y, Marks JR, Waldman F, Wiznerowicz M, Bose R, Chang L-W, Beck AH, Maria Gonzalez-Angulo A, Pihl T, Jensen M, Sfeir R, Kahn A, Chu A, Kothiyal P, Wang Z, Snyder E, Pontius J, Ayala B, Backus M, Walton J, Baboud J, Berton D, Nicholls M, Srinivasan D, Raman R, Girshik S, Kigonya P, Alonso S, Sanbhadti R, Barletta S, Pot D, Sheth M, Demchok JA, Mills Shaw KR, Yang L, Eley G, Ferguson ML, Tarnuzzer RW, Zhang J, Dillon LAL, Buetow K, Fielding P, Ozenberger BA, Guyer MS, Sofia HJ and Palchik JD (2012) Comprehensive molecular portraits of human breast tumours. *Nature* **490**(7418): 61-70.

Kong SL, Li G, Loh SL, Sung W-K and Liu ET (2011) Cellular reprogramming by the conjoint action of ER , FOXA1, and GATA3 to a ligand-inducible growth state. *Mol Syst Biol* **7**: 526.

Kouros-Mehr H, Bechis S, Slorach E, Littlepage L, Egeblad M, Ewald A, Pai S, Ho I and Werb Z (2008) GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell* **13**: 141-152.

Kouros-Mehr H, Slorach E, Sternlicht M and Werb Z (2006) GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127**: 1041-1055.

Kratochwil K (1969) Organ specificity in mesenchymal induction demonstrated in the embryonic development of the mammary gland of the mouse. *Dev Biol* **20**(1): 46-71.

Kratochwil K, Tontsch S and Grosschedl R (2003) Function of LEF1 in early mammary development. *Breast Cancer Res* **5**(Suppl 1): 35.

Kreike B, van Kouwenhove M, Horlings H, Weigelt B, Peterse H, Bartelink H and van de Vijver MJ (2007) Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Res* **9**(5): R65.

Kulessa H, Frampton J and Graf T (1995) GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts. *Genes Dev* **9**(10): 1250-1262.

Lange CA, Richer JK and Horwitz KB (1999) Hypothesis: progesterone primes breast cancer cells for cross-talk with proliferative or antiproliferative signals. *Mol Endocrinol* **13**: 829-836.

Leo C and Chen JD (2000) The SRC family of nuclear receptor coactivators. *Gene* **245**(1): 1-11.

Lim E, Vaillant F, Wu D, Forrest NC, Pal B, Hart AH, Asselin-Labat M-L, Gyorki DE, Ward T, Partanen A, Feleppa F, Huschtscha LI, Thorne HJ, Fox SB, Yan M, French JD, Brown MA, Smyth GK, Visvader JE and Lindeman GJ (2009) Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nature Med* **15**(8): 907-913.

Lim E, Wu D, Pal B, Bouras T, Asselin-Labat M-L, Vaillant F, Yagita H, Lindeman GJ, Smyth GK and Visvader JE (2010) Transcriptome analyses of mouse and human mammary cell subpopulations reveal multiple conserved genes and pathways. *Breast Cancer Res* **12**(2): R21.

Lim K-C, Lakshmanan G, Crawford SE, Gu Y, Grosveld F and Douglas Engel J (2000) Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nat Genet* **25**(2): 209-212.

Lippman M, Bolan G and Huff K (1976) The effects of estrogens and antiestrogens on hormone-responsive human breast cancer in long-term tissue culture. *Cancer Res* **36**: 4595-4601.

Liu S (2008) BRCA1 regulates human mammary stem/progenitor cell fate. *Proc Natl Acad Sci USA* **105**: 1680-1685.

Liu S, Robinson GW, Wagner KU, Garrett L, Wynshaw-Boris A and Hennighausen L (1997) Stat5a is mandatory for adult mammary gland development and lactogenesis. *Genes & Development* **11**(2): 179-186.

