UNIVERSITY OF CALIFORNIA, SAN DIEGO

Links between development and cancer: fetal mammary stem cells and human breast cancer

A dissertation submitted in partial satisfaction of the requirements for the degree Doctor of Philosophy

in

Biology

by

Dannielle Deanna Engle

Committee in charge:

Professor Geoffrey Wahl, Chair
Professor Juan Carlos Izpisúa Belmonte
Professor David Cheresh
Professor Colin Jamora
Professor Randall Johnson
Professor Karl Willert

2011

This Dissertation of Dannielle Deanna Engle is approved, and it is acceptable in quality and form for publication on microfilm and electronically:

_____

_____

_____

_____

_____

_____

Chair

University of California, San Diego

2011

DEDICATION

In memory of my father, Dea H. Engle, for always challenging and encouraging me, for often beating me at my own games, for never giving up, for his dedication to his family, and for his never ending love and support.

To my mother, Chin So Engle, for inspiring me to always see the good in life and people, for finding ways to help regardless of the situation, and for being the bravest woman I know.

To my sisters, for being courageous enough to disagree with me, for being supportive despite our differences, and for always expecting the best from me. To Sophie, the older sister who recognizes when I need help and always lends an ear to my grievances and for whom I print this dissertation in Helvetica; and to Charlene, the younger sister whose perspicacious ennui and irony has always inspired and infuriated me.

And to my husband, Hervé Tiriac, for his patience and support, for helping me find my way when I am lost, for his bravery to point out when I am wrong, and for his love despite my many faults.

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

aMaSC: Adult mammary stem cell

aMRU: Adult mammary repopulating unit

APC: Allophycocyanin

CD: Cluster of differentiation

CFC: Colony forming cell

CHO cells: Chinese hamster ovary cells

Cpm: Counts per minute

Cy: Cyanine

DAPI: 4’,6’-diamidino-2-phenylindole

DMEM: Dulbecco’s modified eagle media

DMSO: Dimethyl sulfoxide

E: Embryonic day

eGFP: Enhanced green fluorescent protein

ER: Estrogen receptor

ErbB2: Her2, neu

ES cell: Embryonic stem cell

FACS: Fluorescence activated cell sorting

FBS: Fetal bovine serum

FDR: False discovery rate

FIAU: Fialuridine

FITC: Fluorescein isothiocyanate

fMaSC: Fetal mammary stem cell

fMRU: Fetal mammary repopulating unit

fSTR: fStromal

fStromal: Fetal stromal

gDNA: Genomic DNA

HEK cells: Human embryonic kidney cells

HeLa cells: Cervical cancer cells derived from Henrietta Lacks

iPS cell: Induced pluripotent stem cell

K: Cytokeratin

LDTA: Limiting dilution transplantation analysis

LUM: Luminal

MARC: Modular activator/reporter cassette

MaSC: Mammary stem cell

MEF: Mouse embryonic fibroblast

miRNA: micro RNA

MRU: Mammary repopulating unit

MYO: Myoepithelial

NOD: Non-obese diabetic mice

NP40: Nonidet P-40

PBS: Phosphate buffered solution

PE: Phycoerythrin

PerCP: Peridinin chlorophyll protein complex

PFA: Paraformaldehyde

PR: Progesterone receptor

qRT-PCR: Quantitative real time polymerase chain reaction

rhEGF: Recombinant human epidermal growth factor

rhFGF: Recombinant human fibroblast growth factor

RMCE: Recombinase mediated cell exchange

rtTA: Reverse tetracycline transactivator

SAM: Significance analysis of microarrays

SCID: Severe combined immune deficiency

SFE: Sphere forming efficiency

ssDNA: Salmon sperm DNA

TK: Thymidine kinase

TRE: Tetracycline transactivator response elements

tTA: Tetracycline transactivator

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I thank Professor Geoffrey Wahl for challenging me as a scientist, for supporting me as a mentor, and for guidance as my friend.

I would also like to acknowledge all my colleagues in the Wahl lab. In particular I would like to thank the following people: Luo-wei Rose Rodewald for her expertise and guidance, Dr. Yao-cheng Leo Li for his insights into molecular cloning, Samantha Cheung, Justin La, and Stephanie Kinkel for their enthusiasm and assistance, Dr. Benjamin Spike for many inspiring conversations, and Dr. Jennifer Lin for technical assistance. In addition, I would like to thank Dr. Mark Wade and Dr. Vivian Wang for their help.

Several aspects of this work involved collaboration. Accordingly, I would like to thank C.M. Perou and A. Prat (University of North Carolina) for careful evaluation and helpful suggestions on bioinformatic analyses; T. Bonnefoix (Institut National de la Sante de la Recherche Medicale) for critical advice and help with LDTA; and G. Cunha and D. Medina for insightful discussions and technical instruction on mammary dissection and transplantation. I would also like to thank D. Chambers and J. Barrie (Salk - Center for Cytometry and Molecular Imaging), Y. Zheng (FACSAria II), N. Webster, J. Valencia and A. Schiller (UCSD-GeneChip Microarray Core); E. Kothari and Y. Zhu (UCSD Transgenic Core), J. Fitzpatrick and Y. Sigal (Salk – Waitt Advanced Biophotonics Core), and D. Medina, G. Smith, C. Eaves and P. Eirew for

critical reagents and protocols. Finally, I would like to thank Prometheus, Inc. and Fluidigm Corporation for our ongoing collaborations.

The work presented here would not have been possible without the following funding contributions: Breast Cancer Research Foundation (BCRF), Department of Defense: Breast Cancer Research Program (BCRP), G. Harold and Leila Y. Mathers Charitable Foundation, NIH Cell and Molecular Genetics training grant T32-GM07240, and California Breast Cancer Research Program pre-doctoral fellowship 15GB-0015. Support for core resources, including the cell sorter, LSR and qRT-PCR instruments, was provided by the Cancer Center Core grant NCI award 5P30CA014195.

Chapter 2, in full, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. * indicates equal contribution. The dissertation author was one of the primary authors of this paper.

Chapter 3, in full, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. * indicates equal contribution. The dissertation author was one of the primary authors of this paper.

Chapter 4, in part, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. * indicates equal contribution. The dissertation author was one of the primary authors of this paper.

Chapter 5, in full, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. * indicates equal contribution. The dissertation author was one of the primary authors of this paper.

VITA

2005 Bachelor of Arts, Northwestern University

2011 Doctor of Philosophy, University of California, San Diego

PUBLICATIONS

Wang, Y.V., Leblanc, M., Fox, N., Mao, J.-H., Tinkum, K.L., Krummel, K., Engle, D., Piwnica-Worms, D., Piwnica-Worms, H., Balmain, A.*, et al.* (2011). Fine-tuning p53 activity through C-terminal modification significantly contributes to HSC homeostasis and mouse radiosensitivity. Genes & Development *25*, 1426-1438.

ABSTRACT OF THE DISSERTATION

Links between development and cancer: fetal mammary stem cells and human breast cancer

by

Dannielle Deanna Engle

Doctor of Philosophy in Biology

University of California, San Diego, 2011

Professor Geoffrey Wahl, Chair

Parallels between stem cells, development and cancer have long been recognized, but few molecular links have been delineated. Here, we identify and characterize a population of fetal mammary stem cells (fMaSCs). fMaSC concentration peaked late in fetal development well after the emergence of the earliest mammary structures. Several characteristics of fMaSCs enabled more extensive purification than previously achieved using adult mammary. Single cells from the fMaSC-enriched population generated multi-lineage spheres.

These cells also revealed properties often associated with cancers. For example, we observed cells co-expressing luminal, myoepithelial and mesenchymal markers. Gene expression profiling revealed significant similarities with archived breast cancer arrays. The ErbB Pathway was heavily represented and fMaSC derived spheres were sensitive to ErbB kinase inhibition. The fMaSC expression profile was associated with aggressive breast cancer subtypes and consisted of unique gene modules with relevance for prognosis and the potential to indicate new therapeutic targets.

The parallels between development and cancer suggest that close relationships may exist between organ specific fetal stem cells and cancer. Here, we overcome the substantial technical challenges of such studies to identify, isolate and characterize fetal mammary stem cells (fMaSCs). We report the surprising finding that fMaSC activity is rare early and undergoes a significant increase late in fetal mammogenesis when the rudiments invade into the mesenchyme and fat pad. We find that regulatory pathways associated with fMaSCs are also differentially expressed in a subset of human breast cancers, suggesting that cancer cells may evolve to recapitulate the fetal stem cell state. Our characterization of the *fetal* mammary stem cell state provides a resource for generating new molecular hypotheses linking development and cancer, identifying candidate therapeutic targets, and developing new diagnostic and prognostic metrics.

Chapter 1. Introduction

Development, Stem Cells, and Cancer

Parallels between development, stem cells, and cancer have long been recognized. Indeed, the imperative of understanding the underlying developmental biology behind the tissue from which cancers arise was recognized as early as 1884 by W.S. Savory: "…before we shall ever be able to understand the question of why or how tumors form … we must be able to solve the problem of normal growth and development (Brewer et al., 2009)."

Most mammalian tissue development operates within a hierarchical organization in which tissue stem cells generate and maintain the organ. During normal development and homeostasis, tissue stem cells self-renew to preserve their numbers and generate lineage committed progenitors that expand to produce the multiple specialized cell types of the tissue (Morrison and Kimble, 2006). This process leads to many diverse cell types within the developing tissue. Similarly, tumors exhibit high levels of heterogeneity, both between different patients and within the tumors themselves. Studies analyzing teratocarcinomas in the mid-1800s raised the possibility that some cancers might originate from a dormant embryonic remnant that persists into adulthood. This embryonal rest hypothesis set forward by Durante and Conheim postulated that changes to these remnants and their environment leads to reactivation of the embryonic remnant and generation of an aberrant

tissue, or in other words, a tumor (Sell, 2010). This idea was further refined to propose that the elements that regain embryonic characteristics are responsible for the generation of malignant tumors (Brewer et al., 2009).

An alternative hypothesis postulates that all cancers arise from a maturation arrest of adult tissue stem cells, which prevents normal differentiation and disrupts the cellular hierarchy (Sell, 2010). Maturation arrest can occur at any point in the cellular hierarchy and controls the state of differentiation in the tumor (Sell, 2010). Arrest early in the hierarchy leads to relatively undifferentiated tumors while a later arrest larger recapitulates the cellular hierarchy and results in a well differentiated tumor maintained by a population of cancer stem cells (CSCs) (Sell, 2010). Periodic entry of such cells into aberrant differentiation programs could generate the cellular heterogeneity evident in a tumor (Sell, 2010). This model allows for the presence of a hierarchical lineage classification, but also includes the possibility of an arrest early in the hierarchy. Arrest within the tissue stem cell component of a tissue would lead to a tumor composed largely of CSCs (Sell, 2010).

Normal cells are progressively transformed through the process of clonal evolution. The progressive accumulation of mutations leads to the selective advantage of certain clones and the progression of disease (Dick, 2008) . Genetic variability acquired through clonal evolution within a tumor can lead to heterogeneity and interclonal cooperativity.

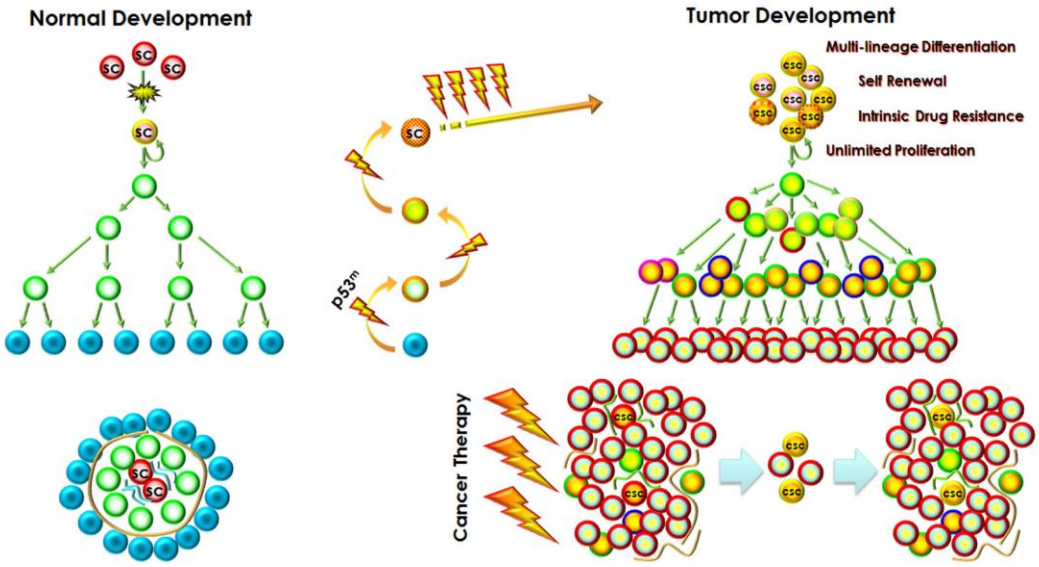


Figure 1-1. Left. Cellular hierarchy during development, with normal stem cells (SC) dividing symmetrically to maintain their numbers and asymmetrically to generate proliferating progenitors. These progenitors will ultimately give rise to the differentiated functional units of the tissue. **Right.** The cellular hierarchy during tumorigenesis is maintained by a population of cancer stem cells (CSCs). CSC identity can be acquired from any point in the cellular hierarchy. The relevance of cancer stem cells to patient therapy is also depicted. Upon acquisition of drug resistance phenotypes sometimes exhibited by normal tissue stem cells, the cancer stem cell population can survive chemo- and radiation therapy and result in tumor relapse.

The Cancer Stem Cell Hypothesis

Not all tumor cells are equal in their ability to grow both *in vitro* and *in vivo*, suggesting that there also exists a cellular hierarchy in cancer in which CSCs are responsible for the maintenance of the tumor (Figure 1-1) (Huntly and Gilliland, 2005a, b). This inequality within tumors was empirically demonstrated in humans in 1961. Investigators demonstrated that injection of one million tumor cells was needed to initiate new tumor formation in the patient from which the cells were isolated (Southam and Brunschwig, 1961). Further evaluation of tumor initiating activity of human cancer cells has been performed in the context of immune compromised mice. It is important to note that the immune system plays an important role in tumorigenesis; therefore, it is unclear how well these xenograft assays recapitulate what occurs in human patients. Studies using these xenograft analyses provided additional support for this hypothesis in which either tissue-specific stem cells are targeted for tumorigenesis or tumorigenic cells gain stem-like properties (Kleinsmith and Pierce, 1964). Subsequently, evidence accumulated in support of this model with the inception of multi-parameter, high throughput fluorescence activated cell sorting (FACS). This new technology allowed the identification and viable purification of distinct cell populations using fluorescently labeled antibodies to detect cell surface expression of different proteins. Upon isolation of different populations, their ability to initiate new tumors and recapitulate the primary

source's phenotypic heterogeneity could be tested in immune compromised xenograft systems.

This strategy was first applied to investigate the CSC hypothesis in hematopoietic malignancies (Bonnet and Dick, 1997) (Huntly and Gilliland, 2005b), in which only a subset of human cells from patients with acute myelogenous leukemia (AML) were able to initiate AML in immune compromised NOD/SCID mice. This subset of cells displayed a cluster of differentiation (CD) phenotype of $CD34^{+}CD38^{-}$. Interestingly, this cell surface expression pattern is associated with the hematopoietic stem cell compartment.

Later, the presence of cells with disproportionate ability to initiate tumors was demonstrated in solid cancers of the breast (Al-Hajj et al., 2003). In this case, the breast CSC populations isolated from eight out of nine patients examined exhibited a lineage negative, $CD44^{+}CD24^{low/-}$ cell surface phenotype and contained all tumor initiating ability in their xenograft model. Aldehyde dehydrogenase activity (ALDH) is another marker that has been demonstrated to further purify the tumor initiating activity within the $CD44^{+}CD24^{low/-}$ population (Croker et al., 2009; Ginestier et al., 2007). $ALDH1^{+}$ breast CSCs are more resistant to chemotherapy and are associated with poor patient outcome (Deng et al., 2010; Ginestier et al., 2007; Morimoto et al., 2009). If CSCs are pervasive among different cancers, then the properties and phenotypes these cells acquire would be expected to be

consistent. However, reproducible markers between different breast cancer patients are lacking. In a recent study of four estrogen receptor negative breast cancer patients, CD44 positivity correlated with tumor initiating ability in immune compromised mice. However, both CD24 positive and negative cells from the CD44 positive fraction could initiate tumors with equal capacity (Meyer et al., 2010). Instead, separation of tumor initiating activity came from fractionation of the $CD44^{+}CD49f^{+}CD133/2^{high}$ population.

The larger applicability of CD44 as a breast CSC marker may arise from its functional significance. CD44 is known to promote survival, invasion, and tumorigenecity and inhibition of its expression attenuates these properties (Bourguignon et al., 2008; Du et al., 2008; Godar et al., 2008; Subramaniam et al., 2007a; Subramaniam et al., 2007b). In addition, expression of CD44 correlates with lymph node metastasis and co-expression of basal and mesenchymal markers, an expression profile associated with basal-like breast cancer (Ricardo et al., 2011). The molecular signatures of the putative breast cancer stem cell populations expressing the markers CD44, CD24, and ALDH were evaluated. The signatures of the populations expressing individual markers or the $CD44^{+}CD24^{low/-}$ population did not yield statistically significant predictions in univariate survival analyses as individual markers (Ricardo et al., 2011). However, upon closer examination within the basal-like subtype of breast cancer, a $CD44^{+}CD24^{low/-}$ expression pattern within more than 10% of

the tumor cells identified patients with a worse prognosis (Ricardo et al., 2011).

A similar situation exists in melanoma, a disease in which detection of a CSC population is context dependent. Upon evaluation in only a partially immune compromised background, there exists only a small population of cells able to re-initiate the disease (Quintana et al., 2010; Quintana et al., 2008). However, upon examination in a more immune compromised model, the CSC frequency approaches 100% (Quintana et al., 2010; Quintana et al., 2008). Furthermore, within this more complete immune compromised model, marker expression of melanoma cells did not correspond to heightened tumor initiating activity (Quintana et al., 2010; Quintana et al., 2008). These results highlight the difficulty in comparing tumor initiating activity between different approaches and emphasize the importance of the immune system in xenograft tumor initiating activity. Altogether, these data suggest that while most cancers are fueled by a population of cancer stem cells, only some are driven by a hierarchical organization (Al-Hajj et al., 2003; Bonnet and Dick, 1997; Clarke et al., 2006). Regardless, the need for functional markers of these populations is paramount.

CSCs have been proposed to retain or adopt some of the putative properties of stem cells such as self-renewal, vast proliferative potential, intrinsic drug resistance, ability to exist in a dormant state, and many others. Unfortunately, these characteristics are acquired without the strict regulation

that accompanies these attributes in the normal setting. Along these lines, reports have identified subpopulations of highly tumorigenic breast CSCs that have displayed chemo-resistance and radio-resistance (Li et al., 2008; Zhang et al., 2008). In light of these findings, it is possible that the presence of a CSC population would persist during treatment of a cancer patient and regenerate the tumor after the therapeutic course is completed (Figure 1-1).

Although normal tissue stem cells are often a rare cell type, CSCs need not follow the same strict cellular hierarchies and cell number regulation. In some cases, such as melanoma, it has been found that the cells responsible for tumor perpetuation can represent the entirety of the tumor (Quintana et al., 2008). Furthermore, although the designation of “cancer stem cell” may seem to indicate a cell of origin, this is not necessarily the case. Whether these cells originate from a stem cell, progenitor, or differentiated cell is unclear and may not be consistent between different tumor types. In acute myelogenous leukemia, the initiating 8;21 translocation that results in the chimeric transcript of AML1/ETO is first detected in hematopoietic stem cells. Although this translocation has been shown to be necessary, it is not sufficient to induce transformation in this stem cell pool (Reya et al., 2001). Indeed, further mutations are required within the progenitor compartment in order to generate a fully transformed population of leukemic stem cells. In the case of blast crisis in chronic myelogenous leukemia (CML), granulocyte-macrophage progenitors appear to be the target of transformation in the generation of a leukemic stem

cell population (Jamieson et al., 2004). The tumorigenic subset of brain tumor cells exhibit the same surface marker expression used to enrich for neural stem cells, CD133 (Singh et al., 2004). Furthermore, these brain CSCs are able to form neurospheres, an assay used to functionally evaluate neural stem cell activity. However, whether this coincidence is meaningful is unclear as the reliability of CD133 as a stem cell marker is in question in other systems. For instance, both CD133+ and CD133- cells from melanoma are able to recapitulate the heterogeneity of the original tumor (Quintana et al., 2008). A similar situation occurs with the expression of JARID1B, a histone demethylase, which correlates with tumorigenecity in melanoma. The expression of JARID1B was unstable and often fluctuated such that single cells could switch their expression status (Roesch et al., 2010).Therefore, it is possible that the search for a CSC should focus on identifying a dynamic and transient state instead of a stable cell type.

Microarray analyses of certain cancers, such as basal-like breast and lung cancers, show a strong enrichment for gene sets found in embryonic stem cells (ES) and induced pluripotent stem cells (iPS) (Ben-Porath et al., 2008; Mizuno et al., 2010). The large contribution of ES and iPS cell genes to these cancer signatures suggests that they have a significant contribution from undifferentiated cell types. The genes underlying these “stemness” signatures transcend the proliferative programs that drive the rapid cell cycles of these stem cells (Mizuno et al., 2010) and do not seem to comprise the core ES/iPS

pluripotency factors (Ben-Porath et al., 2008). Therefore, it is possible that the genes expressed by these cancers represent genes associated with other aspects of stem cells. Further insights into the pathways that underlie the stem cell signatures present in specific types of cancer will likely require more detailed knowledge of the tissue-specific stem cells responsible for the development and maintenance of the corresponding organ. Indeed, recent studies show that cancer-relevant mutations in the stem cells of the small intestine lead to the development of large adenomas, while the same mutation in proliferative progenitors generates micro-adenomas that rarely progress (Barker et al., 2009).