Macias H, Moran A, Samara Y, Moreno M, Compton JE, Harburg G, Strickland P and Hinck L (2011) SLIT/ROBO1 signaling suppresses mammary branching morphogenesis by limiting basal cell number. *Dev Cell* **20**(6): 827-840.

Mallepell S, Krust A, Chambon P and Brisken C (2006) Paracrine signaling through the epithelial estrogen receptor alpha is required for proliferation and morphogenesis in the mammary gland. *Proc Natl Acad Sci USA* **103**(7): 2196-2201.

Mani SA, Guo W, Liao MJ, Eaton EN, Ayyanan A, Zhou AY, Brooks M, Reinhard F, Zhang CC, Shipitsin M, Campbell LL, Polyak K, Brisken C, Yang J and Weinberg RA (2008) The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**(4): 704-715.

Mark S, François V, Kaylene JS, John S, Gordon KS, Marie-Liesse A-L, Li W, Geoffrey JL and Visvader JE (2006) Generation of a functional mammary gland from a single stem cell. *Nature* **439**(7072): 84-88.

Marte BM, Jeschke M, Graus-Porta D, Taverna D, Hofer P, Groner B, Yarden Y and Hynes NE (1995) Neu differentiation factor/heregulin modulates growth and differentiation of HC11 mammary epithelial cells. *Mol Endocrinol* **9**: 14-23.

Masamura S, Santner SJ, Heitjan DF and Santen RJ (1995) Estrogen deprivation causes estradiol hypersensitivity in human breast cancer cells. *J Clin Endocrinol Metab* **80**: 2918-2925.

Mauvais-Jarvis P, Kuttenn F and Gompel A (1986) Estradiol/progesterone interaction in normal and pathologic breast cells. *Ann N Y Acad Sci* **464**: 152-167.

Melton QZ, Juliana B, Andrew K, Jayaraj R and Douglas A (2008) In vivo reprogramming of adult pancreatic exocrine cells to |[bgr]|-cells. *Nature* **455**(7213): 627-632.

Moggs JG and Orphanides G (2001) Estrogen receptors: orchestrators of pleiotropic cellular responses. *EMBO Rep* **2**(9): 775-781.

Moses H and Barcellos-Hoff MH (2011) TGF-beta biology in mammary development and breast cancer. *Cold Spring Harb Perspect Biol* **3**(1): a003277.

Muhammad IM and Mohammad MG (2013) immortality of cell lines: challenges and advantages of establishment. *Cell Biol Int* **37**: 1038-1045.

Naderi A, Meyer M and Dowhan DH (2012) Cross-regulation between FOXA1 and ErbB2 signaling in estrogen receptor-negative breast cancer12. *Neoplasia* **14**(4): 283-296.

Nesbit MA, Bowl MR, Harding B, Ali A, Ayala A, Crowe C, Dobbie A, Hampson G, Holdaway I, Levine MA, McWilliams R, Rigden S, Sampson J, Williams AJ and Thakker RV (2004) Characterization of GATA3 mutations in the hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome. *J Biol Chem* **279**(21): 22624-22634.

Neve RM, Chin K, Fridlyand J, Yeh J, Baehner FL, Fevr T, Clark L, Bayani N, Coppe J-P, Tong F, Speed T, Spellman PT, DeVries S, Lapuk A, Wang NJ, Kuo W-L, Stilwell JL, Pinkel D, Albertson DG, Waldman FM, McCormick F, Dickson RB, Johnson MD, Lippman M, Ethier S, Gazdar A and Gray JW (2006) A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**(6): 515-527.

Neville M, McFadden T and Forsyth I (2002) Hormonal regulation of mammary differentiation and milk secretion. *J Mammary Gland Biol Neoplasia* **7**: 49-66.

Newcomb PA, Storer BE, Longnecker MP, Mittendorf R, Greenberg ER, Clapp RW, Burke KP, Willett WC and MacMahon B (1994) Lactation and a reduced risk of premenopausal breast cancer. *N Engl J Med* **330**(2): 81-87.

Nielsen TO (2004) Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. *Clin Cancer Res* **10**: 5367-5374.

Oakes SR, Gallego-Ortega D and Ormandy CJ (2014) The mammary cellular hierarchy and breast cancer. *Cell Mol Life Sci* **71**(22): 4301-4324.

Obr AE and Edwards DP (2012) The biology of progesterone receptor in the normal mammary gland and in breast cancer. *Mol Cell Endocrinol* **357**(1-2): 4-17.

Oftedal OT (2002) The mammary gland and its origin during synapsid evolution. *J Mammary Gland Biol Neoplasia* **7**(3): 225-252.

Oh H-Y, Jin X, Kim J-G, Oh M-J, Pian X, Kim J-M, Yoon M-S, Son C-I, Lee YS, Hong K-C, Kim H, Choi Y-J and Whang KY (2007) Characteristics of primary and immortalized fibroblast cells derived from the miniature and domestic pigs. *BMC Cell Biol* **8**(1): 20.

Ormandy CJ, Binart N and Kelly PA (1997) Mammary gland development in prolactin receptor knockout mice. *J Mammary Gland Biol Neoplasia* **2**(4): 355-364.

Pandolfi PP, Roth ME, Karis A, Leonard MW, Dzierzak E, Grosveld FG, Engel JD and Lindenbaum MH (1995) Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat Genet* **11**(1): 40-44.

Parker JS, Mullins M, Cheang MCU, Leung S, Voduc D, Vickery T, Davies S, Fauron C, He X, Hu Z, Quackenbush JF, Stijleman IJ, Palazzo J, Marron JS, Nobel AB, Mardis E, Nielsen TO, Ellis MJ, Perou CM and Bernard PS (2009) Supervised risk predictor of breast cancer based on intrinsic subtypes. *J Clin Oncol* **27**: 1160-1167.

Patel M and Yang S (2010) Advances in reprogramming somatic cells to induced pluripotent stem cells. *Stem Cell Rev Rep* **6**: 367-380.

Pedone PV, Omichinski JG, Nony P, Trainor C, Gronenborn AM, Clore GM and Felsenfeld G (1997) The N-terminal fingers of chicken GATA-2 and GATA-3 are independent sequence-specific DNA binding domains. *EMBO J* **16**(10): 2874-2882.

Perou CM, Sørlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, Pollack JR, Ross DT, Johnsen H, Akslen LA, Fluge O, Pergamenschikov A, Williams C, Zhu SX, Lønning PE, Børresen-Dale AL, Brown PO and Botstein D (2000) Molecular portraits of human breast tumours. *Nature* **406**(6797): 747-752.

Petersen OW, Hoyer PE and Deurs B (1997) Frequency and distribution of estrogen receptor-positive cells in normal, nonlactating human breast tissue. *Cancer Res* **47**: 5748-5751.

Pfefferle AD, Herschkowitz JI, Usary J, Harrell J, Spike BT, Adams JR, Torres-Arzayus MI, Brown M, Egan SE, Wahl GM, Rosen JM and Perou CM (2013) Transcriptomic classification of genetically engineered mouse models of breast cancer identifies human subtype counterparts. *Genome Biol* **14**:R125.

Piek E, Moustakas A, Kurisaki A, Heldin C-H and ten Dijke P (1999) TGF-(beta) type I receptor/ALK-5 and Smad proteins mediate epithelial to mesenchymal transdifferentiation in NMuMG breast epithelial cells. *J Cell Sci* **112**: 4557-4568.

Prat A, Karginova O, Parker JS, Fan C, He X, Bixby L, Harrell JC, Roman E, Adamo B, Troester M and Perou CM (2013) Characterization of cell lines derived from breast cancers and normal mammary tissues for the study of the intrinsic molecular subtypes. *Breast Cancer Res Treat* **142**(2): 237-255.

Prat A, Parker JS, Karginova O, Fan C, Livasy C, Herschkowitz JI, He X and Perou CM (2010) Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. *Breast Cancer Res Treat* **12**(5): R68.