There are two predominant models that explain tumor heterogeneity while allowing for the existence of CSCs (Dick, 2008). The stochastic model postulates that all tumor cells have equal probability to become a CSC. In this way, the tumor is biologically homogeneous. The low frequency of tumor cells that are able to perpetuate and initiate cancer is explained by the influence of a culmination of different factors that can bestow and revoke stem-like properties onto cells, thereby establishing functional heterogeneity (Dick, 2008).

On the other hand, the hierarchy model hypothesizes that tumorigenesis is an abnormal re-activation of development in which CSCs maintain themselves and the rest of the “tissue.” In this model, only a specific subset of tumor cells has the potential to behave like CSCs (Dick, 2008). In

order to discern which model is accurate, cell fractions must be isolated prospectively and evaluated by their ability to generate new disease clonally. In the case of melanoma, it seems like the stochastic model prevails as CSC activity is observed in any fraction and cannot be isolated prospectively. Therefore, in this type of cancer, all tumor cells have the same potential for tumor initiation. However, in the case of some leukemias, it seems that the hierarchical model is correct as it is possible to fractionate cells into populations in ways that also segregate CSC activity. Therefore, the relevance of these models depends on the type of tumor being evaluated and it is also possible that both of these models may be operating within the same tumor.

All of these models of tumorigenesis are not mutually exclusive. Indeed, they can be integrated to account for the diverse phenotypes seen in cancer. The accumulation of mutations by clonal evolution can lead to the reactivation of embryonic programs in different cells, leading to a maturation arrest at any level in the cellular hierarchy. Arrest early in the hierarchy can lead to a cancer predominantly composed of CSCs, whose ability to perpetuate the tumor is reliant on stochastic factors such as the immune system. Arrest later in the hierarchy can lead to the aberrant recapitulation of the cellular hierarchy in which only the CSCs have the ability of perpetuating the disease. Understanding how this cellular hierarchy is operates during development will be crucial for determining how it is perturbed during tumorigenesis.

Embryonic Mammogenesis

Fetal mammary development involves proliferation, migration and invasion into the stroma. Detectable mouse mammogenesis begins on embryonic day 11.5 (E11.5) with thickening of the ventral ectoderm into five pairs of mammary placodes (Figure 1-2) (Veltmaat et al., 2003). The following developmental paradigm proceeds asynchronously, with the third pair of placodes forming first, followed by pairs 4, 5 and 1, and finally 2. By E12.5, the placodes become spherical buds and a condensed layer of mammary mesenchyme has surrounded the invaginating bud by E13.5. Sexual differentiation of the mammary rudiments begins at E13.5 with the androgen dependent degeneration of the male rudiments, leading to their eventual disconnection from the overlying epidermis. Elongated ductal sprouts are observed in females by E16 (Veltmaat et al., 2003) and after breaking through the condensed mammary mesenchyme, invasion of the extending rudiment into the fat pad precursor begins by E16.5. At E18.5, the mammary rudiment consists of a primitive branched network within the developing mammary fat pad (Veltmaat et al., 2003). These cellular processes are similar to those observed during breast tumorigenesis and metastasis, but are rarely observed in the adult gland outside of pregnancy.

Specific examples of the processes mirrored between mammary development and tumorigenesis include signal transduction pathways that are activated during development and often co-opted during cancer progression

(Beachy et al., 2004). As an example, Wnt signaling is necessary for the initiation of mammary gland development in the mouse. This was demonstrated by the inhibition of Wnt signaling, for example through the knockout of the transcription factor Lef1 or the over-expression of the negative regulator Dkk1, which precludes the formation of this tissue (van Genderen et al., 1994) (Boras-Granic et al., 2006; Chu et al., 2004). It is important to note that each pair of mammary glands differs slightly in their reliance on particular signaling pathways. In the case of Lef1 knockout, pair number 4 initiates, although they subsequently regress (van Genderen et al., 1994). In addition to the importance of Wnt signaling to development, its relevance to cancer was demonstrated by its discovery as one of the first breast cancer proto-oncogenes by the frequent pro-viral integration of the mouse mammary tumor virus (MMTV) into this locus, resulting in over-expression of this proto-oncogene and breast tumor development (Nusse and Varmus, 1982). The connections between embryonic development and cancer continue as many embryonic transcription factors have also been shown to promote breast tumorigenesis (Briegel, 2006). In addition, splice isoforms found in the embryo are often found in cancer. For instance, the V7 isoform of CD44 correlates with breast CSCs (Ponta et al., 1998). Another example of this is the M2 embryonic splice isoform of Pyruvate Kinase, which is associated with the altered metabolism of many cancers (Christofk et al., 2008). Finally, many embryonic miRNAs have been identified in cancers (Nimmo and Slack, 2009). Together,

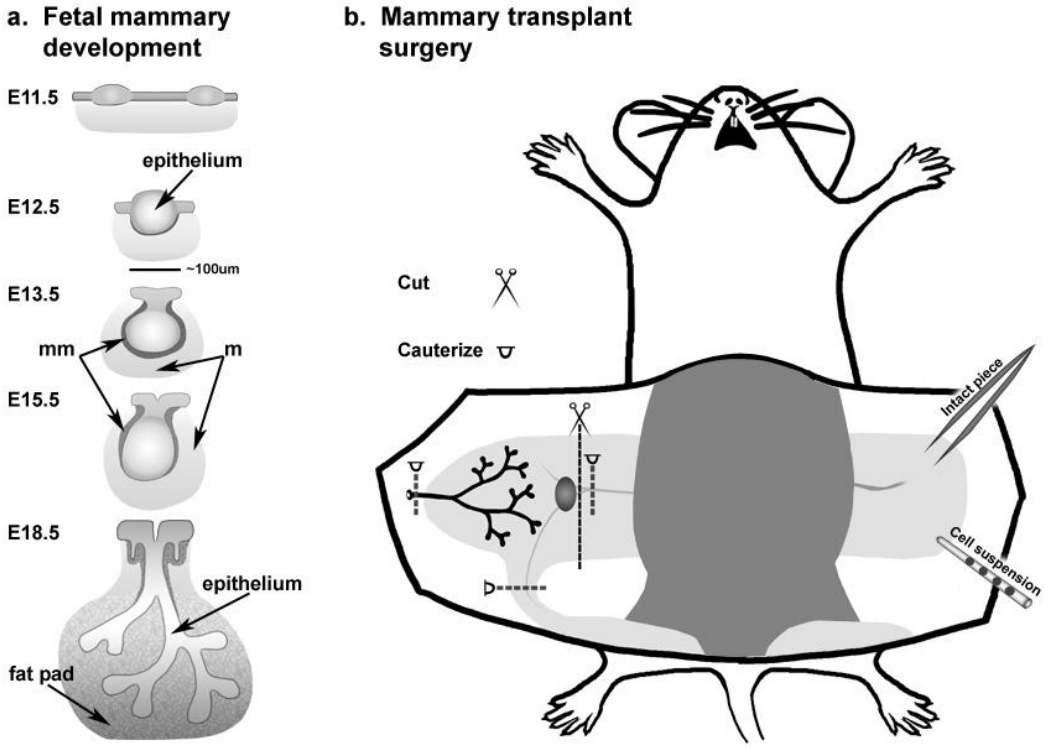


Figure 1-2. A. Embryonic mammogenesis (m: dermal mesenchyme, mm: condensed mammary mesenchyme)**. B.** The mammary transplantation procedure**.** By removing the fat pad from the major lymph node to the nipple in a 21 day old mouse (P21), all of the endogenous epithelial ducts can be removed. This leaves an empty fat pad as an orthotopic transplantation site for either implantation of tissue fragments of injection of cell suspensions.

these observations provide a compelling rationale for determining whether specific types of breast cancer share properties of normal mammary stem cells (MaSCs) during embryonic mammogenesis.

Mammary Stem Cells

The massive cellular proliferation and destruction seen in the mammary gland during each cycle of pregnancy suggests the existence of a robust MaSC population. MaSCs are defined by their ability to self-renew, generate the multiple lineages of the tissue, and continuously repopulate the functional units of the organ in which they reside. These cells have extensive proliferative potential, but are ultimately restricted in their ability to divide. Adult mouse MaSCs were first inferred from landmark studies demonstrating serial regeneration of functional mammary glands from transplanted adult mammary fragments (Figure 1-2) (Deome et al., 1959). These studies were later extended to transplantation of cell suspensions into the number four fat pad after removal of its endogenous epithelial ducts (Figure 1-2) (Smith, 1996). It was demonstrated that MaSCs are dispersed throughout the mammary gland by transplanting random fragments from different locations and that the mammary gland retains MaSC activity throughout the lifespan of the organism (Smith and Medina, 1988). MaSCs are quite rare in the adult, representing 1/1,400 – 1/4,900 total mammary cells when evaluated in syngeneic transplant models using limiting dilution transplantation analyses (LDTA) (Shackleton et

al., 2006; Smith, 1996; Stingl et al., 2006b). The number of MaSCs is thought to remain constant except upon pregnancy, when an eleven fold increase in their numbers can be measured at mid-pregnancy by LDTA, after which their numbers begin to decline and return to normal levels ten days after involution begins (Asselin-Labat et al., 2010; Smith and Boulanger, 2003). In addition, ovarectomy decreases the number of MaSCs by approximately four fold. Altogether, these data demonstrate the steroid hormone regulation of MaSCs and upon further evaluation, reveal the importance of the microenvironment, such as estrogen receptor (ER) positive luminal progenitors, in mediating these changes in ER negative MaSC numbers (Asselin-Labat et al., 2010). In contrast, the embryonic mammary gland develops even in the absence of both estrogen receptors or the progesterone receptor (PR) and estrogen delivery from the mother is thought to be prevented due to its capture by α-fetoprotein (Veltmaat et al., 2003). At the same time, embryonic mammary rudiments cultured *in vitro* show a clear response to estrogen treatment (Veltmaat et al., 2003). Therefore, it is unclear whether steroid hormone signaling affects embryonic mammogenesis in ways similar to its effect on adult MaSCs.

Stem cells by definition can self-renew and generate the multiple lineages of their respective tissue. The ability of MaSCs to fulfill these criteria is evaluated using transplantation analyses. Upon transplantation, MaSCs are capable of generating of both epithelial lineages of the gland (myoepithelial muscle-like cells and milk-producing luminal epithelial cells) and fully filling the

fat pad. The progeny of the MaSCs also produce alveolar structures and secrete milk upon pregnancy, and ultimately recapitulate the phenotypic heterogeneity of a nulliparous mammary gland. MaSCs are capable of repeating this entire regenerative process serially for 4 to 7 generations (Daniel et al., 1968). Interestingly, this is a key difference between normal and cancer cells as unlike their normal counterpart, cancer cells have unlimited proliferative potential in this mammary transplantation model (Smith, 1996).

MaSCs have been significantly enriched by flow sorting with cell surface markers such as CD24 (Heat Stable Antigen), α6 integrin (CD49f) and β1 integrin (CD29) (Shackleton et al., 2006; Stingl et al., 2006b). Typically, such enrichments result in ~1/60 cells being a stem cell after depletion of hematopoietic and endothelial cell lineages as estimated by LDTA. This enrichment strategy enabled the regeneration of an entire functional mammary gland from the transplantation of a single cell from these enriched populations of adult MaSCs (aMaSCs) (Shackleton et al., 2006; Stingl et al., 2006b). Although this represents a significant enrichment over unsorted mammary, it has not been possible to definitively identify and characterize the repopulating cell within such a heterogeneous population using existing techniques. Indeed, it remains unclear whether aMaSCs are a homogeneous population, or whether this tissue is similar to the blood, hair follicle, and intestine, which contain separate subsets of quiescent and cycling stem cells (Barker et al.,

2007; Jaks et al., 2008; Sangiorgi and Capecchi, 2008; Tumbar et al., 2004; Wilson et al., 2008; Wilson et al., 2007).

Fetal MaSCs (fMaSCs) have been suggested to exist as intact mammary rudiments obtained from as early as embryonic day 12.5 (E12.5) can reconstitute the mammary gland (Sakakura et al., 1979). In addition, experiments recombining embryonic mammary epithelial buds with salivary mesenchyme demonstrated that mammary cell identity is acquired very early in mammogenesis as their cell fate cannot be redirected into the salivary lineage (Sakakura et al., 1976). However, direct evidence that such structures contain specific cells able to regenerate functional mammary glands, let alone their quantification, purification, or molecular characterization, has yet to be presented. This represents a substantial gap in our understanding of mammary biology as it relates to breast cancer. The experiments presented here define the kinetics of MaSC activity during development and also reveal novel molecular links between fMaSCs and human breast cancer.

Chapter 2. Mammary stem cell activity peaks at E18.5 and serves as a naturally enriched source of mammary stem cells.

Background

Classic studies demonstrated that intact embryonic mammary epithelium from E13 to E17 generate mammary outgrowths when multiple epithelial rudiments were implanted within the epithelial-divested fat pads of recipient adult mice (Sakakura et al., 1979). Although these studies suggest the presence of fetal mammary stem cells (fMaSCs), it is not possible to accurately measure the number of stem cells from this type of experiment. Transplanted intact rudiments may not contain intrinsic MaSC activity at the initial point of transplantation. Instead, it is possible that these rudiments continue to develop normally until MaSCs become specified at a later time point. MaSC quantification requires the transplantation of dissociated cell populations at limiting numbers in order to determine whether individual cells have the intrinsic ability to regenerate a mammary gland. This type of analysis is referred to as limiting dilution transplantation analysis, or LDTA. In our studies, we refer to the units able to generate full, functional mammary outgrowths as mammary repopulating units (MRU) (Bonnefoix et al., 1996; Stingl et al., 2006b). According to this terminology, the stem cell is one cell, but not necessarily the only cell type, within the MRU.

Previous studies have focused on identification, enrichment, and characterization of adult mammary stem cells (aMaSCs) and therefore many of the tools we will utilize have already been validated on adult MaSC-enriched populations. aMaSCs have been enriched using hematopoietic and endothelial lineage depletion followed by sorting for CD24 and either CD49f or CD29 (Shackleton et al., 2006; Stingl et al., 2006a). Utilizing CD24 and CD29, the population enriched with aMaSCs reside in the $CD24^{+}CD29^{high}$ fraction at a frequency of about 1 in 590 or 1 in 64 upon double-sorting, however, MaSC activity was still detectable in two other fractions, the $CD24^{-}CD29^{high}$ and $CD24^{-}CD29^{low}$ populations, albeit at frequencies 5 – 250 fold lower than in the designated aMaSC-enriched population (Shackleton et al., 2006). When using a different fractionation strategy consisting of CD24 and CD49f expression after lineage depletion, the $CD24^{med}CD49f^{high}$ population has the highest MaSC frequency, which was estimated to be 1 in 60 cells (Stingl et al., 2006a). However, the $CD24^{low}CD49f^{low}$ designated population also contains stem cell activity, although this activity is approximately 50 fold more rare (Stingl et al., 2006a). Fractionations using these markers results in significant enrichment relative to unsorted adult mammary cells, but these markers are also expressed extensively in the surrounding stroma and terminally differentiated epithelial cells (Shackleton et al., 2006; Stingl et al., 2006a). In addition, upon knock out of these genes in mice, the mammary gland is still able to form, indicating that these markers do not have functional relevance for mammary

gland development. Specifically, upon knock out of CD24 in mice, not only does the mammary gland still form and function, the extent of branching was significantly elevated in the knockout mice (Cremers et al., 2010; Nielsen et al., 1997). In addition, targeted ablation of β1 Integrin in the mammary gland did not affect its development (White et al., 2004).

Although enrichment for MaSCs has been achieved in the adult gland, segregation of all MaSC activity into one population has not been possible. In addition, the characterized populations are heavily contaminated with both committed myoepithelial and stromal cells. Therefore, comparisons between two different functional groups have been obscured on more than one front; differentiated cell contamination and lack of distinguishing features.

Results

In 1979, Sakakura demonstrated that embryonic mammary rudiments could regenerate an entire mammary gland. However, these studies utilized transplantation of multiple rudiments (3-5) into each fat pad (Sakakura, 1979). We show that transplanting *single* intact mammary rudiments consisting of both epithelial and stromal compartments from E12.5 to E18.5 repopulated fat pads with complete and functional mammary ductal outgrowths at very high efficiency (70-100%, Figure 2-1, Table 2-1). In contrast, tissue fragments from E11.5 that were positionally equivalent to the thoracic or inguinal mammary rudiments rarely generated outgrowths, and E10.5 fragments failed to

regenerate mammary glands (Figure 2-1A-B and Table 2-1). These data demonstrate that single intact mammary anlagen from as early as E12.5 can regenerate the mammary gland, suggesting the existence of at least one MRU per rudiment.

The onset of limited regenerative ability of intact mammary rudiments at E11.5 and robust regenerative ability at E12.5 led us to hypothesize that fetal MRUs (fMRUs) may arise between these times. Whole rudiment transplant studies cannot be used to determine the frequency of fetal mammary repopulating units (fMRUs) during embryonic development. Therefore, we determined the MRU frequency during fetal mammary development by performing LDTA with cell suspensions derived from rudiments at different developmental stages. Surprisingly, we were unable to detect fMRUs in *dissociated*, bulk mammary cell populations obtained on or before E15.5 despite transplantation of cell populations equivalent to multiple rudiments (i.e., up to 10,000 viable cells/fat pad) in 2-3 independent experiments (Figure 2-2 and Table 2-2 column 3). It is possible that dissociating the rudiments and transplanting cell suspensions affects measurements of MRU abundance, especially if contextual cues, analogous to niche interactions in other systems (Scadden, 2006), affect stem cell activity. However, identically dispersed bulk cells from E18.5 rudiments generated mammary outgrowths from as few as 100 cells. This striking difference indicates that a significant increase in MRU

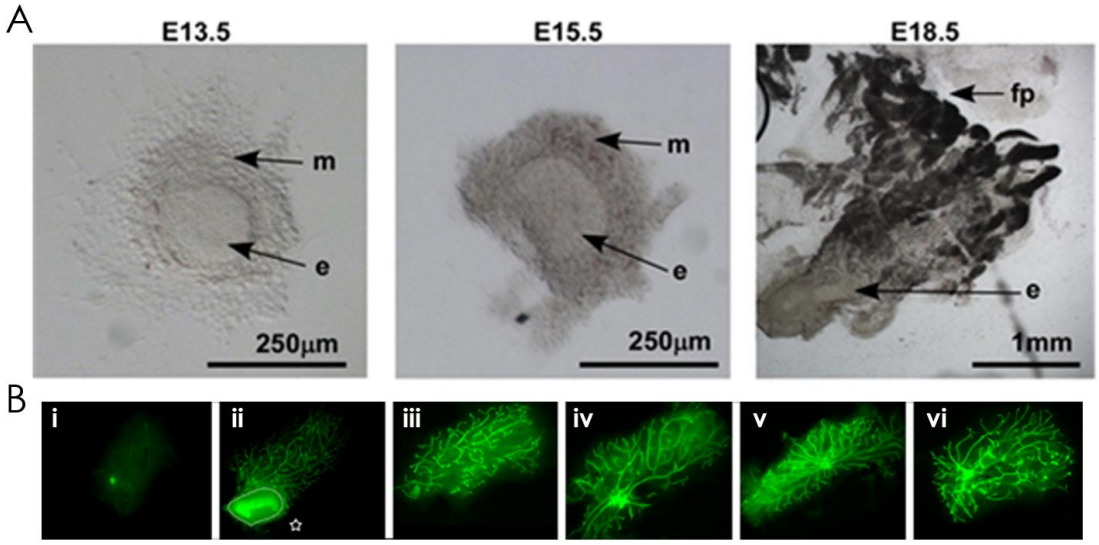


Figure 2-1. **A.** Brightfield images of dissected intact rudiments from E13.5, E15.5 and E18.5 indicating contribution by: m=mesenchyme, e=epithelium, fp=fat pad identified according to morphology. **B.** Intact fetal pieces including mammary epithelium and associated stroma (from actin-eGFP+ donors) can repopulate cleared fat pads with mammary epithelial outgrowths *in vivo*. Transplanting intact pieces from E10.5 did not result in mammary gland reconstitution (i), while pieces with mammary epithelium plus associated stroma from developmental stage E11.5 (ii), E12.5 (iii), E13.5 (iv), E15.5 (v), and E18.5 (vi) resulted in reconstitution of mammary gland and sometimes the growth of hair (ii). The star symbol (ii) indicates levels adjusted separately for outlined area due to overexposure.

Table 2-1.

Mammary outgrowths from intact Actin-eGFP+ fetal tissues including mammary rudiments.

Fetal stage	Repopulated fat pad	Total fat pads transplanted	Transplant efficiency	Engraftment
E10.5	0	9	0	100%
E11.5	3	24	13%	100%
E12.5	5	7	70%	100%
E13.5	8	11	72%	N.D.
E15.5	10	10	100%	N.D.
E18.5	11	12	92%	N.D.

Engraftment = successful donor tissue implantation determined by visualization of GFP-expressing tissues in non-GFP expressing fat pads 6-10 weeks post-surgery. N.D. = not determined.

frequency, as measured by LDTA, occurs during fetal development (Figure 2-2 and Table 2-2 column 3).