Prat A and Perou C (2011) Deconstructing the molecular portraits of breast cancer. *Mol Oncol* **5**: 5-23.

Rajaram RD and Brisken C (2012) Paracrine signaling by progesterone. *Mol Cell Endocrinol* **357**: 80-90.

Ranganath S and Murphy KM (2001) Structure and specificity of GATA proteins in Th2 development. *Mol Cell Biol* **21**(8): 2716-2725.

Ray A, Prefontaine KE and Ray P (1994) Down-modulation of interleukin-6 gene expression by 17 beta-estradiol in the absence of high affinity DNA binding by the estrogen receptor. *J Biol Chem* **269**(17): 12940-12946.

Rios AC, Fu NY, Lindeman GJ and Visvader JE (2014) In situ identification of bipotent stem cells in the mammary gland. *Nature* **506**(7488): 322-327.

Rogier EW, Frantz AL, Bruno ME, Wedlund L, Cohen DA, Stromberg AJ and Kaetzel CS (2014) Lessons from mother: Long-term impact of antibodies in breast milk on the gut microbiota and intestinal immune system of breastfed offspring. *Gut Microbes* **5**(5): 663-668.

Ross DT and Perou CM (2001) A comparison of gene expression signatures from breast tumors and breast tissue derived cell lines. *Dis Markers* **17**(2): 99-109.

Sandel MJ (2004) Embryo ethics: The moral logic of stem-cell research. *N Engl J Med* **351**(3): 207-209.

Schmeichel KL and Bissell MJ (2003) Modeling tissue-specific signaling and organ function in three dimensions. *J Cell Sci* **116**(12): 2377-2388.

Sell S (2013) *Stem cells handbook*. Springer New York.

Shackleton M (2006) Generation of a functional mammary gland from a single stem cell. *Nature* **439**: 84-88.

Shehata M, Teschendorff A, Sharp G, Novcic N, Russell IA, Avril S, Prater M, Eirew P, Caldas C and Watson CJ (2012) Phenotypic and functional characterisation of the luminal cell hierarchy of the mammary gland. *Breast Cancer Res Treat* **14**(5): R134.

Silberstein GB and Daniel CW (1987) Reversible inhibition of mammary gland growth by transforming growth factor-beta. *Science* **237**(4812): 291-293.

Sleeman KE, Kendrick H, Ashworth A, Isacke CM and Smalley MJ (2006) CD24 staining of mouse mammary gland cells defines luminal epithelial, myoepithelial/basal and non-epithelial cells. *Breast Cancer Res Treat* **8**(1): R7.

Sleeman KE, Kendrick H, Robertson D, Isacke CM, Ashworth A and Smalley MJ (2007) Dissociation of estrogen receptor expression and in vivo stem cell activity in the mammary gland. *J Cell Biol* **176**(1): 19-26.

Smalley MJ (2015) Isolation of mouse mammary epithelial subpopulations. *J Mammary Gland Biol Neoplasia* **17**(2): 91-97.

Smalley MJ, Titley J and O'Hare M (1998) Clonal characterization of mouse mammary luminal epithelial and myoepithelial cells separated by fluorescence-activated cell sorting. *In Vitro Cell Dev Biol Anim* **34**(9): 711-721.

Smith G (1996) Experimental mammary epithelial morphogenesis in anin vivo model: evidence for distinct cellular progenitors of the ductal and lobular phenotype. *Breast Cancer Res Treat* **39**(1): 21-31.

Smith GH and Boulanger CA (2003) Mammary epithelial stem cells: transplantation and self-renewal analysis. *Cell Prolif* **36**: 3-15.

Sridharan R, Tchieu J, Mason MJ, Yachechko R, Kuoy E, Horvath S, Zhou Q and Plath K (2009) Role of the murine reprogramming factors in the induction of pluripotency. *Cell* **136**(2): 364-377.

Stadtfeld M and Hochedlinger K (2010) Induced pluripotency: history, mechanisms, and applications. *Genes Dev* **24**(20): 2239-2263.

Stein B and Yang MX (1995) Repression of the interleukin-6 promoter by estrogen receptor is mediated by NF-kappa B and C/EBP beta. *Mol Cell Biol* **15**(9): 4971-4979.

Stein T, Morris JS, Davies CR, Weber-Hall SJ, Duffy MA, Heath VJ, Bell AK, Ferrier RK, Sandilands GP and Gusterson BA (2004) Involution of the mouse mammary gland is associated with an immune cascade and an acute-phase response, involving LBP, CD14 and STAT3. *Breast Cancer Res* **6**(2): R75-R91.

Sternlicht MD and Sunnarborg SW (2008) The ADAM17-amphiregulin-EGFR axis in mammary development and cancer. *J Mammary Gland Biol Neoplasia* **13**(2): 181-194.

Stingl J, Eaves CJ, Zandieh I and Emerman JT (2001) Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67**: 93-109.

Stingl J, Eirew P, Ricketson I, Shackleton M, Vaillant F, Choi D, Li H and Eaves C (2006) Purification and unique properties of mammary epithelial stem cells. *Nature* **439**: 993-997.

Stricker P and Grueter R (1928) Action du lobe anterieur de l'hypophyse sur la montee laiteuse. *CR Soc Biol* **99**(1978): 80.

Takahashi K and Yamanaka S (2006) Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126**(4): 663-676.

Tallkvist J, Yagdiran Y, Danielsson L and Oskarsson A (2015) A model of secreting murine mammary epithelial HC11 cells comprising endogenous Bcrp/Abcg2 expression and function. *Cell Biol Toxicol* **31**: 111-120.

Tanos T, Rojo LJ, Echeverria P and Brisken C (2012) ER and PR signaling nodes during mammary gland development. *Breast Cancer Res* **14**(4): 210.

Taylor RA, Cowin PA, Cunha GR, Pera M, Trounson AO, Pedersen J and Risbridger GP (2006) Formation of human prostate tissue from embryonic stem cells. *Nat Meth* **3**(3): 179-181.

Thangaraju M, Rudelius M, Bierie B, Raffeld M, Sharan S, Hennighausen L, Huang A-M and Sterneck E (2005) C/EBPδ is a crucial regulator of pro-apoptotic gene expression during mammary gland involution. *Development* **132**(21): 4675-4685.

Theodorou V, Stark R, Menon S and Carroll JS (2012) GATA3 acts upstream of FOXA1 in mediating ESR1 binding by shaping enhancer accessibility. *Genome Res* **23**(1): 12-22.

Thorner AR, Parker JS, Hoadley KA and Perou CM (2010) Potential tumor suppressor role for the c-Myb oncogene in luminal breast cancer. *PLoS ONE* **5**:e13073.

Tonner E, Barber MC, Allan GJ, Beattie J, Webster J, Whitelaw CB and Flint DJ (2002) Insulin-like growth factor binding protein-5 (IGFBP-5) induces premature cell death in the mammary glands of transgenic mice. *Development* **129**(19): 4547-4557.

Toouli CD, Huschtscha LI, Neumann AA, Noble JR, Colgin LM, Hukku B and Reddel RR (2002) Comparison of human mammary epithelial cells immortalized by simian virus 40 T-Antigen or by the telomerase catalytic subunit. *Oncogene* **21**(1): 128-139.

Troester MA (2004) Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res* **64**: 4218-4226.

Turner NC, Reis-Filho JS, Russell AM, Springall RJ, Ryder K, Steele D, Savage K, Gillett CE, Schmitt FC, Ashworth A and Tutt AN (2006) BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene* **26**(14): 2126-2132.

Ubil E, Duan J, Pillai IC, Rosa-Garrido M, Wu Y, Bargiacchi F, Lu Y, Stanbouly S, Huang J, Rojas M, Vondriska TM, Stefani E and Deb A (2014) Mesenchymal-endothelial transition contributes to cardiac neovascularization. *Nature* **514**: 585-590.