The quantification of MaSC activity using LDTA is based on a linear model assuming a single-hit Poisson distribution (Bonnefoix et al., 1996). A single-hit assumption indicates that the regeneration of the mammary gland was due to clonal regrowth by a single cell. Use of this model does not always generate reliable estimations depending on the frequency of successful outgrowths at limiting cell numbers (Bonnefoix and Callanan, 2009). We note which frequencies are accurate statistical estimations and also recognize the likely contribution of other extrinsic factors and cell types to stem cell activity despite the one hit assumption.

The dissimilarity between the results of intact and dissociated rudiment transplantation analyses raised the possibility that dissociation may have disrupted interactions critical for stem cell function. As Matrigel, a reconstituted basement membrane matrix rich with laminin and collagen, increases transplantation efficiency of adult mammary epithelial cells (Lim et al., 2009), we evaluated its effects on transplantation of dissociated embryonic mammary cells. Matrigel greatly increased the estimated MRU frequency of adult mammary cells, resulting in a 100 fold increase within bulk, unsorted cells and a 25 fold increase within the adult MYO population (Table 2-3). Interestingly, we also detected repopulating activity within the luminal CFC population, but the MRU frequency showed little change upon addition of Matrigel (Table 2-3).

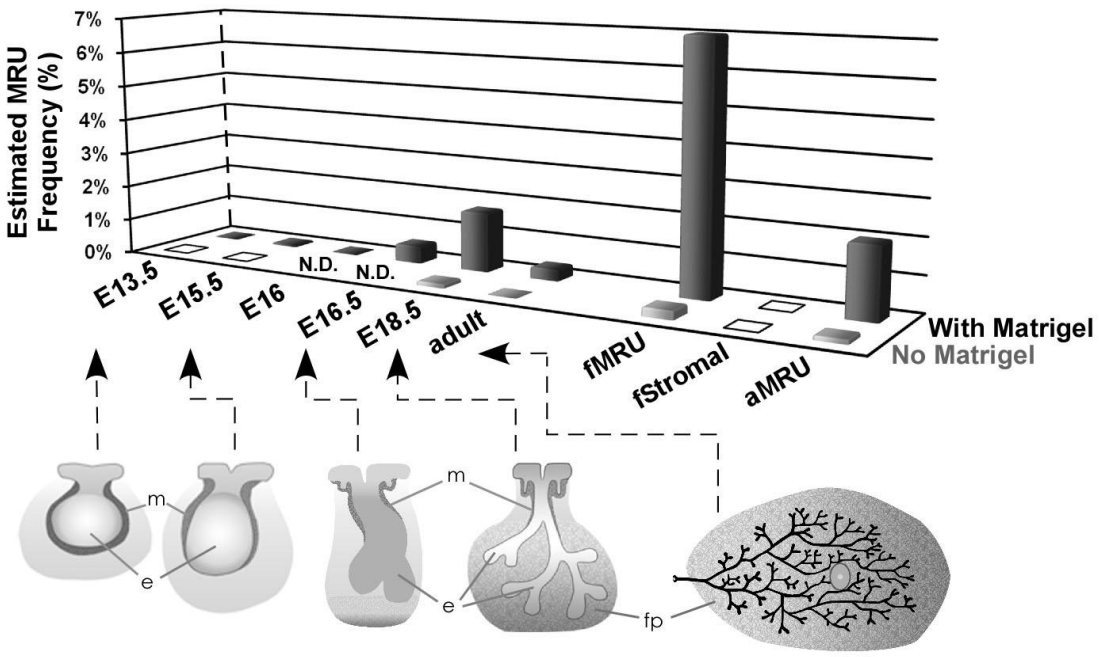


Figure 2-2. Bar chart showing MRU frequency estimates at various stages of fetal development and in the adult in the presence or absence of Matrigel. Empty squares indicate no MRUs detected. N.D. = not determined. Gross morphological appearance of the gland at various stages is illustrated. e=epithelium, m=mesenchyme and fp=fat pad.

Table 2-2. Dissociated cells from fetal mammary rudiments are enriched in mammary repopulating units late in embryogenesis.

p values > 0.05 indicate the data are consistent with single-hit Poisson model
† Rough MRU frequency estimate (see Statistical Analyses under Methods for detail). N.D.= not determined.

Stage	Cells injected per fat pad	Repopulated fat pad #	
		No Matrigel	Co-injected with Matrigel
E13.5	25	N.D.	0/2
	100	0/4	0/8
	1000	0/4	2/6
	10000	0/11	4/9
	MRU freq. (95% CI)	**<1/40,000**	**1/12,000 (1/5000 - 1/30,000)**
	p Value		0.2
E15.5	10	N.D.	0/4
	100	N.D.	0/11
	500	0/5	4/9
	1000	0/10	2/7
	MRU freq. (95% CI)	**<1/4,000**	**1/1,800 (1/800-1/4,000)**
	p Value		0.9
E16	50	N.D.	0/6
	100	N.D.	0/6
	500	N.D.	1/6
	1000	N.D.	1/6
	MRU freq. (95% CI)		**1/4,600† (1/1,000 - 1/18,000)**
	p Value		0.8
E16.5	50	N.D.	2/3
	100	N.D.	3/9
	1000	N.D.	6/6
	MRU freq. (95% CI)		**1/200 (1/70 - 1/400)**
	p Value		0.68
E18.5	10	N.D.	3/15
	20	N.D.	2/11
	25	N.D.	10/27
	50	0/6	18/30
	100	1/5	13/17
	1000	5/8	12/12
	2000	4/4	N.D.
	MRU freq. (95% CI)	**1/830 (1/400 - 1/1,600)**	**1/60 (1/40 - 1/80)**
	p Value	0.8	0.7
Adult	10	N.D.	0/10
	100	0/11	2/9
	500	N.D.	8/9
	1,000	2/17	8/8
	10,000	2/11	2/2
	20,000	N.D.	4/4
	MRU freq. (95% CI)	**1/30,000† (1/11,000-1/80,000)**	**1/300 (1/150 - 1/500)**
	p Value	0.2	0.2

Table 2-3. Limiting dilution mammary transplantations of freshly dissociated and fractionated adult mammary cells. Note: CB17-SCID recipients were used for all transplants, except where indicated.

Fractionated mammary cells	Cells injected per fat pad	Repopulated fat pad #	
		No Matrigel	co-injected with Matrigel
Adult MYO ($CD24^{low}CD49f^{low}$) (Stingl et al., 2006b)	10	1/2	1/2
	50	0/4	2/2
	100	0/5	3/5
	500	0/2	4/4
	1,000	3/6	6/6
	MRU freq. (95% CI)	**1/1,500† (1/500 - 1/4,000)**	**1/60 (1/25 - 1/150)**
	p Value	0.2	0.18
Adult LUM CFC ($CD24^{high}CD49f^{low}$) (Stingl et al., 2006b)	50	0/2	0/2
	100	0/4	1/6
	1,000	1/4	0/5
	2,000	0/2	N.D.
	MRU freq. (95% CI)	**1/8,000† (1/1,000 - 1/58,000)**	**1/5700† (1/700 - 1/43,000)**
	p Value	0.8	0.06
Adult MRU-enriched population ($CD24^{med}CD49f^{high}$)* (Stingl et al., 2006b)	20	4/7	N.D.
	50	4/6	N.D.
	100	4/6	N.D.
	MRU freq. (95% CI)	**1/50† (1/30-1/100)**	
	p Value	0.07	

p values > 0.05 indicate the data are consistent with single-hit Poisson model.
† Rough MRU frequency estimate (see Statistical Analyses under Methods for detail).
N.D. = not determined.
* Performed syngeneic mammary transplantations using FVB immune-competent hosts.

We found that Matrigel also increased the sensitivity of fMRU detection such that we were able to detect MRU activity as early as E13.5 (Figure 2-2 and Table 2-2 column 4). However, even with Matrigel, MRU frequency prior to E16.5 remained so low that we estimate that there may be, at most, one MRU per rudiment at E13.5-E15.5 (Figure 2-2 and Table 2-2 column 4). Interestingly, the narrow developmental window between E15.5-E16.5 revealed a dramatic inflection point in transplant efficiency (>9-fold). This period corresponds to the time when the elongating rudimentary duct invades through the adjacent mammary mesenchyme and into the fat pad precursor (Veltmaat et al., 2003). Matrigel also increased the transplantation efficiency of E18.5 dissociated bulk mammary cells, increasing MRU frequency nearly 15-fold to 1 in 60 (Figure 2-2 and Table 2-2 column 4). These data are the first quantitative measurements of MRU frequency during embryogenesis. They demonstrate that fully functional MaSCs are rare early in mammogenesis (E13.5-E16). Our results show that MaSC activity dramatically increases later in embryogenesis (i.e. after E16.5) and reaches a peak in the unfractionated, dissociated E18.5 rudiment, when the MRU frequency is ~40-fold higher than in the bulk adult mammary population when evaluated in the absence of Matrigel and still ~5-fold higher in its presence (Table 2-2). These observations indicate that the E18.5 mammary rudiment is a naturally enriched source of MRUs.

Expression of CD24 and CD49f facilitate fMRU enrichment.

The high concentration of fMRUs in the E18.5 mammary rudiment served as a natural enrichment step. In order to further purify fMRUs, we determined whether the strategy first pioneered in the hematopoietic system and later adopted in the adult mammary gland would be applicable in the embryonic mammary gland. Enrichment of MRUs from the adult mammary gland involves cell sorting based on expression of cell surface markers such as CD24 and CD49f (Stingl et al., 2006b). We detected expression of these markers in both the stromal and epithelial components of E13.5 and E15.5 mammary rudiments. In contrast, high expression of CD24 and CD49f at E18.5 identify basal-epithelial staining with very little staining in the fat pad and surrounding mesenchyme (Figure 2-3). Consistent with *in situ* immunofluorescence, flow cytometric analyses demonstrate that most cells from E13.5 and E15.5 mammary rudiments express CD24 and CD49f (Figure 2-4). These markers delineate a distinct subpopulation at E18.5 that comprises approximately 5% of mammary cells following endothelial and hematopoietic lineage exclusion (Figure 2-4). We examined co-processed eGFP negative adult and eGFP positive fetal mammary tissue and discovered that the distinct E18.5 population expresses higher levels of CD24 than the aMRU-enriched population ($CD24^{med}CD49f^{high}$) and higher levels of CD49f than the adult luminal CFC population (Colony Forming Cells, $CD24^{high}CD49f^{low}$) (Figure 2-5 and 2-6) (Stingl et al., 2006b). Based on this

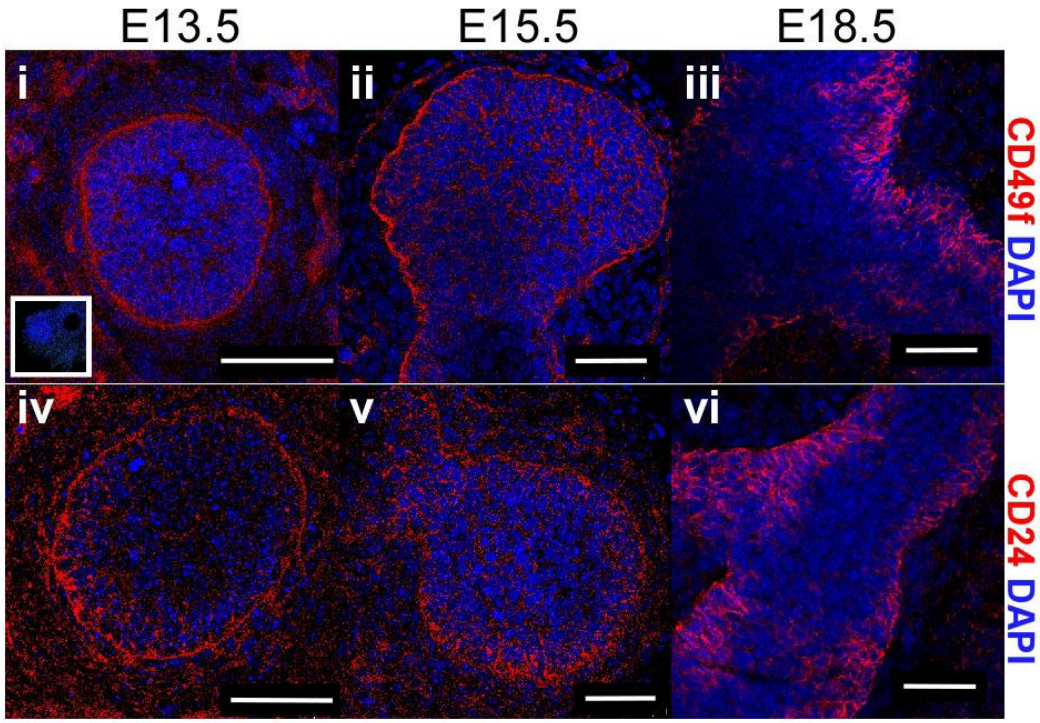


Figure 2-3. Confocal images of immunofluorescence on fetal mammary whole mounts showing the expression of CD49f (i-iii) and CD24 (iv-vi) in red at E13.5, E15.5 and E18.5. Blue indicates nuclear counterstain (DAPI). Inset, secondary antibody control. Scale bars, 50µm.

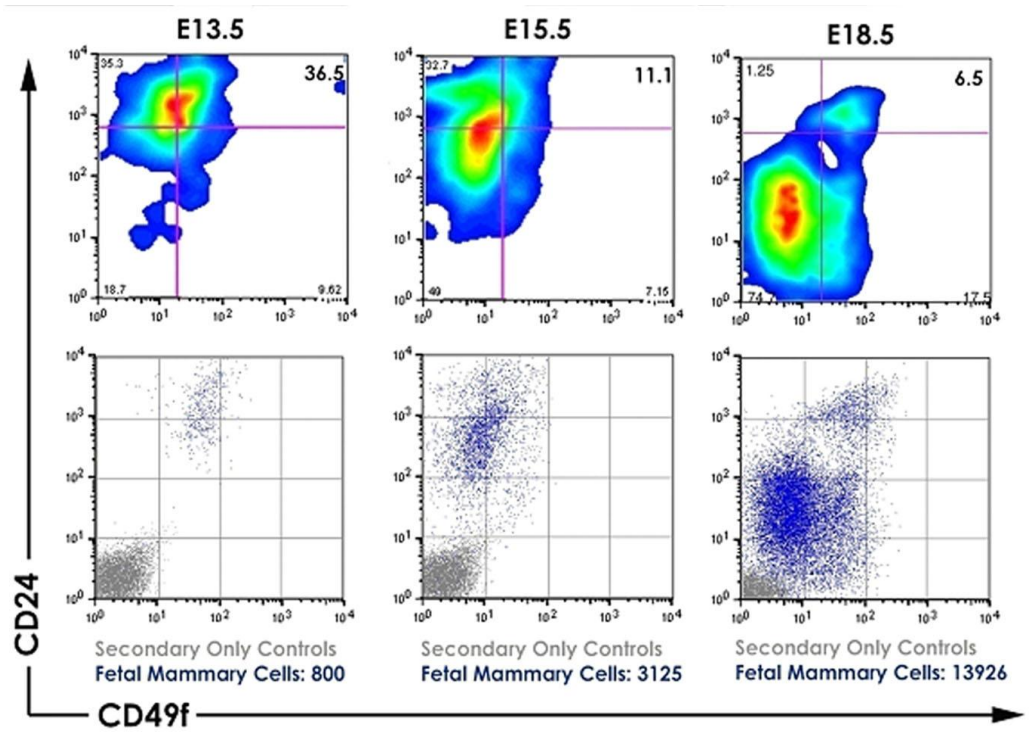


Figure 2-4. Characterization of CD24 and CD49f expression in the mammary gland during fetal development. (A) FACS contour plots showing population centers of LIN$^-$ mammary rudiment cells according to their expression of CD24 and CD49f at E13.5, E15.5 and E18.5. (B) FACS dot plots showing the distribution of cells expressing CD24 and CD49f in the LIN- mammary rudiments (blue) relative to unlabeled controls (gray).

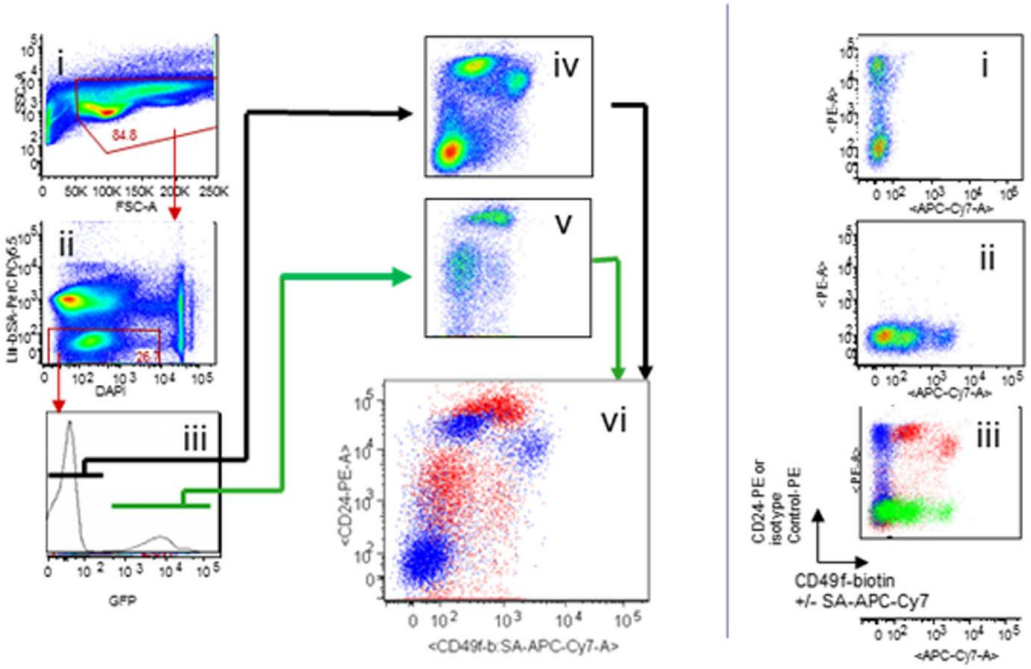


Figure 2-5. Typical gating and control set for flow cytometry and sorting. We employ standard controls to reduce the contribution of signals from debris, non-viable cells and nonspecific staining. This example shows a gating and control scheme for co-processed (dissociated, labeled, and evaluated together) adult and fetal (actin-eGFP) mammary as in Figure 2-6. (A) (i) Debris is reduced through size exclusion and (ii) viable lineage negative cells are identified through DAPI exclusion and lack of reactivity with a lineage antibody set (CD31, CD45, Ter119). (iii) In this overlay experiment, co-processed Actin-eGFP cells from fetal mammary rudiments are gated away from CD1 adult mammary cells (iv-v) and levels of CD49f and CD24 are examined. (vi) Staining of the eGFP positive fetal mammary cells are directly compared.to eGFP negative adult cells (B) Representative control set to ensure staining specificity. (i) no CD49f-biotin primary antibody. (ii) Isotype control-PE instead of CD24-PE (also controls for matched CD49f isotype). (iii) Superimposition of panels B-i, B-ii, and A-iv staining specificity and proper fluorescence compensation.

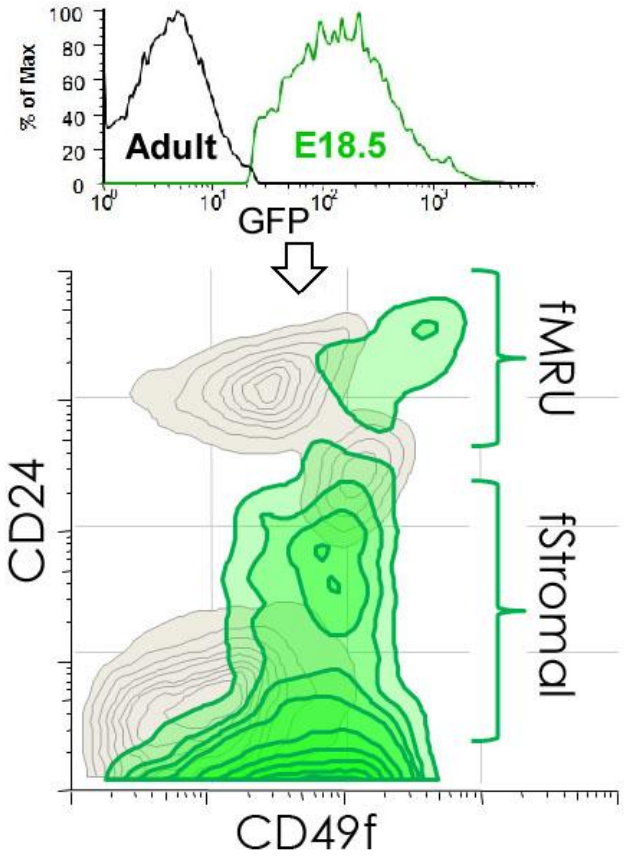


Figure 2-6. Histogram and FACS contour plot showing the distribution of cells expressing CD24 and CD49f in the LIN$^-$ population (DAPI$^-$CD31$^-$CD45$^-$ TER119$^-$) in mammary glands from a nulliparous adult mouse and actin-eGFP E18.5 female embryos. Adult mammary and actin-eGFP E18.5 fetal mammary cell suspensions were mixed, co-stained and analyzed together. The fMRU-enriched population expresses high levels of CD24 while the fStromal-enriched population expresses a range of CD24, from medium to negative relative to isotype controls.

Table 2-4. Fractionation of E18.5 dissociated mammary rudiments using CD24 and CD49f reveal a population highly enriched in mammary repopulating units.

Fractionated mammary cells	Cells injected per fat pad	Repopulated fat pad #	
		No Matrigel	co-injected with Matrigel
fetal MRU-enriched population ($CD24^{high}CD49f^{high}$)	5	N.D.	3/17
	10	N.D.	11/16
	20	1/10	14/20
	25	N.D.	12/14
	50	1/5	4/4
	100	0/4	4/4
	MRU freq. (95% CI)	**1/400† (1/100 - 1/1,700)**	**1/14 (1/10 - 1/18)**
	p Value	0.2	0.7
fetal Stromal-enriched population ($CD24^{med/low/neg}$)	50-300	0/11	0/9
	1000-3000	0/8	0/9
	MRU freq. (95% CI)	**<1/9,000**	**<1/5,000**
	p Value		
Adult MRU-enriched population ($CD24^{med}CD49f^{high}$)	10	0/6	1/6
	50	1/6	4/6
	100	1/7	6/7
	1000	1/2	2/2
	MRU freq. (95% CI)	**1/800† (1/200 - 1/2,700)**	**1/50 (1/30 - 1/100)**
	p Value	0.4	0.98

p values > 0.05 indicate the data are consistent with a single-hit Poisson model.
† Rough MRU frequency estimate (see Statistical Analyses under Methods for detail).
N.D. = not determined.

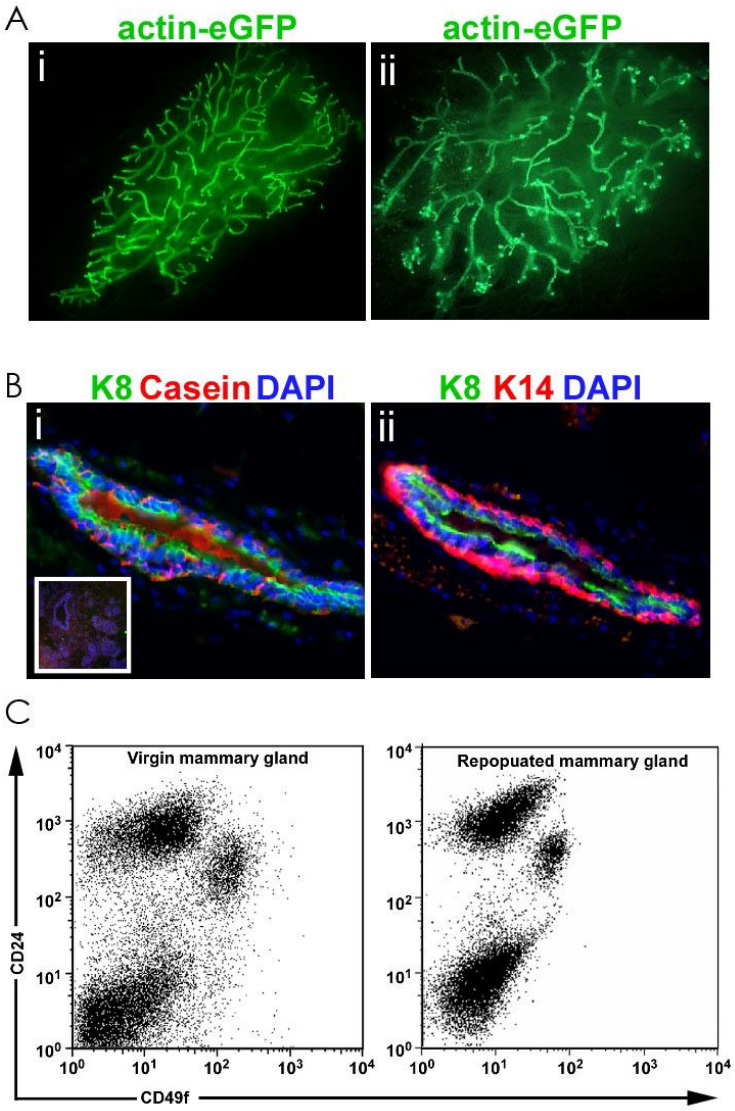


Figure 2-7. A. Representative whole mount of actin-eGFP mammary outgrowth arising from transplantation of cells from the fMRU-enriched population ($Lin^{-}CD24^{high}CD49f^{high}$) isolated from E18.5 embryos. The mammary glands were harvested from a primary recipient (i) and a secondary recipient (ii) 12 weeks after transplant. **B.** Immunofluorescence on paraffin sections of regenerated mammary from a parous recipient showing Casein/K8 (i) and K14/K8 (ii). Inset, secondary antibody control. **C.** Representative FACS dot plots showing the expression of CD24 and CD49f in viable lineage depleted mammary cells from a nulliparous adult mouse (left panel) and from a mammary gland regenerated by the fMRU-enriched population (right panel). The mammary glands were harvested from recipients 12 weeks after transplant.

direct comparison with adult mammary cells, we designate the E18.5 population as $CD24^{high}CD49f^{high}$. The remaining fetal population exhibiting lower CD24 staining is significantly enriched in stromal gene expression (see chapter 5) and is therefore designated as the fStromal-enriched population.

LDTA reveals that all fMRUs reside within the E18.5 $CD24^{high}CD49f^{high}$ population, subsequently referred to as the fMRU-enriched population. As few as 5-10 cells from the fMRU-enriched population reproducibly enabled robust mammary gland repopulation, while as many as 3,000 cells from the fStromal-enriched population consistently failed to generate outgrowths, even with Matrigel addition (Table 2-4). We estimate the fMRU frequency in the fMRU-enriched population to be 1 in 14 with Matrigel and roughly 1 in 400 in its absence (Table 2-4 and Figure 2-2). This is an approximately four-fold enrichment over the aMRU frequency we observe using LDTA within immune-compromised hosts in the presence of Matrigel (Table 2-3 and 2-4). fMRU-derived mammary outgrowths were morphologically normal, fully arborized within the fat pad, and formed casein positive alveolar structures upon induction of pregnancy (Figure 2-7A and B). In addition, they exhibited typical luminal and myoepithelial architecture illustrated by localized expression of luminal and myoepithelial cytokeratins (K8 and K14, respectively) (Figure 2-7B) and recapitulated the phenotypic heterogeneity of wild-type adult mammary glands (Figure 2-7C). Furthermore, serial transplantation demonstrated that mammary glands generated by fMRUs contain stem cells

able to self-renew (Figure 2-7A). Thus, this fetal cell population exhibits the multi-lineage cell differentiation and self-renewal characteristics expected of MaSCs and demonstrated previously for aMRU-enriched populations (Shackleton et al., 2006; Stingl et al., 2006b). These studies demonstrate that the E18.5 fMRU-enriched population contains repopulating activity at concentrations higher than previously achieved using the adult mammary gland and that cells within this population fulfill the functional criteria for MaSCs.

Discussion

This study presents the first quantitative analysis of MaSC activity during fetal mammary development. We show that a single *intact* mammary rudiment from as early as E12.5 reconstituted functional mammary glands. This led us to expect that these structures would contain MaSCs. Our results present the surprising finding that *dissociated* cells from rudiments obtained until approximately E16.5 show little capacity for mammary gland regeneration. In contrast, stem cell activity was readily detectable by E16.5 and increased dramatically by E18.5. The low concentration of cells capable of regenerating mammary glands early in development may indicate that MaSCs are present but very rare at these stages. The fMRU frequency increases approximately 9-fold over a period of 24 hours between E15.5 and E16.5. Therefore, MaSC expansion by cell division alone is unlikely to account for this

dramatic surge in stem cell activity. Alternatively, it is conceivable that MaSC precursors are present from the beginning of development, but require exposure to cues that may arise later in development to engender stem cell function as measured by transplantation. In this model, the elaboration of such cues correlates with when cells in the rudiments invade through the mesenchyme and contact the underlying fat pad. This raises the possibility that extrinsic cues may trigger cellular programming events to increase MaSC abundance and/or to convert MaSC precursors into MaSCs. Along these lines, we demonstrated that Matrigel greatly affects the sensitivity of the transplantation assay. The addition of this basement membrane matrix increases detectable stem cell activity anywhere between 14 – 100 fold in both the embryonic and adult mammary gland cell suspensions. Altogether, our study demonstrates that fMaSCs are highly enriched in late embryogenesis and their repopulating activity is significantly influenced by context.

The un-fractionated, dissociated E18.5 mammary rudiment has a higher MRU frequency relative to the un-fractionated, dissociated adult mammary gland and an equivalent MRU frequency relative to the aMRU-enriched population. Therefore, we utilized this time point to serve as a naturally enriched reservoir of stem cell activity for further purification. Application of an adult MaSC enrichment strategy to the E18.5 rudiment revealed that unlike the adult gland, expression of the markers CD24 and CD49f is tightly restricted to the basal layer of the epithelial ducts, with very little or low staining in the

surrounding stroma. This effectively reduced the amount of stromal contamination in the fMRU-enriched population, which is usually present in the adult MRU-enriched population. In addition, in the adult gland, CD24 expression usually corresponds to the luminal layer of the adult mammary epithelial ducts, which is very distinct from its expression pattern in the embryonic rudiment. Not only is CD24 expressed in the basal layer, its expression seems to coincide with CD49f expression, which corresponds to the basal layer of the adult gland. This could indicate a more plastic state evident in the embryonic gland that is either lost in the adult or masked in the adult MRU-enriched population by contaminating cell types. The novel expression pattern of CD24 and CD49f in the E18.5 rudiment resulted in a four-fold increase in the purity of MRUs than previously achieved in the adult mammary gland. Furthermore, all repopulating activity is segregated into the fMRU-enriched population. By increasing the purity level and decreasing the stromal contamination, it is possible that we will be able to discern unique properties correlating with the MaSC activity observed within the fMRU-enriched population. In addition, it may be possible to identify characteristics associated with the fMRU-enriched population that are either intrinsically different in the adult population or identify a core MaSC phenotype that is obscured by the presence of other cell types and incomplete segregation of MaSC activity.

Acknowledgements

This chapter, in full, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. * indicates equal contribution. The dissertation author was one of the primary authors of this paper.

Chapter 3. Individual cells from the fMRU-enriched population are multi-potent and co-express markers of multiple lineages

Background

Formation of mammospheres as an *in vitro* surrogate assay for stem cell activity

The *in vitro* culture of mammary epithelial cells from both rodent and human primary tissue on solid substrates results in differentiation and senescence (Dontu et al., 2003b). This has precluded the *in vitro* propagation, characterization, and manipulation of MaSC populations. Advances made in the culture of neural stem cells in non-adherent growth conditions allowed their maintenance in an undifferentiated state. These principles were applied to create the mammosphere assay, which has become a popular *in vitro* surrogate assay for stem cell activity. In general, sphere formation in suspension has not only been a method for *in vitro* propagation of stem cells, but also a means by which they can be enriched (Dontu et al., 2003b). Cells from these mammospheres were able to generate the multiple lineages of the mammary gland upon different culture conditions *in vitro*. In addition, cells from mammospheres were able to serially generate spheres in these non-adherent conditions, which has been suggested to be analogous to self-renewal. However, it is important to note that aggregation and migration can contribute to mammosphere formation and the impact of these processes to

this culture strategy need to be evaluated. More stringent analyses can be carried out either by performing single cell mammosphere formation assays or by examining the clonality of the mammospheres using fluorescently labeled cells mixed at different ratios. The ability of cells to migrate in the mammosphere assay can also be mitigated by the addition of methyl cellulose to the media.

A single mammosphere has also been demonstrated to regenerate the mammary gland upon transplantation with 15 – 33% efficiency (Moraes et al., 2007). More recently, mammosphere culture has been used to identify slowly cycling human and mouse mammary cells using a membrane binding dye, PKH-26, which serves as a label retention system *in vitro* (Cicalese et al., 2009a; Pece et al., 2010). By isolating the label retaining, infrequently dividing cells from the mammosphere cultures, MaSCs were purified to a frequency of one in three from mouse cultures and one in twenty-six from human cultures as estimated by LDTA. However, in order to isolate these cells, they are exposed to *in vitro* culture and the effects of this dramatic change in environment are unclear. Although the MaSC frequencies from this culture enrichment strategy are at best approximately 30%, the authors claim to have relatively pure populations due to the stress of the cell isolation procedure. In addition, in the human mammary gland, they found the *in vivo* markers, $CD24^{high}CD49f^{high}$, Dll1 and DNER, to correlate with PKH-26 positivity *in vitro*. However, the MRU frequencies of these *in vivo* isolated populations were not

determined.

Expression of markers indicative of a multi-potent state

MaSCs have previously been proposed to co-express markers of both the myoepithelial and luminal epithelial lineages, such as cytokeratins K14/5 and K8/18/19, respectively (Petersen and Polyak, 2010; Sun et al., 2010; Villadsen et al., 2007). The enriched population of aMaSCs from the mouse has been published to lack any cells co-expressing markers of multiple lineages and contains approximately 27% K14 positive cells, 18% K18 positive cells, with the remaining cells within this population expressing neither of these markers (Stingl et al., 2006a). In another study using a different approach for the enrichment of human aMaSCs, this population was reported to exclusively include K14, K19 double positive cells (Villadsen et al., 2007).

Human breast cancer cells that co-express epithelial cytokeratins and vimentin have been correlated with poor prognosis and tumor-initiating activity in xenograft models (Creighton et al., 2009; Thomas et al., 1999). The expression of vimentin has been previously documented in human myoepithelial cells; however, these results have not been consistent between different samples and antibodies (Asch et al., 1981; Mork et al., 1990). In the mouse, vimentin expression is strictly observed in the stroma, but similar to the human mammary gland, expression of vimentin in rat myoepithelial cells has also been demonstrated (Asch et al., 1981; Warburton et al., 1989). However, the expression of vimentin within epithelial cells in the human

mammary gland is restricted to the myoepithelial lineage and is rarely observed in concert with luminal keratin expression (Mork et al., 1990). Interestingly, forced co-expression of the luminal keratins 8 and 18 with Vimentin induced motility, invasiveness, proliferation, and tumorigenecity of MCF7 human breast cancer cells *in vitro* (Hendrix et al., 1997). The motile and invasive phenotypes of MCF7 and MDA-MB-231 cells that co-expressed luminal cytokeratins and vimentin could be transiently abrogated by the addition of antisense vimentin oligonucleotides. However, metastatic potential of these cell lines was ultimately not affected by the co-expression of vimentin and luminal cytokeratins. These data suggest that a plastic phenotype as indicated by the co-expression of luminal epithelial cytokeratins and the mesenchymal marker vimentin correlate with a more aggressive phenotype in human breast cancer.

Beyond invasion and proliferation, an epithelial to mesenchymal transition could signify a return to an embryonic tissue stem cell phenotype. In normal human mammary epithelial cells, induction of EMT by inducible expression of the transcription factors Twist or Snail or by exposure to TGFβ1 induced the acquisition of the cell surface phenotype, $CD44^{high}CD24^{low}$, which is associated with breast CSC activity. Moreover, induction of EMT greatly enhanced the ability of these cells to generate mammospheres *in vitro* in the presence of methyl cellulose and also promoted their tumorigenecity *in vivo* (Mani et al., 2008).

Results

There are a plethora of stem cell associated enrichment and depletion markers. Therefore, we sought to better characterize the fMRU-enriched population to determine their cell cycle status and whether any of these other markers correlated with stem cell activity (Figure 3-1). For example, expression of multi-drug transporters measured by dye efflux and Sca1 positivity have been correlated with stem cell activity in adult tissues including the mammary gland (Li and Rosen, 2005). However, the E18.5 mammary gland does not contain significant Hoechst 33342 dye efflux activity and Sca1 positivity is just as prevalent in the non-stem cell containing fStromal-enriched population as the fMRU-enriched population (Figure 3-1 B and C). Another property reported to be enriched in normal and malignant adult breast stem cells is aldehyde dehydrogenase (ALDH) activity (Ginestier et al., 2007). However, ALDH activity was also equally distributed between the fMRU- and fStromal-enriched populations in which approximately 5% of the cells are labeled in each population.

We used *in vitro* mammosphere formation to determine whether individual cells from the fMRU-enriched population exhibit multi-lineage capacity (Dontu et al., 2003a). Surprisingly, using the conventional sphere forming protocol in non-adherent culturing conditions at low seeding density (~1000 cells per cm^2), the fMRU-enriched population had negligible sphere forming efficiency (SFE) (SFE ≈ 0.1%, Figure 3-2). In contrast, the fStromal-

enriched population formed spherical clusters at much higher efficiency using identical conditions (SFE=1.35%) (Figure 3-2).

The discordance between the low *in vitro* sphere forming capacity and the high *in vivo* transplantation efficiency prompted us to identify conditions under which the fMRU-enriched population generated spheres *in vitro* with higher efficiency. Just as Matrigel profoundly increased transplantation efficiency, adding a low percentage of Matrigel (2%) to the culture medium enabled the fMRU-enriched population to form spheres when plated at low density in semi-non-adherent growth conditions. The spheres generated by the fMRU-enriched population were morphologically similar to those reported for aMRU-enriched populations, which had smooth boundaries (Figure 3-2) (Stingl et al., 2006b). On the other hand, the fStromal-enriched population did not generate spheres in the presence of Matrigel, but instead produced cultures of dispersed cells resembling fibroblasts and neurons (Figure 3-2). In the absence of Matrigel, the fStromal spheres had irregular boundaries in which individual cells could be visualized. It is possible that these spheres were formed by aggregation in the absence of Matrigel and that the motility of these cells is restricted upon the addition of low concentrations of Matrigel to the media, thereby preventing sphere formation.

We next used two independent strategies to determine whether spheres derive from a single cell or from cell aggregation. We first seeded single cells from the fMRU-enriched population into individual wells in the

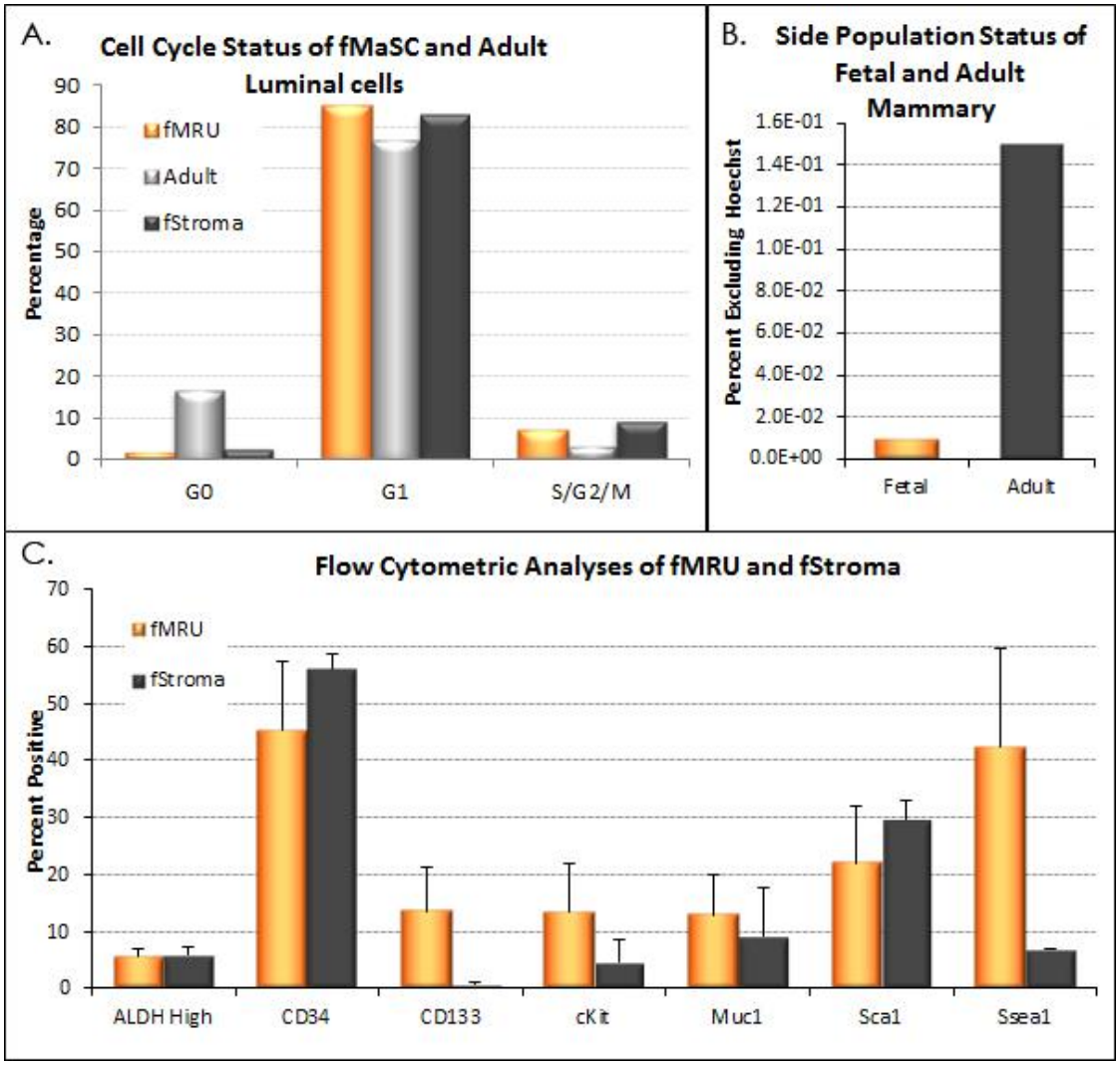


Figure 3- 1. A. Percentage of fMRU ($LIN^-CD24^{high}CD49f^{high}$), fStromal ($LIN^-CD24^{med/low/neg}CD49f^{neg}$) and adult luminal cells ($LIN^-CD24^+CD49f^{low}$, Stingl et al., *Nature* 2006) in the G0, G1 and S/G2/M fractions as determined by pyronin Y/Hoechst33342 staining. Inset shows representative staining. **B.** Percentage of unfractionated E18.5 and adult mammary cells that excluded Hoechst3342 (side population) as measured by Hoechst3342 fluorescence. **C.** Percentage of fMRU and fStromal-enriched populations expressing stem and lineage markers by flow cytometry.