Usary J, Llaca V, Karaca G, Presswala S, Karaca M, He X, Langerod A, Karesen R, Oh D, Dressler L, Lonning P, Strausberg R, Chanock S, Borresen-Dale A and Perou C (2004) Mutation of GATA3 in human breast tumors. *Oncogene* **23**: 7669-7678.

Vaillant F, Lindeman GJ and Visvader JE (2011) Jekyll or Hyde: does Matrigel provide a more or less physiological environment in mammary repopulating assays? *Breast Cancer Res* **13**(3): 108.

van Doorninck JH, van Der Wees J, Karis A, Goedknegt E, Engel JD, Coesmans M, Rutteman M, Grosveld F and De Zeeuw CI (1999) GATA-3 is involved in the development of serotonergic neurons in the caudal raphe nuclei. *J Neurosci* **19**(12): Rc12.

Van Esch H, Groenen P, Nesbit MA, Schuffenhauer S, Lichtner P, Vanderlinden G, Harding B, Beetz R, Bilous RW, Holdaway I, Shaw NJ, Fryns JP, Van de Ven W, Thakker RV and Devriendt K (2000) GATA3 haplo-insufficiency causes human HDR syndrome. *Nature* **406**(6794): 419-422.

Verbeke S, Richard E, Monceau E, Schmidt X, Rousseau B, Velasco V, Bernard D, Bonnefoi H, MacGrogan G and Iggo RD (2014) Humanization of the mouse mammary gland by replacement of the luminal layer with genetically engineered preneoplastic human cells. *Breast Cancer Res* **16**: 504.

Vierbuchen T, Ostermeier A, Pang ZP, Kokubu Y, Sudhof TC and Wernig M (2010) Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463**(7284): 1035-1041.

Vierbuchen T and Wernig M (2012) Molecular roadblocks for cellular reprogramming. *Mol Cell* **47**(6): 827-838.

Virmani A, Rathi A, Sathyanarayana U, Padar A, Huang C, Cunnigham H, Farinas A, Milchgrub S, Euhus D, Gilcrease M, Herman J, Minna J and Gazdar A (2001) Aberrant methylation of the adenomatous polyposis coli (APC) gene promoter 1A in breast and lung carcinomas. *Clin Cancer Res* **7**: 1998-2004.

Visvader ACR, Nai Yang F, Geoffrey JL and Jane E (2014) In situ identification of bipotent stem cells in the mammary gland. *Nature* **506**: 322-327.

Visvader JE (2009) Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev* **23**: 2563-2577.

Visvader JE, Crossley M, Hill J, Orkin SH and Adams JM (1995) The C-terminal zinc finger of GATA-1 or GATA-2 is sufficient to induce megakaryocytic differentiation of an early myeloid cell line. *Mol Cell Biol* **15**(2): 634-641.

Visvader JE and Stingl J (2014) Mammary stem cells and the differentiation hierarchy: current status and perspectives. *Genes Dev* **28**(11): 1143-1158.

Wang RH (2006) The new portrait of mammary gland stem cells. *Int J Biol Sci* **2**(4): 186-187.

Watson CJ (2006) Involution: apoptosis and tissue remodelling that convert the mammary gland from milk factory to a quiescent organ. *Breast Cancer Res* **8**(2): 203.

Watson CJ and Khaled WT (2008) Mammary development in the embryo and adult: a journey of morphogenesis and commitment. *Development* **135**(6): 995-1003.

Welshons WV, Wolf MF, Murphy CS and Jordan VC (1988) Estrogenic activity of phenol red. *Mol Cell Endocrinol* **57**: 169-178.

Williams RL, Hilton DJ, Pease S, Willson TA, Stewart CL, Gearing DP, Wagner EF, Metcalf D, Nicola NA and Gough NM (1988) Myeloid leukaemia inhibitory factor maintains the developmental potential of embryonic stem cells. *Nature* **336**(6200): 684-687.