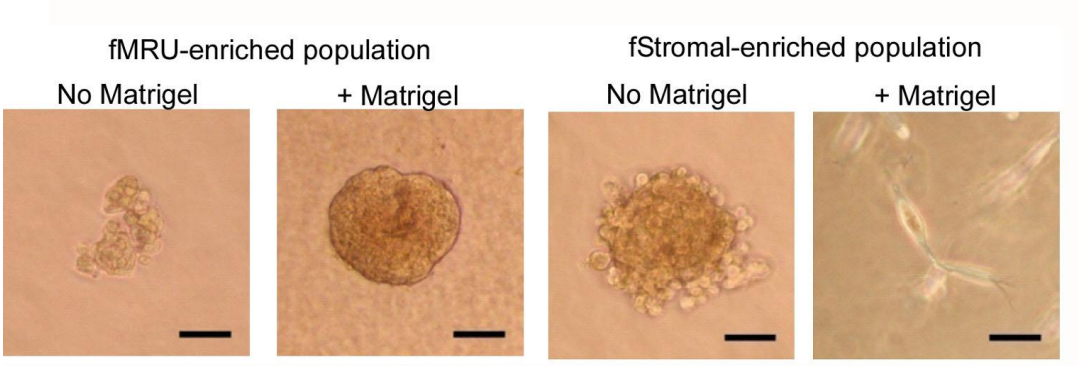


Figure 3-2. Morphology of outgrowths from fMRU- and fStromal-enriched populations grown under non-adherent conditions *in vitro*, in the presence and absence of Matrigel. Scale bar, 50μm.

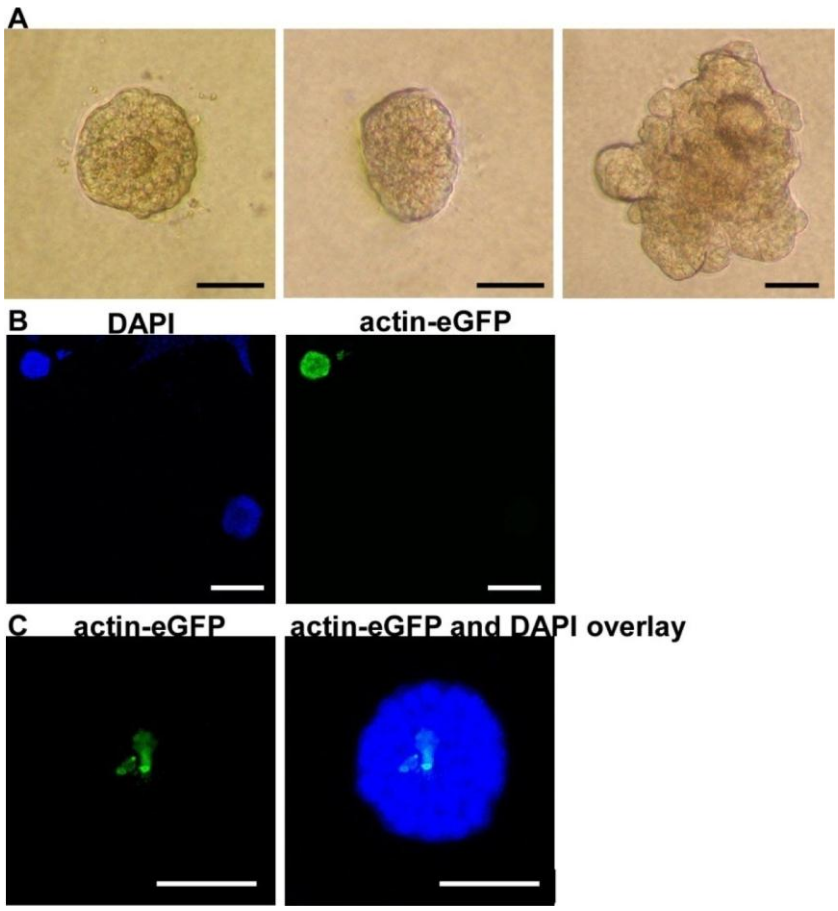


Figure 3-3. Individual cells from the fMRU-enriched population generate clonal primary spheres. **A.** Brightfield images of primary spheres generated from single cells cultured in individual wells of 96 well plates (n=3, at least 3 96 well plates per experiment, sphere forming efficiency = 10.7 ± 1.5%). **B and C.** Cells from the fMRU-enriched populations were independently prepared from Actin-eGFP mice and wild type mice (CD1 strain), mixed in a 1:1 ratio (B) and a 4:1 ratio (C), and seeded on top of Matrigel. 207 spheres were a single color and one sphere was chimeric (shown in C). Scale bars, 50µm.

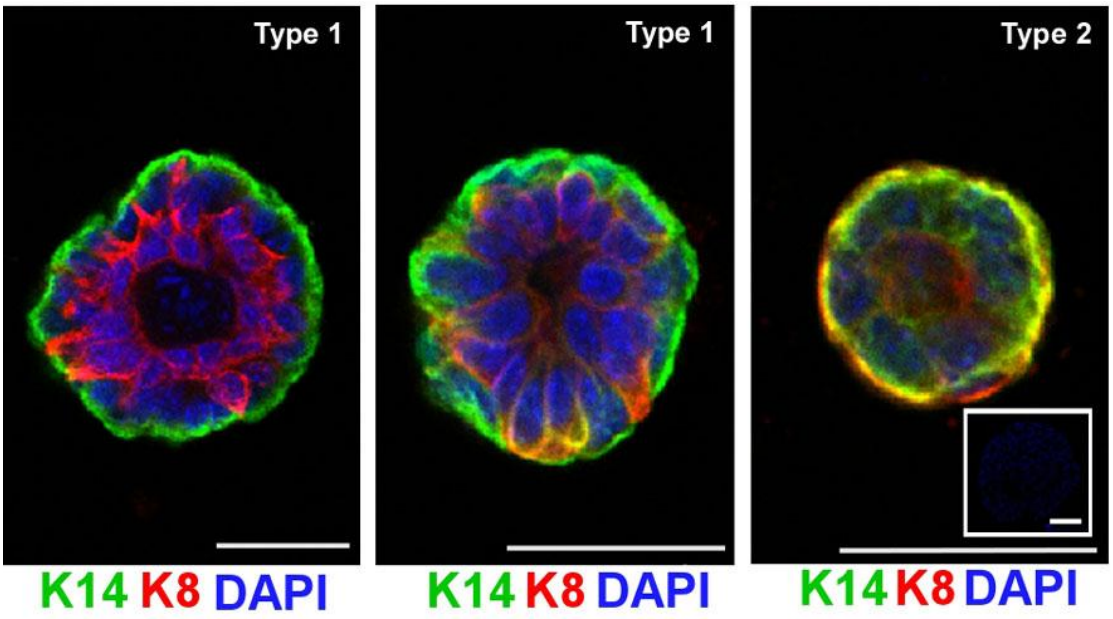
presence of low concentrations of Matrigel. Spheres were formed with a SFE of 10.7 ± 1.5%, similar to that observed when the cells were plated at low density in the presence of Matrigel (SFE = 10%). Interestingly, the *in vitro* SFE is similar to the ~7% fMRU frequency determined by transplantation (Figure 3-2 and 3-3A), suggesting that both assays provide a measure of stem cell frequency. The second strategy to assess sphere clonality involved mixing single cell suspensions consisting of $eGFP^+$ and $eGFP^-$ cells from the fMRU-enriched population, and then growing them at low density (Figure 3-3A and B). We rarely observed chimeric spheres (1 chimera per 200 spheres analyzed), indicating that the vast majority of spheres represent clonal growth from a single initiating cell. The fMRU-enriched population generated spheres that almost exclusively expressed both K14 and K8, either within the same cell or in separate cells within the same sphere (Figure 3-4). Since spheres generated from the fMRU-enriched population arise from single cells, we infer that the vast majority (>99%) of the sphere initiating cells were multipotent, which indicates that ~10% of the cells from the fMRU-enriched population have this property *in vitro*.

The relationship between cells co-expressing K14 and K8 and stem cell activity has not been analyzed throughout fetal development. Our immunofluorescence analyses revealed numerous K14 and K8 double positive cells in intact rudiments and dissociated cells at E13.5 and E15.5 despite the low abundance of MRUs during this period (Figure 3-5 and 3-6). The

percentage of double positive cells decreased dramatically by E18.5 in un-fractionated, dissociated rudiments. In contrast, the percentage of K14+/K8+ cells remained high in the stem cell containing fMRU-enriched population (~30% K14+K8+, Figure 3-6). We also detected cytokeratin negative cells within the fMRU-enriched population whose fate and function is unknown (Figure 3-6). Thus, our data show that double positivity for myoepithelial and luminal epithelial markers does not correlate exclusively with stem cell activity during development, even though double positive cells are evident within stem-enriched populations.

Discussion

The fMRU-enriched population exhibits several unique characteristics. Previous studies proposed that co-expression of markers of both myoepithelial (K14) and luminal epithelial (K8) specification would be indicative of multi-potency in mammary cells (Petersen and Polyak, 2010; Sun et al., 2010; Villadsen et al., 2007). While we detected such cells, their abundance did not correlate with MaSC activity throughout development. However, these cells were significantly enriched in the fMRU-enriched population and also, many of the spheres generated from the fMRU-enriched population expressed markers of both lineages within the same cell. It is possible that double positivity of these lineage markers correlates with acquisition of MaSC identity, but not competency. As discussed in the prior chapter, it is conceivable that MaSC



Sphere Phenotype

K14	K8	Frequency Observed
+	+	99.7% (72% Type 1, 27.7% Type 2)
-	+	0%
+	-	0.3%
-	-	0%

Figure 3-4. Confocal images of immunofluorescence on spheres derived from the fMRU-enriched population showing the expression of K8, K14 or both, and tabular summary of the types of spheres observed. Type 1 spheres consist of cells expressing either K14 (i.e. sphere periphery) or K8 (i.e. middle of the sphere), while type 2 spheres consist mainly of cells co-expressing K8 and K14 (yellow cells). The clonal sphere forming efficiency is 10.7 ± 1.5%. Inset, secondary antibody control. Scale bars, 25µm.

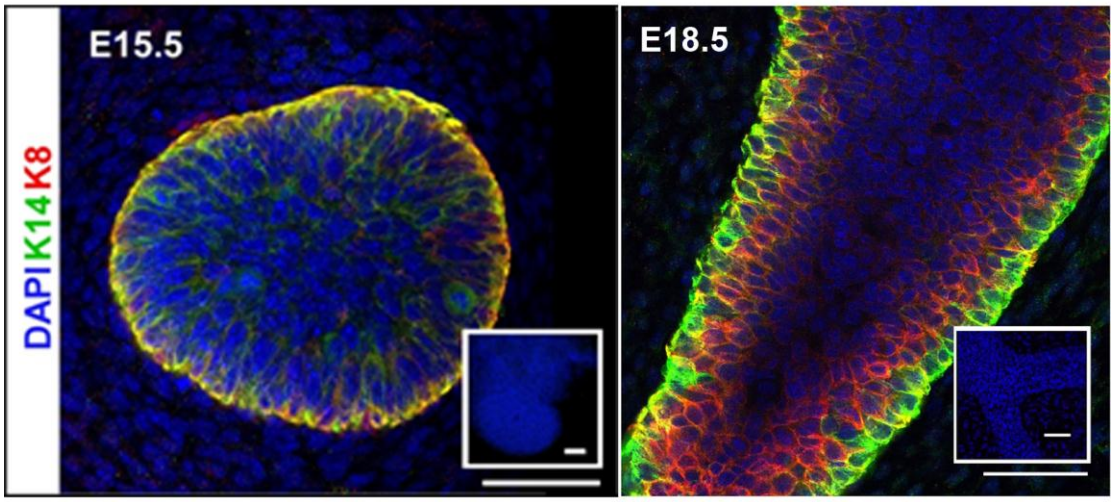


Figure 3-5. Confocal images of immunofluorescence on E15.5 and E18.5 mammary rudiment showing cells expressing K8, K14, both and neither. Insets, secondary antibody only. Scale bar, 50µm.

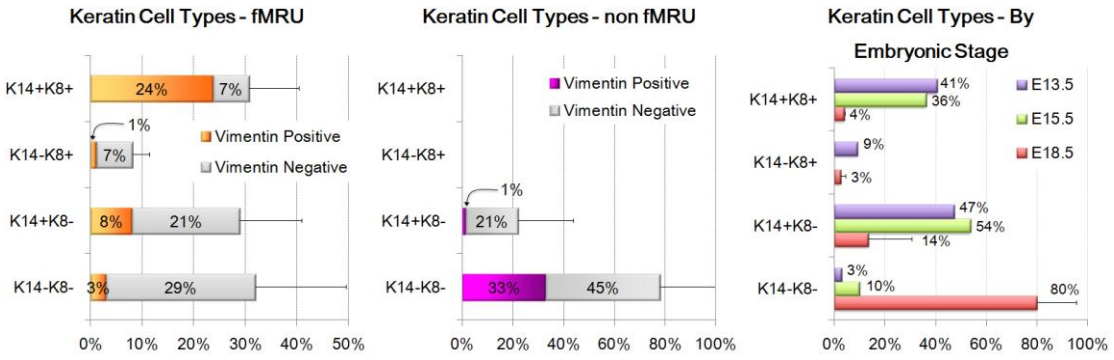


Figure 3-6. Bar charts summarizing the percentage of cells in the fMRU- and fStromal-enriched populations expressing K8, K14, and/or Vimentin. Bar chart summarizing the percentage of K14 and K8 positive and negative cells in E13.5, E15.5 and E18.5 bulk mammary cell populations.

precursors are present in the early mammary rudiments, but await an inductive cue to trigger full MaSC competence. One possibility is that that double positivity for myoepithelial and luminal epithelial markers indicates the capacity for MaSC activity, but is not sufficient. This would include the consideration that although the fMRU-enriched population is enriched for such double positive cells, their numbers exceed the estimation of MRU frequency within this population (30% double positive, 7% pure).

The co-expression of not only myoepithelial and luminal markers, but also mesenchymal markers, indicates an uncommitted state not typically observed in the adult outside of the tumorigenic context. The co-expression of mesenchymal and epithelial markers was also observed during sphere formation as will be discussed in later chapters. Of note, these results raise the possibility that the apparent epithelial to mesenchymal transition (EMT) commonly observed during aggressive tumorigenesis may represent a reversion to an embryonic-like state. Indeed, EMT as an embryonic process has long been recognized and without the process of EMT, development cannot proceed beyond the blastula stage (Briegel, 2006; Hay, 1995; Perez-Pomares and Munoz-Chapuli, 2002; Thiery and Sleeman, 2006). Furthermore, epithelial tissues, such as the kidney, are derived from the mesoderm and vice versa, mesodermal cells can be derived from epithelial endodermal cells (Davies, 1996; Gershengorn et al., 2004). It has been reported that an EMT can induce a stem cell state, but it is also possible that a plastic EMT like

phenotype is an intrinsic property of developing tissue stem cells (Mani et al., 2008).

In light of these findings, our results suggest the utility of studying programs that are present within the fMRU-enriched population, not detected in the normal adult, and deregulated during tumorigenesis. This strategy could be extended to fetal stem cells from other tissues and their corresponding tumors. The fMRU-enriched population exhibits cancer relevant properties that have not been reported for normal aMaSCs either because they are unique to the fetal state or developmental context, or because they have been obscured by contaminating cells in isolates from adult tissue.

Acknowledgements

This chapter, in full, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. * indicates equal contribution. The dissertation author was one of the primary authors of this paper.

Chapter 4. Neuronal cell representation during mammary gland development

Background

Development requires fine-tuned coordination between many heterogeneous cell types. Studies on development often recognize the contribution of the stroma to the developing epithelial structure. However, although the mesenchymal cell contribution is well recognized, the input of other cell types has not been explored in detail, yet may also be important. Indeed, recent studies of the submandibular mouse salivary gland have demonstrated that upon removal of parasympathetic innervation of the tissue, the branching morphology and number of progenitors is severely affected (Knox et al., 2010). The loss of its innervation source can be replaced by an Acetylcholine analog, carbachol, which rescues the number of salivary gland epithelial progenitors. The development of the salivary gland follows a very similar progression as the mammary gland. The epithelium next to the tongue thickens into a structure akin to the placodes of the mammary gland and at the same time point of E11.5. This thickened protrusion invaginates into the underlying mesenchyme to form a bud by E12.5 (Tucker, 2007). Branching of the salivary gland occurs by E13.5, at which point there are between 4 – 5 buds, which continue to branch until there is a multi-lobed gland by E14.5 (Tucker, 2007). The ducts within the salivary gland form lumens beginning at

E15.5 and becoming well developed by E17.5, although the lumens are not yet continuous (Tucker, 2007). As both the mammary gland and salivary gland follow a similar developmental paradigm, it is possible that like the salivary gland, the mammary gland and its progenitors may not only depend on coordinated development of the epithelium and mesenchyme, but also neuronal cells.

Neural Cell Adhesion Molecule

Neural cell adhesion molecule, NCAM, was highly up-regulated in the fMRU-enriched population relative to the fStromal-enriched population by microarray analysis that will be discussed in following chapters. It is unclear whether detection of this and many other related neuronal associated genes that are expressed in this population represent contribution from a neuronal cell population or co-expression of these genes within the epithelial cells. Therefore, the contributions of this gene to the fMRU-enriched population were studied in greater detail.

NCAM can be alternatively spliced into three different class sizes (Bonfanti, 2006). The 120kDa isoform is most often found during embryogenesis (Goridis and Brunet, 1992; Zecchini and Cavallaro, 2010). Upon reaching adulthood, the splicing switches predominantly to the 140 and 180 kDa forms. NCAM can mediate trans-cellular interactions through homophilic adhesion through its extracellular domains, which also contain

domains that bind heparin-sulfates, collagens, and other extracellular matrix factors (Ronn et al., 1998). In addition, NCAM binding can instigate signal transduction and has been demonstrated to induce neurite extension through increased intracellular calcium that is dependent on $p59^{fyn}$ tyrosine kinase activity (Ronn et al., 1998). NCAM can also bind FGF receptors and activate their downstream signaling and mediate L1CAM homophilic binding through heterophilic interactions (Ronn et al., 1998). Despite these functions, mice develop relatively normally upon removal of NCAM from the genome (Cremer et al., 1997).

NCAM proteins can be post-translationally modified with polysialic acid (Bonfanti, 2006; Ronn et al., 1998). Polysialic acid (PSA) is a large, negatively charged and highly hydrated glycan that accounts for approximately 30% of the NCAM protein's molecular mass during embryonic and early postnatal brain development, but only 10% in the adult brain (Bonfanti, 2006; Goridis and Brunet, 1992). Poly-sialylation has been demonstrated to change cellular interactions by decreasing homophilic binding and attenuating cell adhesion (Ronn et al., 1998; Sadoul et al., 1983). In this way, it is thought that PSA-NCAM can promote cellular migration (Cremer et al., 1997; Yoshida et al., 1999) and allow plasticity of cellular interaction (Rutishauser and Landmesser, 1996). Altogether, these data demonstrate the importance of both stable interactions and dynamic adhesion during tissue development and structural plasticity (Bonfanti, 2006).

During tumorigenesis, a switch back to the embryonic isoform is commonly observed (Zecchini and Cavallaro, 2010). The expression of NCAM has been correlated with tumor progression and metastasis (Christofori, 2003) and the expression of a α2,6-sialyltransferase has been found to mediate passage of breast cancer cells through the blood brain barrier, thereby mediating brain metastasis (Bos et al., 2009). Furthermore, the forced expression of NCAM has demonstrated that this protein is both necessary and sufficient to induce an EMT phenotype in epithelial cells *in vitro* (Lehembre et al., 2008). However, the role of the PSA modification during mammary development and breast tumorigenesis remains unclear.

Single Cell Analysis

Given that the normal breast is extremely heterogeneous in its cellular makeup, it follows that breast cancer is also a very heterogeneous disease, both from patient to patient as well as within individual patients themselves (Navin et al., 2010). Each of the cell types present in a normal mammary gland contribute to the overall function of the tissue. Therefore, it is also likely that the presence of multiple cell types in a tumor play a role in its maintenance, perpetuation, and spread. Along these lines, recent studies have reported the polyclonality and heterogeneity of not only breast cancer, but also of glioblastoma (Bonavia et al., 2011) and small cell lung cancer (Calbo et al., 2011). This heterogeneity is likely to reflect the interplay between the evolving

clones and their progeny and these interactions have been designated as interclonal cooperativity (Lyons et al., 2008). These results emphasize the need to elucidate the individual cell contributions first within the normal breast to understand how these cell types are perturbed to contribute to the evolution of a malignant tumor. However, until recently, single cell analysis has not been possible in a high throughput and reliable manner (Warren et al., 2006; Zhong et al., 2008). Recently, new technology to screen single cells for 96 different transcripts in a high throughput manner has become more widespread, enabling the evaluation of heterogeneity in normal development, stem cell populations, and cancer (Diehn et al., 2009; Guo et al., 2010).

Stem cells function in close association with the supporting cell types and factors that compose their niche. Although the contribution of the mesenchyme in supporting mammary gland development has been well studied, the involvement of neuronal cells to MaSC activity is unclear. It is conceivable that like the salivary gland, innervation of the mammary gland influences stem/progenitor cell activity and specification. Using single cell analysis, we dissect the components of the fMRU-enriched population. These analyses revealed the presence of neuronal-like, NCAM positive cells within the population enriched for MRUs. Whether any of the cell types present, including the NCAM positive subpopulation, support MRU function throughout development remains unclear.