Wilson B and Giguere V (2008) Meta-analysis of human cancer microarrays reveals GATA3 is integral to the estrogen receptor alpha pathway. *Mol Cancer* **7**: 49.

Wilson CL, Sims AH, Howell A, Miller CJ and Clarke RB (2006) Effects of oestrogen on gene expression in epithelium and stroma of normal human breast tissue. *Endocr Relat Cancer* **13**(2): 617-628.

Wu D, Sunkel B, Chen Z, Liu X, Ye Z, Li Q, Grenade C, Ke J, Zhang C, Chen H, Nephew KP, Huang TH, Liu Z, Jin VX and Wang Q. (2014) Three-tiered role of the pioneer factor GATA2 in promoting androgen-dependent gene expression in prostate cancer. *Nucleic Acids Res* **42**(6): 3607-3622.

Wu K, Li A, Rao M, Liu M, Dailey V, Yang Y, Di Vizio D, Wang C, Lisanti MP, Sauter G, Russell RG, Cvekl A and Pestell RG (2006) DACH1 is a cell fate determination factor that inhibits cyclin D1 and breast tumor growth. *Mol Cell Biol* **26**: 7116-7129.

Wyszomierski SL and Rosen JM (2001) Cooperative effects of STAT5 (signal transducer and activator of transcription 5) and C/EBPbeta (CCAAT/enhancer-binding protein-beta) on beta-casein gene transcription are mediated by the glucocorticoid receptor. *Mol Endocrinol* **15**(2): 228-240.

Xie X, Lu J, Kulbokas EJ, Golub TR, Mootha V, Lindblad-Toh K, Lander ES and Kellis M (2005) Systematic discovery of regulatory motifs in human promoters and 3' UTRs by comparison of several mammals. *Nature* **434**: 338-345.

Yan W, Cao QJ, Arenas RB, Bentley B and Shao R (2010) GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *J Biol Chem* **285**(18): 14042-14051.

Yang ZR, Thomson R, McNeil P and Esnouf RM (2005) RONN: the bio-basis function neural network technique applied to the detection of natively disordered regions in proteins. *Bioinformatics* **21**(16): 3369-3376.

Yap SKW, Guillermo AG, Magdalene M, Suzie V, Hayley LC, James GL, Robert GP, Nicholas AH, Zoltan N and Alpha S (2014) Cortical F-actin stabilization generates apical-lateral patterns of junctional contractility that integrate cells into epithelia. *Nature Cell Biol* **16**: 167-178.

Yeh S (2003) Abnormal mammary gland development and growth retardation in female mice and MCF7 breast cancer cells lacking androgen receptor. *J Exp Med* **198**: 1899-1908.

Yin P, Roqueiro D, Huang L, Owen JK, Xie A, Navarro A, Monsivais D, Coon V JS, Kim JJ, Dai Y and Bulun SE (2012) Genome-wide progesterone receptor binding: cell type-specific and shared mechanisms in T47D breast cancer cells and primary leiomyoma cells. *PLoS ONE* **7**:e29021.

Yoon N, Maresh E, Shen D, Elshimali Y, Apple S, Horvath S, Mah V, Bose S, Chia D, Chang H and Goodglick L (2010) Higher levels of GATA3 predict better survival in women with breast cancer. *Hum Pathol* **41**: 1794-1801.

Yu J, Vodyanik MA, Smuga-Otto K, Antosiewicz-Bourget J, Frane JL, Tian S, Nie J, Jonsdottir GA, Ruotti V, Stewart R, Slukvin II and Thomson JA (2007) Induced pluripotent stem cell lines derived from human somatic cells. *Science* **318**(5858): 1917-1920.

Zhou H, Wu S, Joo JY, Zhu S, Han DW, Lin T, Trauger S, Bien G, Yao S, Zhu Y, Siuzdak G, Scholer HR, Duan L and Ding S (2009) Generation of induced pluripotent stem cells using recombinant proteins. *Cell Stem Cell* **4**(5): 381-384.