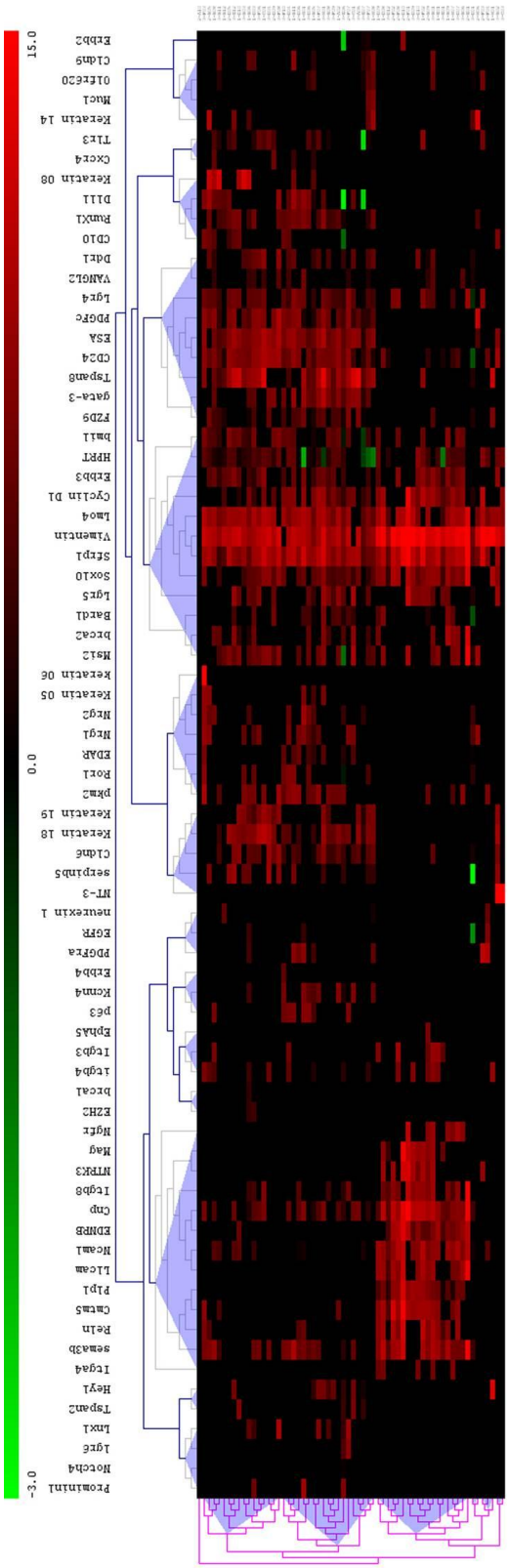
Results

The fMRU-enriched population may be composed of cells other than the MaSC. These cells may be important for the support of the MaSC, or result from contamination with unnecessary cell types. We used a high throughput single-cell qRT-PCR platform to characterize the cell types within the fMRU-enriched population. We confirmed that each sample contained a single cell by titrating expression of housekeeping genes (data not shown, see methods). Expression of epithelial and mesenchymal markers was examined in addition to genes revealed by microarray analyses to be highly expressed in the fMRU-enriched population (Figure 5-1). Hierarchical clustering of the expression patterns of the 74 genes scored enabled reproducible segregation of the fMRU–enriched population into different groups upon normalization to either HPRT (data not shown) or GAPDH (Figure 4-1). This approach confirmed the existence of the cytokeratin subtypes identified using immunofluorescence, and also verified that individual fetal cells co-express myoepithelial and luminal epithelial cytokeratins along with the mesenchymal marker, Vimentin (Figure 4-1). The first obvious bifurcation of different cell types within the fMRU-enriched population is the presence of a keratin negative population by mRNA expression that expresses a cluster of neuronally associated genes, such as NGFR, NCAM, L1CAM and MAG (Figure 4-1). Although these cells were originally sorted based on their high levels of CD24 protein expression by flow

cytometry, they appear to have lower levels of CD24 expression at the transcript level.

Fractionation of the fMRU-enriched population was possible by utilizing a PSA-NCAM targeted antibody (Figure 4-2). PSA-NCAM positive cells isolated from the fMRU-enriched population maintained a Keratin negative phenotype by immunofluorescence, matching the results obtained by single cell qRT-PCR (Figure 4-3). In addition, upon recovery after FACS and adherence to a solid substrate, these cells appeared neuronal in morphology with an occasional fibroblastic cell (Figure 4-3). We evaluated the ability of the NCAM positive and negative fractions from within the fMRU-enriched population to form spheres *in vitro*. Normally, the fMRU-enriched population forms spheres at a frequency of approximately 10% when plated in low-adherence plates at low densities in the presence of low concentrations of Matrigel. Only the NCAM negative fraction from the fMRU-enriched population was able to generate spheres under these conditions (8.2%) while both the NCAM positive fraction of the fMRU-enriched population and the fStromal-enriched population failed to generate any spheres (Figure 4-4). The spheres generated by the NCAM negative fraction of the fMRU-enriched population included cells single and double positive for myoepithelial and luminal markers K14 and K8 as well as cells that were triple positive for the two epithelial markers and Vimentin (Figure 4-5). However, upon removal of the Matrigel

Figure 4-1. Single-cell, qRT-PCR analyses on the fMRU-enriched population hierarchically clustered using un-centered Pearson correlation. Each column represents a gene product while each row represents a single cell. Black indicates no detectable signal and increasing intensity of red indicates higher expression relative to a corrected GAPDH value.


15.0
0.0
-3.0
Erbb2
Cldn9
01Ffe30
Muc1
Keratin 14
Tlr3
Cxcr4
Keratin 08
D111
Runx1
CD10
Ddr1
VANGL2
Lgr4
PDGFC
ESA
CD24
Tspan8
gata-3
FZD9
bmi1
HPRT
Erbb3
Cyclin D1
Lmo4
Vimentin
Sfrp1
Sox10
Lgr5
Bard1
brca2
M312
Keratin 06
Keratin 05
Ntrg2
Ntrg1
EDAR
Ror1
pkm2
Keratin 19
Keratin 18
Cldn6
serpinb5
NT-3
neurexin 1
EGFR
PDGFra
Erbb4
Kcnn4
p63
EphA5
Itgb3
Itgb4
brca1
EZH2
Mgfr
Mag
NTRK3
Itgb8
Cnp
EDNRB
Mcam1
L1cam
P1p1
Cmtm5
Reln
sema3b
Itga4
Hey1
Tspan2
Lmx1
Lgr6
Notch4
Frounin1

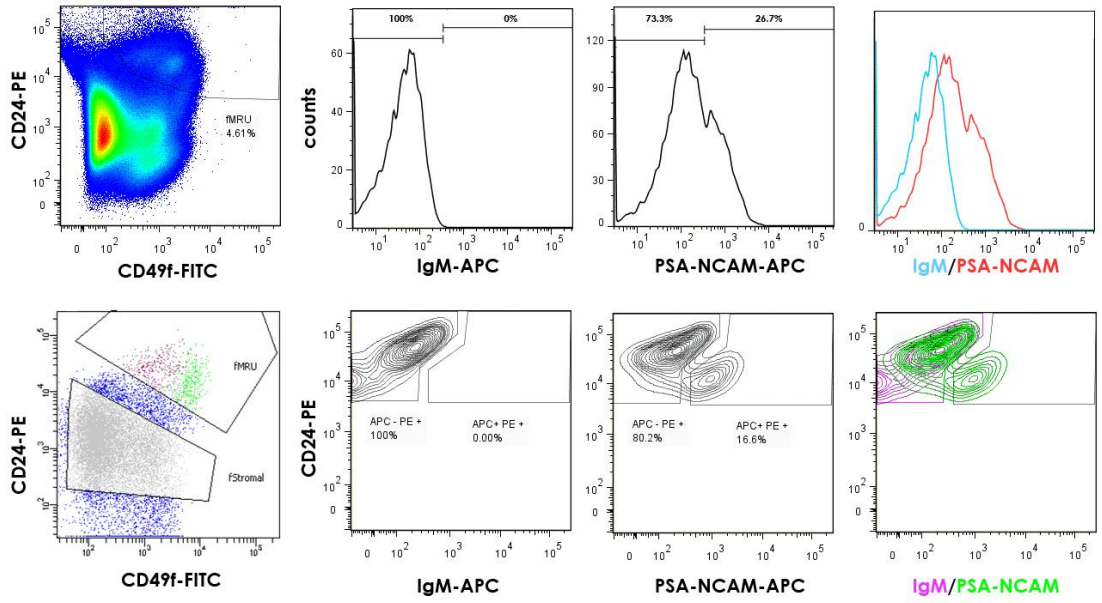
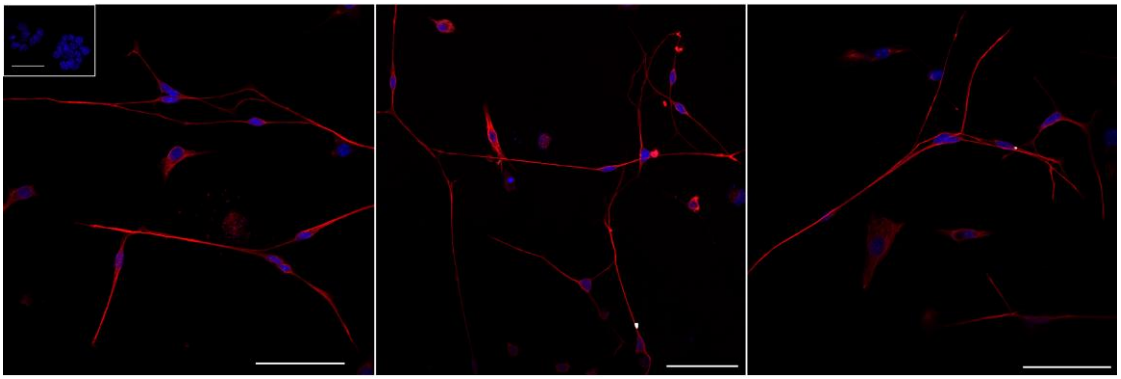


Figure 4-2. Flow cytometric analysis of E18.5 fetal mammary cells after size exclusion, DAPI (dead cell) exclusion, hematopoietic and endothelial lineage exclusion and sorting for CD24, CD49f, and PSA-NCAM. Gating parameters were determined relative to isotype controls as shown.



DAPI Keratin 14 Keratin 8 Vimentin

Figure 4-3. Immunofluorescence analysis on the NCAM+ fraction of the fMRU-enriched population after 3 days of culture on a cell culture coated coverslip. DAPI indicates nuclei (blue), Keratin 14 indicates myoepithelial cells (white), Keratin 8 indicates luminal epithelial cells (green), and Vimentin indicates mesenchymal cells (red). Scale bar represents 50µm. Inset, secondary antibody only control.

from the media, the NCAM negative fraction of the fMRU-enriched population lost its ability to generate spheres. However, in the absence of Matrigel, the NCAM positive fraction of the fMRU-enriched population and the fStromal-enriched population formed spheres at efficiencies of 0.9% and 0.8%, respectively (Figure 4-4). The behavior of the NCAM positive and negative cells from the fMRU-enriched population *in vivo* was similar to the *in vitro* results. Repopulating activity of the NCAM negative fraction of the fMRU-enriched population was 20 fold higher than the corresponding NCAM positive fraction when transplanted in the presence of Matrigel (p value = $5.64x10^{-9}$, Table 4-1). Altogether these data suggest that at this time point, the NCAM positive cells are dispensable for fMRU function. Interestingly, the MRU frequency within non-polysialylated NCAM positive cells from the aMRU-enriched population was 20 fold higher than in the negative fraction (p value = 0.000174, Figure 4-6, Table 4-1). These results highlight yet another difference between the MRU-enriched fractions of the embryonic and adult mammary gland.

NCAM positive cells during embryonic mammogenesis

As described in an earlier chapter, MaSC activity is very rare early in mammogenesis. NCAM positive cells can be detected within the ductal architecture of the mammary rudiments at E18.5 (Figure 4-7). Earlier in development, the close association of these cells with the epithelial rudiment is

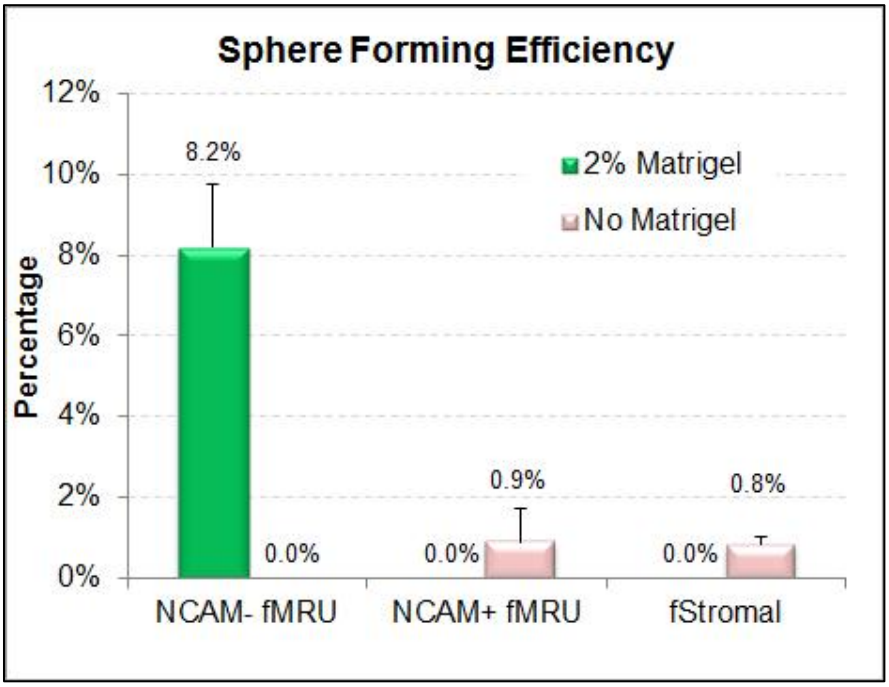


Figure 4-4. Graph showing the sphere forming efficiencies of the fStromal-enriched population and the NCAM positive and negative fractions of the fMRU-enriched population s in the presence of absence of Matrigel.

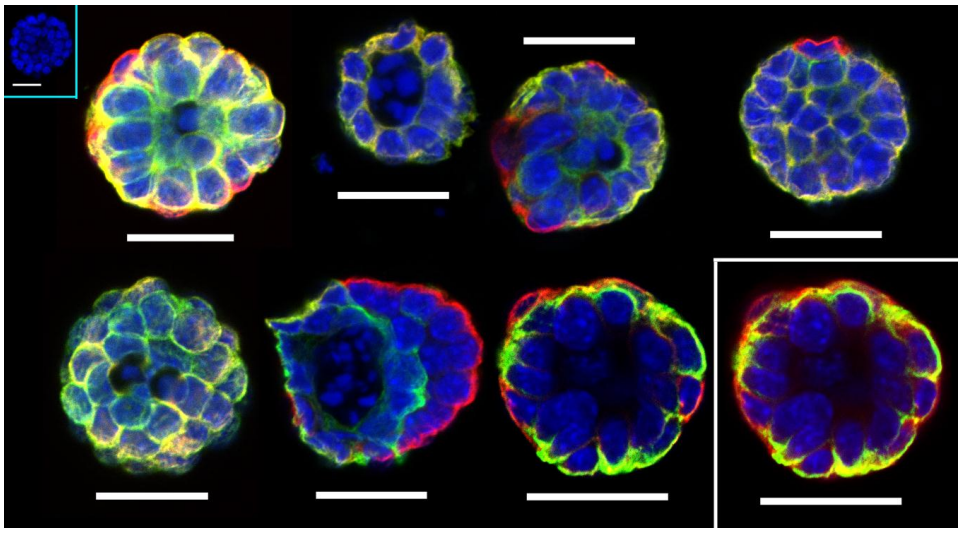


Figure 4-5. Confocal immunofluorescence images of spheres generated from the Ncam negative fraction of the fMRU-enriched population. DAPI (Blue), Keratin 14 (Red), and Keratin 8 (Green). White inset, DAPI (Blue), Vimentin (Red), Keratin 8 (Green). Blue inset, secondary antibody only control. Scale bar represents 20µm.

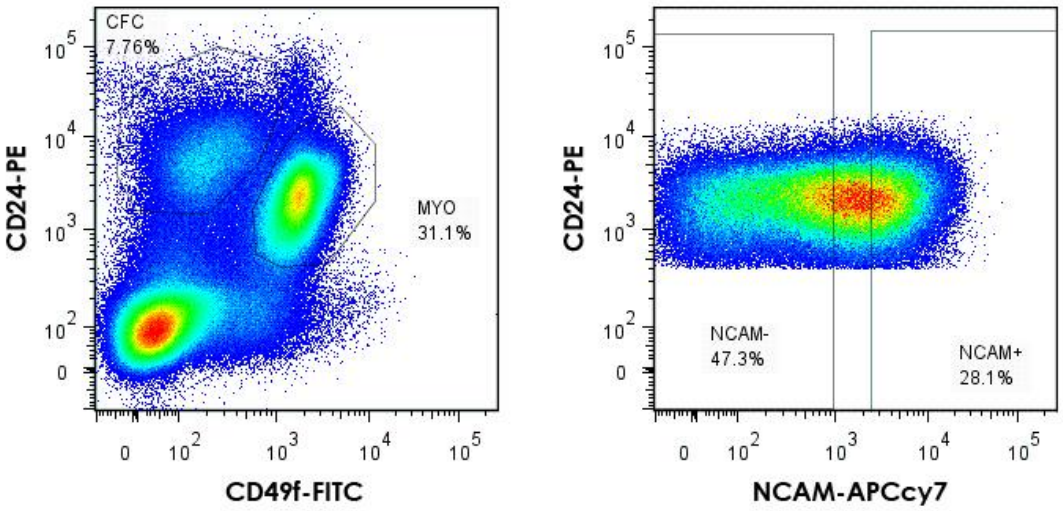


Figure 4-6. Adult flow cytometric analysis detecting presence of un-polysialylated NCAM in the adult mammary gland. Fractionation of the adult Lin^- mammary gland into the MRU-enriched MYO population **(Left).** Fractionation of the MYO population into NCAM positive and negative fractions **(Right)**.

Table 4-1. Mammary repopulation frequency of the NCAM positive and negative fractions of the fMRU and aMRU-enriched population transplanted in the presence of Matrigel into immune-compromised hosts. The values within parentheses represent the 95% confidence interval and the italicized values indicate p value. A p value greater than 0.05 indicates that the data are consistent with a single hit model.

Cell Type	Cell #	Repopulation	MRU Frequency	Full Regrowth
Ncam$^-$ fMRU	1000	2/2	1/37.2 (73.9 – 18.7) *0.584*	2/2
	100	10/11		9/11
	10	3/10		1/10
Ncam$^+$ fMRU	1000	0/2	1/882 (284 – 2743) *0.000936*	0/2
	100	2/15		2/15
	10	2/14		2/14
aMRU	100	4/5	1/62.1 (184.7 – 20.91)	1/5
Ncam$^+$ aMRU	100	5/5	1/20.6 (80.8 – 5.27) *0.738*	5/5
	10	1/3		1/3
Ncam$^-$ aMRU	1000	1/2	1/423.4 (1551.4 115.55) *0.0652*	1/2
	100	3/5		1/5
	10	0/3		0/3

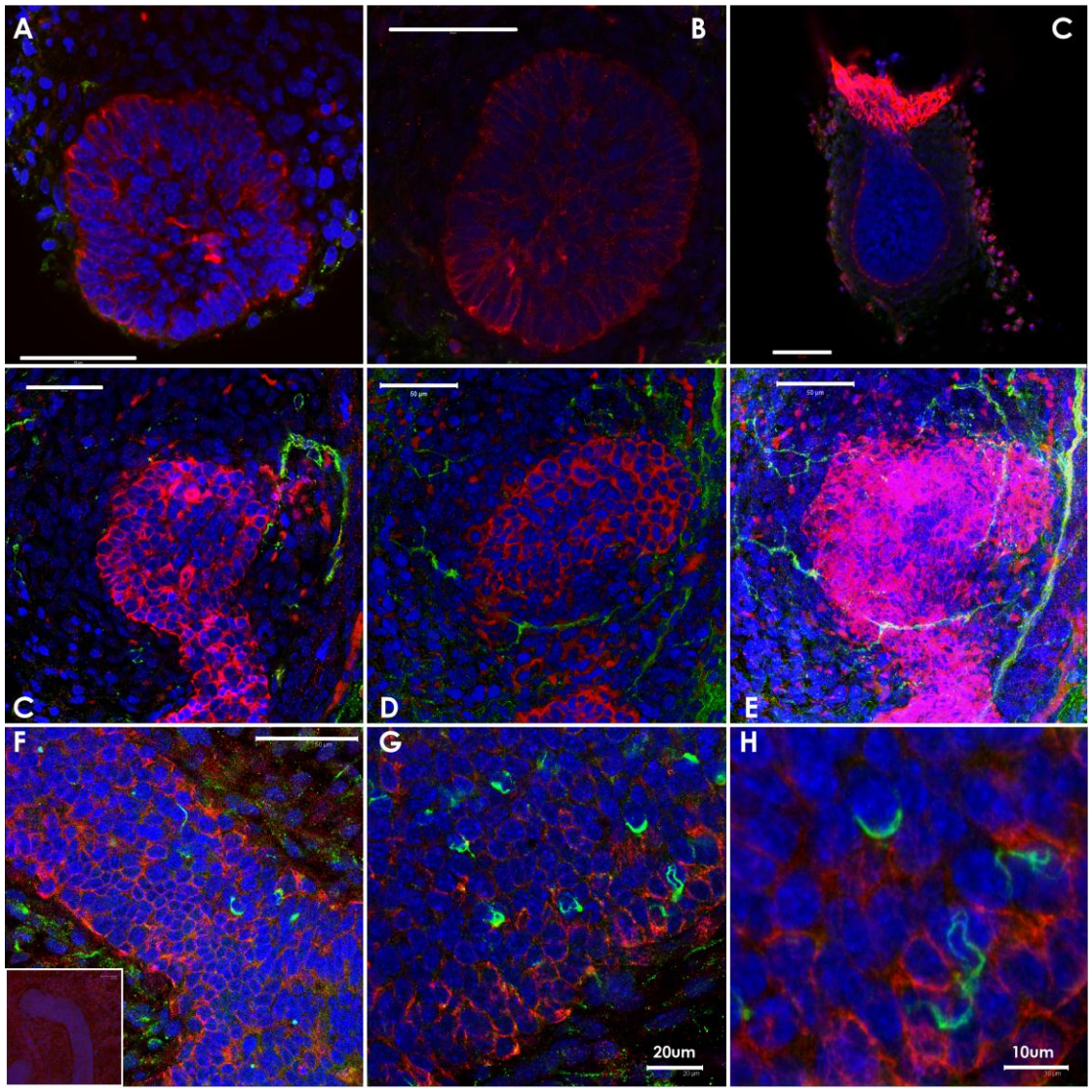


Figure 4-7. Confocal immunofluorescence images. DAPI (Blue), Pan-cytokeratin (Red) and PSA-NCAM (Green) staining on mammary rudiments from E15.5 **(A-B),** E13.5 **(C),** and E16.5 **(C, D** Average Projection**, E –** Maximum Projection**).** DAPI (Blue), Keratin 14 (Red), and PSA-NCAM (Green) staining on E18.5 mammary rudiments **(F – H).** Scale bars represent 50µm unless otherwise indicated. Inset, secondary antibody only control.

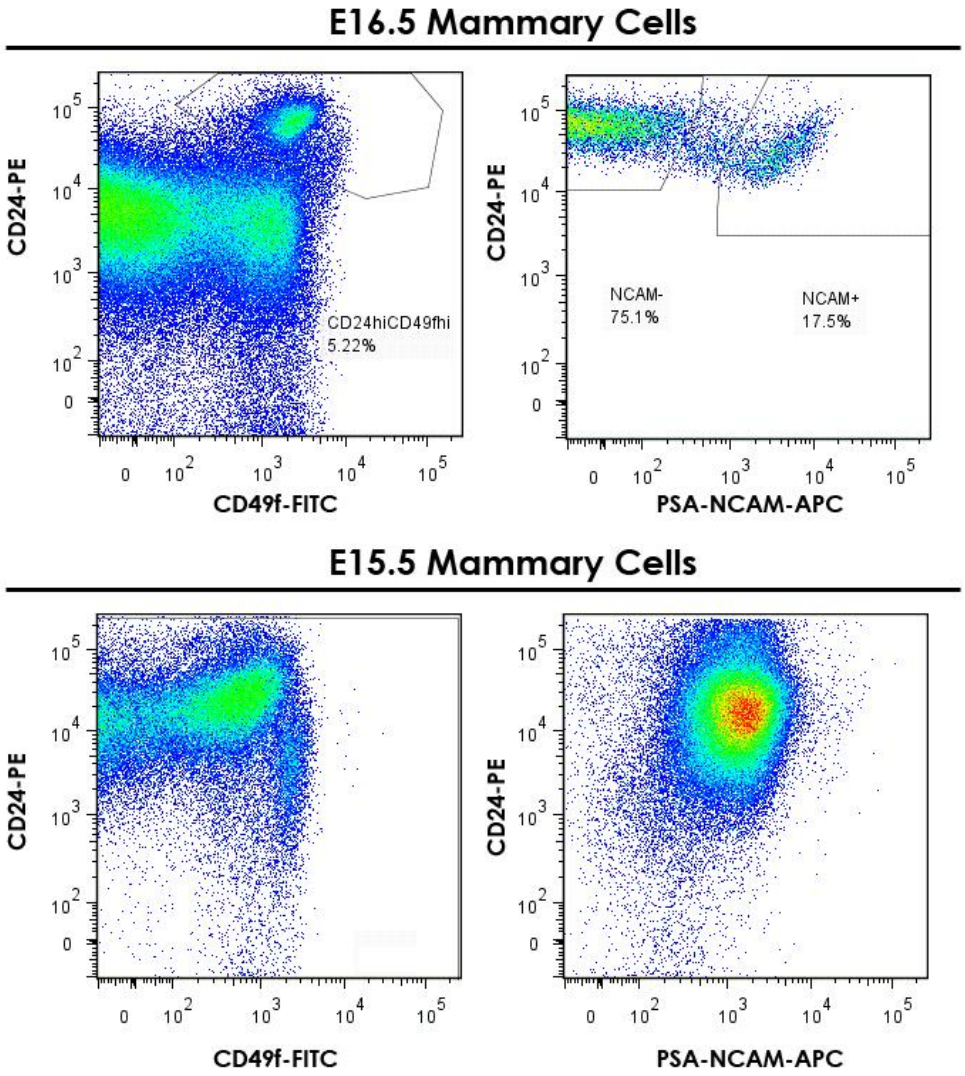


Figure 4-8. Flow cytometric analysis of embryonic mammary cells from E15.5 and E16.5 to detect the presence of PSA-NCAM.

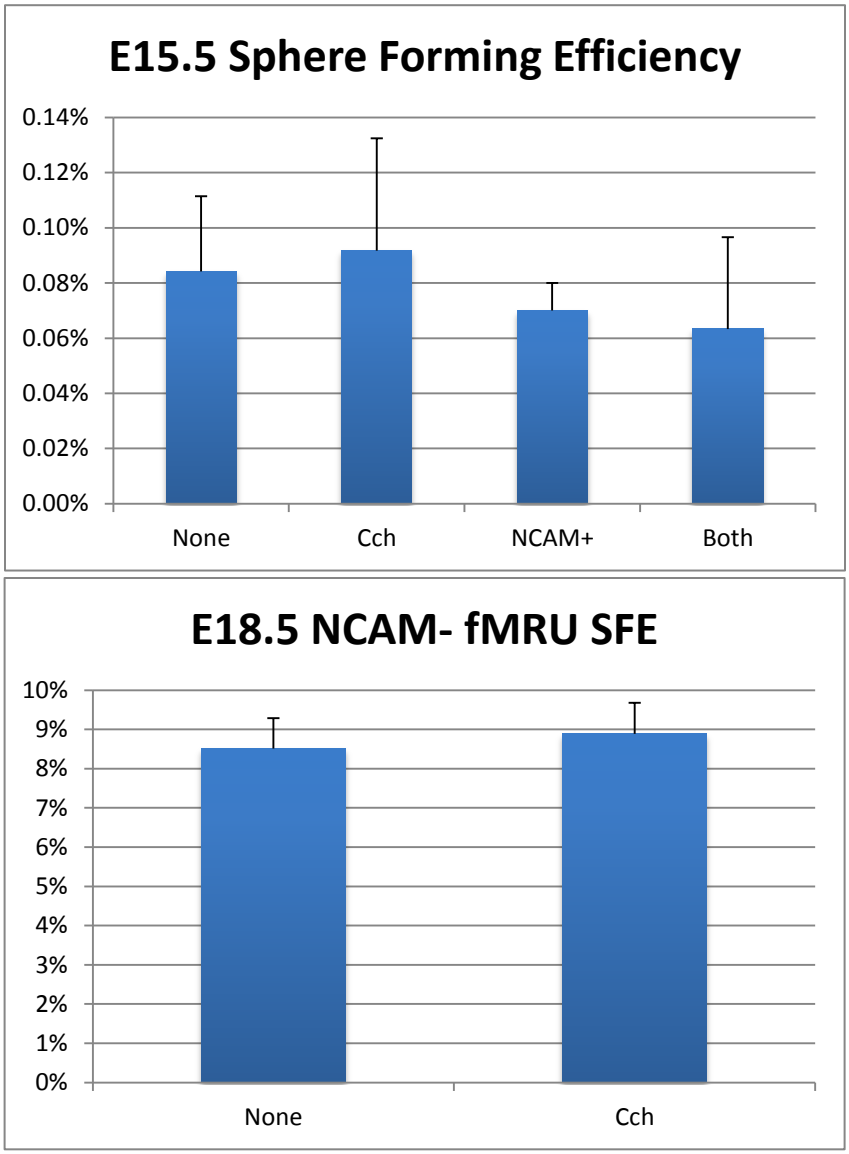


Figure 4-9. Evaluation of E15.5 and fMRU-enriched population sphere forming efficiency. E15.5 mammary cells were treated with 50nM Carbachol, an acetylcholine mimetic, 100 NCAM positive cells, or both. The NCAM negative fraction of the fMRU-enriched population was treated with 50nM Carbachol.

Table 4-2. Mammary repopulation frequency of E15.5 mammary cells transplanted with or without NCAM positive cells from the E18.5 fMRU-enriched population in the presence of Matrigel into immune-compromised hosts. The values within parentheses represent the 95% confidence interval and the italicized values indicate p value. A p value greater than 0.05 indicates that the data are consistent with a single hit model. Difference between E15.5 bulk with and without NCAM positive cells p = 0.0123. Difference between NCAM positive cells with and without E15.5 cells p = 0.056.

Cell Type	Cell #	Repopulation	MRU Frequency
$NCAM^+$ fMRU	1000	0/2	1/882 (284 – 2743) *0.000936*
	100	2/15	
	10	2/14	
E15.5 Bulk	10000	1/1	1/1232 (621 – 2444) *0.149*
	1000	2/7	
	500	4/9	
	100	3/19	
	10	0/4	
E15.5 + $NCAM^+$ fMRU	100/10	4/10	1/215 (80-580)

lacking until at approximately E16.5, when they become visible as a distinct population by immunofluorescence and flow cytometry (Figures 4-7 and 4-8).This stage corresponds to the time at which MaSC activity first begins to expand. Therefore, it is possible that similar to the inductive capabilities of the parasympathetic innervation on the progenitors of the salivary gland (Knox et al., 2010), NCAM positive cells may induce MaSC activity. Accordingly, we hypothesized that introduction of these cells to cell suspensions from E15.5 rudiments, when there is very little MaSC activity, may induce higher repopulation frequencies. Therefore, we evaluated whether NCAM positive cells from the fMRU-enriched population had an effect on E15.5 mammary cells ability to form spheres *in vitro* or their ability to regenerate mammary glands *in vivo*. E15.5 mammary cells have a very low SFE *in vitro*, which was not affected by addition of NCAM positive cells, carbachol, or both (Figure 4-9). Similarly, the SFE of the NCAM negative fraction of the fMRU-enriched population was not affected by the addition of carbachol (Figure 4-9). Limiting numbers of cells from E15.5 mammary rudiment dissociations and NCAM positive cells from the fMRU-enriched fractions were transplanted alone or together into the cleared fat pads of immune compromised animals (Table 4-2). Upon the addition of 10 NCAM positive cells to 100 E15.5 mammary cells, the MRU frequency was estimated to be 1 in 215, an approximately six fold increase over the frequency observed upon transplantation of E15.5 mammary cells alone (Table 4-2). These preliminary data demonstrate that there is an

association between the presence of NCAM positive cells and mammary repopulating activity.

Discussion

Every stem cell population characterized to date includes contributions from multiple cell types. We performed single cell qRT-PCR profiling of a highly enriched population of MaSCs in order to characterize the level of heterogeneity. This analysis revealed that despite achieving higher levels of purification than previously possible using the same strategy in the adult gland, there are several different cell types within the fMRU-enriched population. Hierarchical organization revealed a subset of cells with a gene expression pattern associated with neuronal cells. The genes expressed by these cells have been well characterized in the nervous system, but whether these cells play a functional role in MaSC specification and activity is unclear. These cells are dispensable for MaSC activity in the E18.5 rudiment. As PSA-NCAM positive cells are not detectable in early mammogenesis when MaSC activity is low, but are present upon MaSC activity expansion at E16.5, it is possible that PSA-NCAM positive cells play a role in instructing mammary cells to become competent as MaSCs.

Although NCAM positive cells were inessential for fMaSC activity at E18.5, an emerging association with MaSC activity was observed in the adult. The adult NCAM positive and negative cells from the adult MRU-enriched

population were both able to repopulate the mammary gland, however, the NCAM negative cells had a lower regrowth frequency and often were only able to generate small outgrowths that did not fully fill the fat pad, a phenotype commonly observed with transplanting the same number of total cells from the aMRU-enriched population. On the other hand, transplantation of NCAM positive cells from the adult MRU-enriched population yielded outgrowths more frequently that also reliably repopulated the entire fat pad. It is conceivable that the aMRUs either acquire neuronal like characteristics or their activity depends on input from NCAM positive cells. Upon co-transplantation of NCAM positive cells from the fMRU-enriched population with mammary cells from E15.5, an increase in MRU frequency was observed. It is conceivable that NCAM positive cells have a role in activating a MaSC state such that in the E18.5 mammary gland when the MaSCs are already activated, their presence is dispensable. However, in situations where MaSCs have not yet become active, such as the E15.5 rudiment, or states in which the MaSCs may be dormant (the homeostatic adult gland), association with NCAM positive cells may induce stem cell activity. As this data is preliminary, further dissection of the contribution of these previously uncharacterized, NCAM positive, neuronal-like cells to MaSC activity during development and adulthood is required.

Although microarray analyses are powerful, a significant amount of information is lost through population averaging. These studies demonstrate

the influence of multiple cell types on the fMRU-enriched population gene expression profile. Examination of this population as an average precludes the ability to determine whether genes that are highly expressed within a heterogeneous population are due to many cells coordinately up-regulating gene expression to a moderate level, or a few cells expressing very high levels of different transcripts. The high expression levels of neuronal associated genes detected in the fMRU-enriched population by microarray analysis resulted from the presence of a distinct subset of cells expressing these genes and exhibiting lower stem cell activity than the rest of the population. Therefore, it is difficult to deduce what is a property of a stem cell in contrast to a contaminating cell type with no stem cell activity from microarray analysis alone. Single cell analysis of tissue stem cell pools will be critical in dissecting the different cell type contributions to these heterogeneous populations.

Acknowledgements

This chapter, in part, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. * indicates equal contribution. The dissertation author was one of the primary authors of this paper.

Chapter 5. Molecular links between fMaSCs, development and human breast cancer

Background

Breast cancer mortality has decreased significantly due to numerous factors including the cessation of replacement hormone therapy, implementation of screening programs, and new targeted treatment strategies (Jemal et al., 2009). Unfortunately, breast cancer still causes approximately 40,000 deaths in the United States every year (Jemal et al., 2009). Breast cancer is an extremely heterogeneous disease and clinical metrics to evaluate the distinct phenotypes present within this disease and to suggest treatment options are limited (Prat and Perou, 2011). Although parameters including grade, lymph node status, and tumor size are clear predictors of prognosis, they do not offer any information to guide individual patient treatment. Pathological markers such as Estrogen Receptor, Progesterone Receptor, and ErbB2 (EGFR2, Her2, Neu) offer more information in terms of prescribing anti-estrogen and anti-ErbB targeted therapies, but patients continue to succumb to clinical progression of the disease. Little is known in terms of how to stratify patients within the same clinico-pathological class, as patients with the same parameters can have significantly different clinical courses.

Dissecting the biological complexity of breast cancer with global gene expression analysis

Breast cancer encompasses a wide extent of biologic diversity. Recent approaches to stratify breast cancer patients have utilized microarray analyses to query global gene expression patterns in individual tumors. Microarray analyses offer comprehensive molecular profiling and have the potential to elucidate meaningful biological classifications that are more clinically useful, both prognostically and therapeutically. Over the years, microarray measurements have been established as both reproducible and reliable within and between different microarray technologies (Sotiriou and Piccart, 2007). In fact, the variation among microarray analyses is comparable to those reported for immunohistochemical assessment of the hormone receptors (Sotiriou and Piccart, 2007).

Using unsupervised cluster analysis, distinct subtypes of breast cancer can be distinguished by correlating gene expression patterns with signatures developed from breast myoepithelial and luminal epithelial cell lines and subsequent association with clinical outcomes (Perou et al., 2000b). This clustering revealed several different molecular intrinsic subtypes of breast cancer, consisting of Luminal A, Luminal B, Her2-enriched, Normal-like, Claudin-Low, and Basal-like, which reveal critical differences in survival and response to treatment (Prat and Perou, 2011). Interestingly, basal-like breast cancer includes a subset of tumors characterized as triple negative (ER^-PR^-

HER2$^-$) and is often characterized by its undifferentiated stem-like phenotype in addition to frequently co-expressing myoepithelial and luminal epithelial keratins as well as the mesenchymal marker vimentin by immunohistochemistry (Livasy et al., 2006). Several prognostic signatures have been refined and subsequently adopted in the clinic, including Mammaprint (Agendia, NKI 70 gene signature) (van 't Veer et al., 2002; van de Vijver et al., 2002a) and OncotypeDX Recurrence Score (RS, Genomic Health) (Fan et al., 2011; Paik et al., 2004). Although these analyses have revealed a wealth of information, they correlate closely with hormone receptor status and proliferative markers (Desmedt et al., 2008). In addition, these signatures are similar in terms of their outcome predictions although they do not share gene identities. This suggests that although these signatures are composed of unique gene sets, they report similar biologic programs (Fan et al., 2006; Haibe-Kains et al., 2008; Prat and Perou, 2011). Ways to stratify and treat patients who are triple negative for ER, PR, and Her2, which have very aggressive disease and poor prognosis, are limited. Altogether, these studies represent a retrospective analysis of archival breast cancer microarray data that have yielded informative signatures for early stage, ER positive breast cancer patients. However, signatures that identify patients with aggressive disease that is likely to metastasize are needed. Indeed, these signatures have had limited success in predicting optimal therapeutic strategies in patients of diverse breast cancer subtypes.

Biology driven prospective signature identification has also been applied to breast cancer. These signatures represent biological processes including cell proliferation, cell differentiation, and wound responses (Prat and Perou, 2011). Other signatures have been generated to report association of particular cell types, such as immune cells and fibroblasts, and phenotypes including therapeutic response (Prat and Perou, 2011). Pathway activation status has also been utilized for the development of prognostic signatures (Prat and Perou, 2011). However, most of these biological process driven signatures funnel into the same set of parameters in that they are composed predominantly of proliferation associated genes. Regardless, they are useful for gauging the risk of recurrence in some breast cancer subtypes (Sotiriou and Piccart, 2007).

Out of all of these predictive gene signatures, there has been very little progress in terms predicting which individual patients need systemic adjuvant therapy and the likelihood of different treatment effectiveness. For example, the development of anti-Her2 therapeutics has significantly increased the survival of Her2 positive breast cancer patients. However, as a single line agent, Trastuzumab, a Her2 targeted antibody, has a clinical response rate of only 26% in patients overexpressing this protein (Vogel et al., 2002). Moreover, some patients benefit from adding Trastuzumab to their treatment regimen despite being clinically designated and validated as Her2 negative (Paik et al., 2008). Her2 copy number does not predict response and even

patients with normal gene copy numbers benefited from Trastuzumab (NSABP B-31, (Paik et al., 2008)). Therefore, by using this criterion to select patients eligible for Trastuzumab treatment, certain patients who could benefit from targeted therapy are effectively prevented from pursuing this treatment option. Therefore, it is of utmost importance to continue to develop signatures to explain the underlying biology behind these outcomes.

Results

We performed microarray expression analyses on fMRU-, fStromal-, and aMRU-enriched populations to gain insight into molecular pathways with potential importance for fetal mammary development, stem cell biology, and cancer. We obtained reproducible expression profiles from independent biological replicate pools representing each population. This facilitated the identification of many differentially expressed genes whose relative expression was corroborated by independent qRT-PCR based analyses (Figure 5-1 and 5-2). In addition to many genes whose expression and function in the developing mammary has not been previously studied, we also recognized many genes and ontological categories expressed in a manner consistent with the cell types analyzed. For instance, we identified augmented expression of cell cycle and synthetic metabolism genes in the growing fetal tissue relative to the adult (Figure 5-3). Comparisons of the profiles obtained from the fMRU- and fStromal-enriched populations showed the expected segregation of

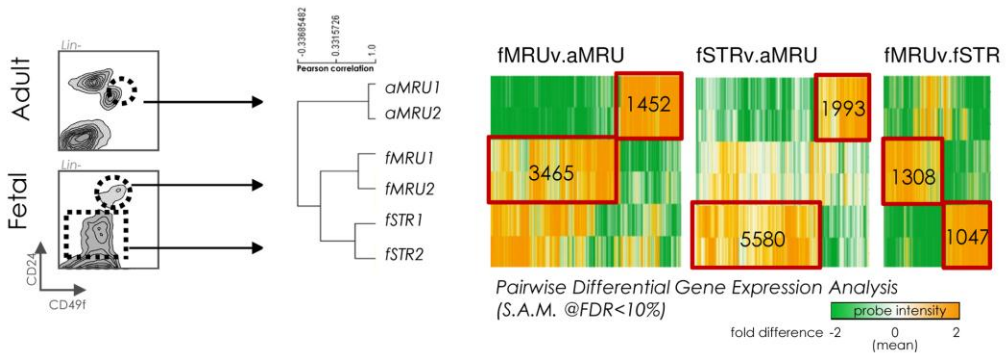


Figure 5-1. Illustration of sorted populations, correlation among biological replicates for each cell type and heat maps illustrating the identification of differentially expressed genes.

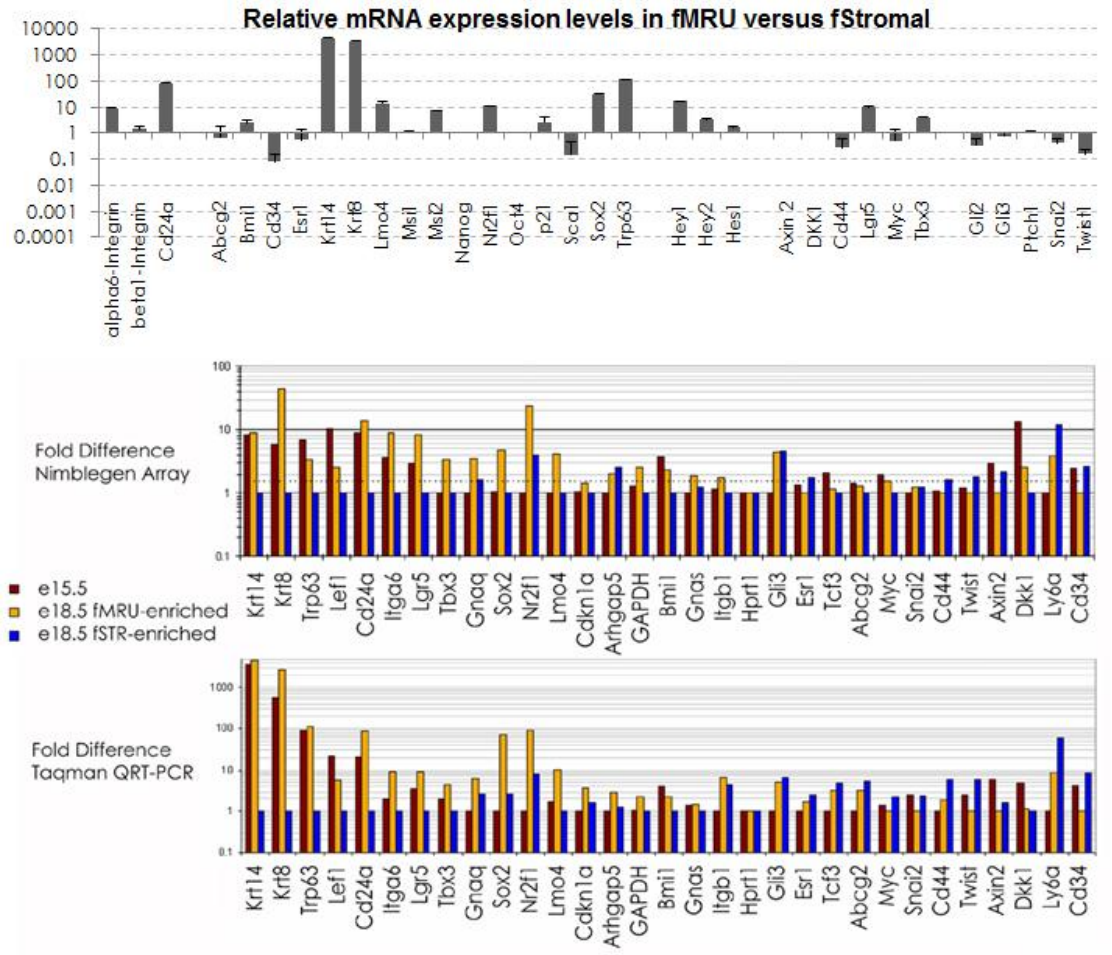
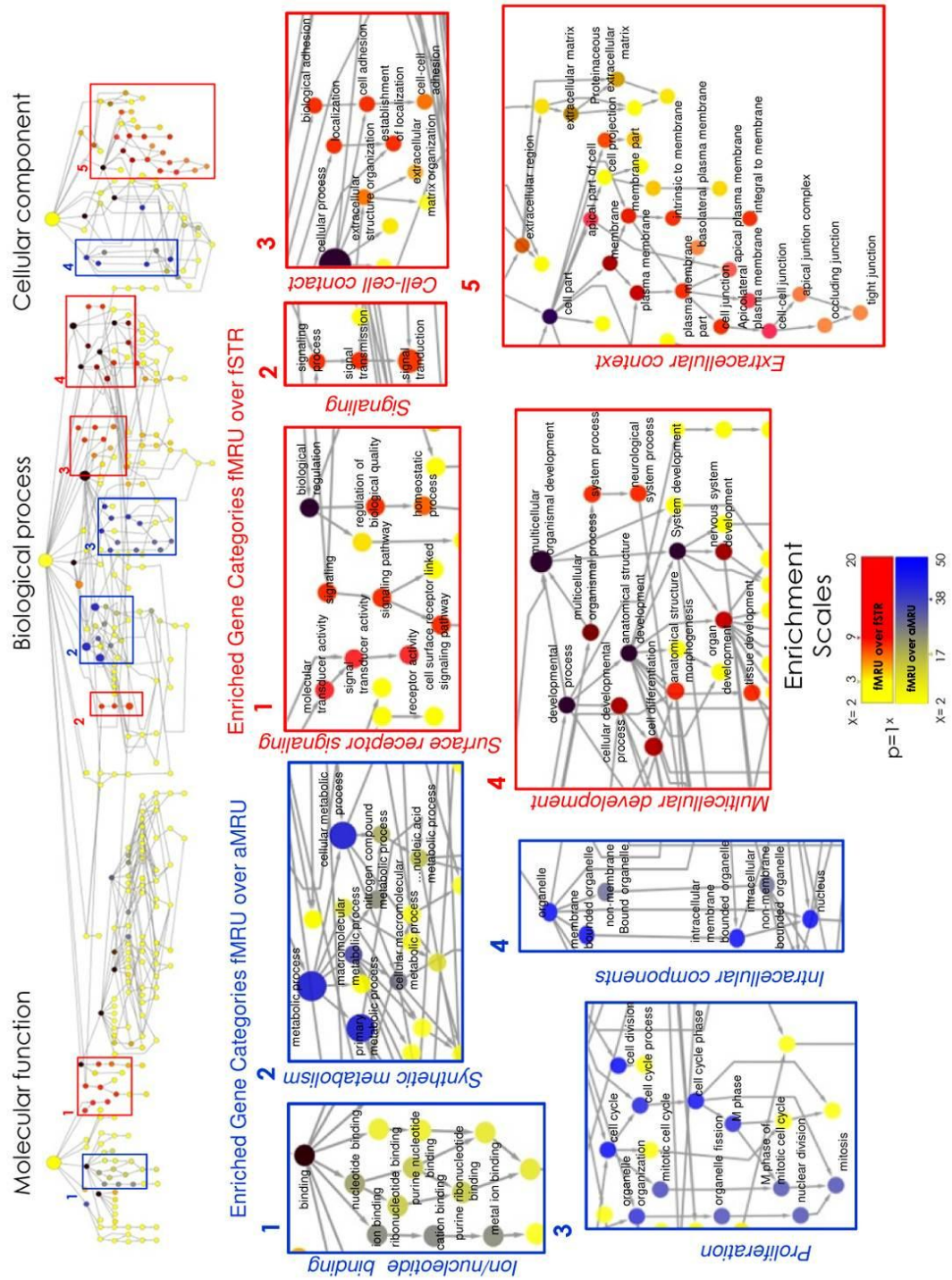


Figure 5-2. qRT-PCR analyses confirm gene expression differences between fMRU- and fStromal-enriched populations. **Top Panel.** Select genes were analyzed by qRT-PCR to validate their gene expression differences. All data are normalized to HPRT. Comparisons are reported as expression in the fMRU-enriched population over the fStromal-enriched population. Fold differences in gene expression were calculated assuming ideal amplification (fold change= $2^{\Delta Ct}$). **Bottom Panels.** Fold expression is presented relative to the sample with the lowest expression of each gene. The same process was carried out using probe intensities from Nimblegen array data (top panel). The direction of change was highly consistent between assay methods. The estimated magnitude of gene expression may vary due to differences between biological replicates, assay platforms, primer versus probe location, and the less than log2 RT-PCR fidelity.

Figure 5-3. Gene Ontology Analysis reveals prominent characteristics defining the fMRU-enriched population. The fMRU-signature exhibits augmented expression of proliferation and synthetic metabolism genes relative to adult and shows over representation of genes involved in extracellular signaling relative to both the aMRU- and fStromal-enriched populations.


Molecular function
Biological process
Cellular component
Enriched Gene Categories fMRU over aMRU
Enriched Gene Categories fMRU over fSTR
Ion/nucleotide binding
Synthetic metabolism
Surface receptor signaling
Signaling
Cell-cell contact
Extracellular context
Multicellular development
Intracellular components
Proliferation
Enrichment Scales
p=1^x
fMRU over fSTR
fMRU over aMRU

epithelial gene expression with the fMRU-enriched population and elevated expression of many mesenchymal genes in the fStromal-enriched population (Figure 5-3). We also noted the presence of genes corresponding to several molecular pathways commonly associated with early mammary development (Figure 5-2 and 5-3) (Veltmaat et al., 2003). We found 1308 probes to be more highly expressed in the fMRU-enriched population (designated as the fMRU signature) and 1047 probes to be more highly expressed in the fStromal-enriched population (the fSTR signature) when these two populations were compared (Figure 5-1). The comparative signatures associated with the fMRU- and fStromal-enriched population exclude genes shared by both populations. Theoretically, this comparative signature generation should at least partially exclude contribution from shared cell types, including contaminating cells with no stem cell activity in the fMRU-enriched population. Of the 1308 fMRU signature probes, 530 were significantly overexpressed in the fMRU-enriched population relative to the aMRU population. The fMRU signature represents 869 unique genes and 654 known human orthologues, and the fSTR signature represents 812 unique genes and 722 human orthologues that were used for subsequent enrichment analyses.

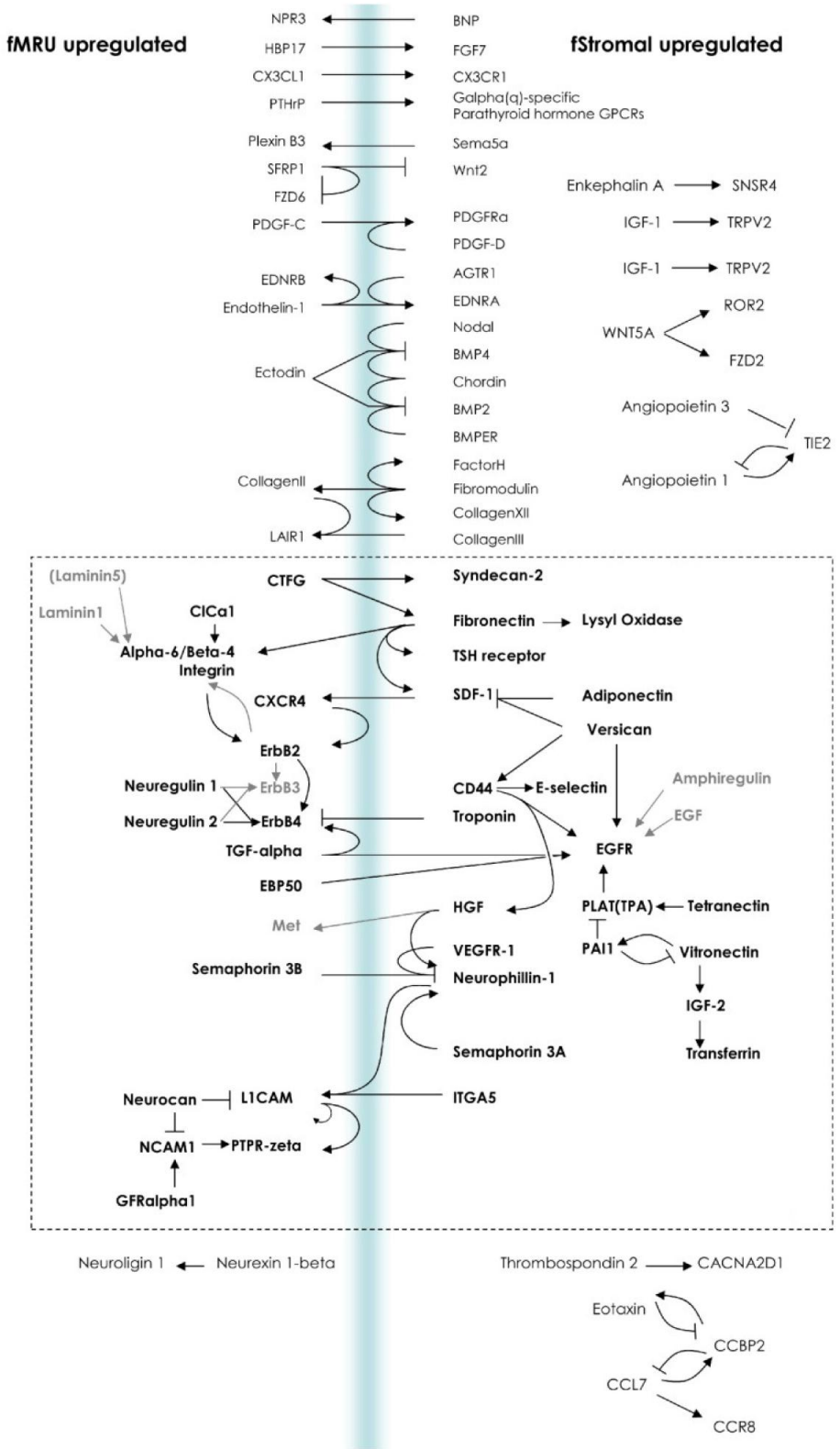
Gene ontology enrichment analysis of fetal signatures supports a prominent role for cell-cell and cell-matrix interactions as well as for cell surface receptor signaling (Figure 5-3) in the fMRU-enriched population. These interactions are consistent with the sensitivity of MaSC activity to

developmental and microenvironmental contexts described above. We used the GeneGo pathway analysis platform to organize the genes comprising the fMRU and fSTR signatures into potential receptor-ligand interactions. The fMRU- and fStromal-enriched populations reciprocally express many receptor-ligand pairs recorded in the GeneGo database, enabling construction of a hypothetical interaction map based on reported receptor-ligand interactions (Figure 5-4). We note that many of these receptor-ligand pairs have been implicated in tumor growth, invasion and chemo-resistance (Figure 5-4) (Luker and Luker, 2006; Sebens Muerkoster et al., 2006).

ErbB signaling in the fMRU-enriched population

We focused our immediate attention on ErbB family members because of the relevance of this pathway to cellular transformation (Korkaya et al., 2008; Lenferink et al., 2000) and cancer treatment (Hynes and Lane, 2005). Furthermore, the contribution of ErbB signaling to normal MaSC function remains to be resolved (Asselin-Labat et al., 2010; Cicalese et al., 2009b; Pece et al., 2010; Stern, 2003). ErbB2 (also known as Her2 or Neu) and ErbB3 mRNA expression were significantly elevated in the fMRU-enriched population relative to the fStromal- and aMRU-enriched populations as determined by qRT-PCR (Figure 5-5). ErbB2 protein was also readily detectable *in situ* in E18.5 mammary rudiments (Figure 5-5). Furthermore, we found that *in vitro* sphere formation by the fMRU-enriched population was

Figure 5-4. A model constructed from fetal gene signatures filtered for receptors and ligands using the GeneGo pathway analysis platform. The model illustrates candidate protein-protein interactions including receptor-ligand pairs expressed reciprocally in the fMRU- and fStromal-enriched populations. Additional gene products of interest predicted to interact with the network are also indicated (gray). The map suggests that ErbB signaling, among other pathways, may play a prominent role in fMRU function.


fMRU upregulated
fStromal upregulated
NPR3
BNP
HBP17
FGF7
CX3CL1
CX3CR1
PTHrP
Galpha(q)-specific
Parathyroid hormone GPCRs
Plexin B3
Sema5a
SFRP1
Wnt2
FZD6
Enkephalin A
SNSR4
PDGF-C
PDGFRa
PDGF-D
IGF-1
TRPV2
IGF-1
TRPV2
EDNRB
AGTR1
Endothelin-1
EDNRA
ROR2
WNT5A
Nodal
FZD2
Ectodin
BMP4
Chordin
Angiopoietin 3
BMP2
TIE2
BMPER
Angiopoietin 1
FactorH
CollagenII
Fibromodulin
CollagenXII
LAIR1
CollagenIII
(Laminin5)
CTFG
Syndecan-2
Laminin1
ClCa1
Fibronectin
Lysyl Oxidase
Alpha-6/Beta-4
Integrin
TSH receptor
CXCR4
SDF-1
Adiponectin
Versican
ErbB2
Neuregulin 1
ErbB3
CD44
E-selectin
Amphiregulin
Neuregulin 2
ErbB4
Troponin
EGF
TGF-alpha
EGFR
EBP50
HGF
PLAT(TPA)
Tetranectin
Met
PAI1
VEGFR-1
Vitronectin
Semaphorin 3B
Neurophillin-1
IGF-2
Semaphorin 3A
Transferrin
Neurocan
L1CAM
ITGA5
NCAM1
PTPR-zeta
GFRalpha1
Neuroligin 1
Neurexin 1-beta
Thrombospondin 2
CACNA2D1
Eotaxin
CCBP2
CCL7
CCR8

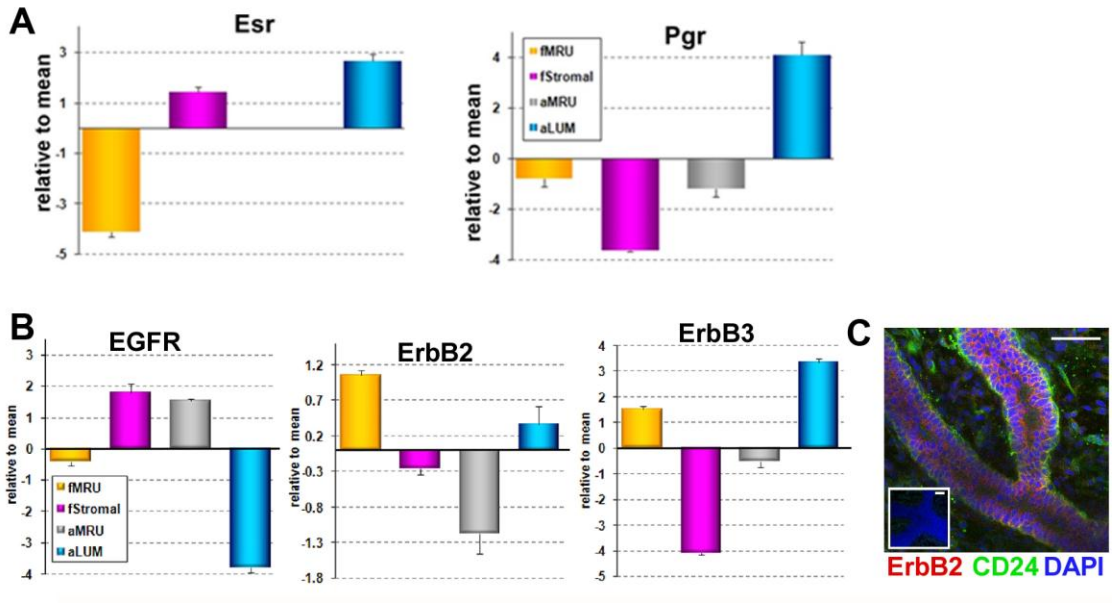


Figure 5-5. A. Bar chart showing qRT-PCR analysis of Estrogen and Progesterone receptors. **B.** Bar chart showing qRT-PCR analysis of the ErbB family receptors in fMRU-enriched, fStromal-enriched, adult MRU-enriched (aMRU), and adult luminal (aLUM) cells. **C.** A confocal image of immunofluorescence on an E18.5 mammary rudiment showing cells expressing ErbB2 and CD24. Inset, secondary antibody control. Scale bar, 50µm.

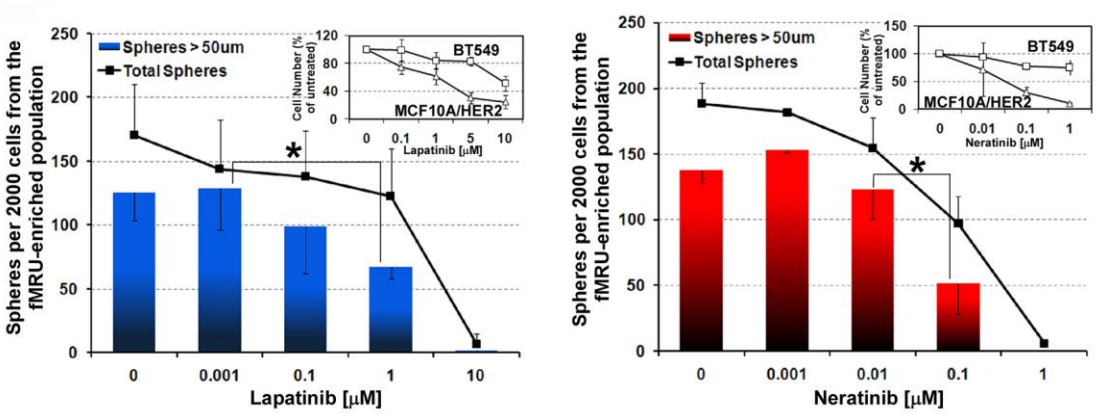


Figure 5-6. Bar charts showing the quantification of fMRU-derived sphere growth upon inhibition of ErbB1/2 signaling by either Lapatinib or Neratinib. Inset, dose response curves to Lapatinib and Neratinib in both resistant BT549 and sensitive MCF10A/HER2cell lines (Wang et al., 2006; Weigelt et al., 2010). *p<0.05, Student's t test. Error bars, standard deviation.

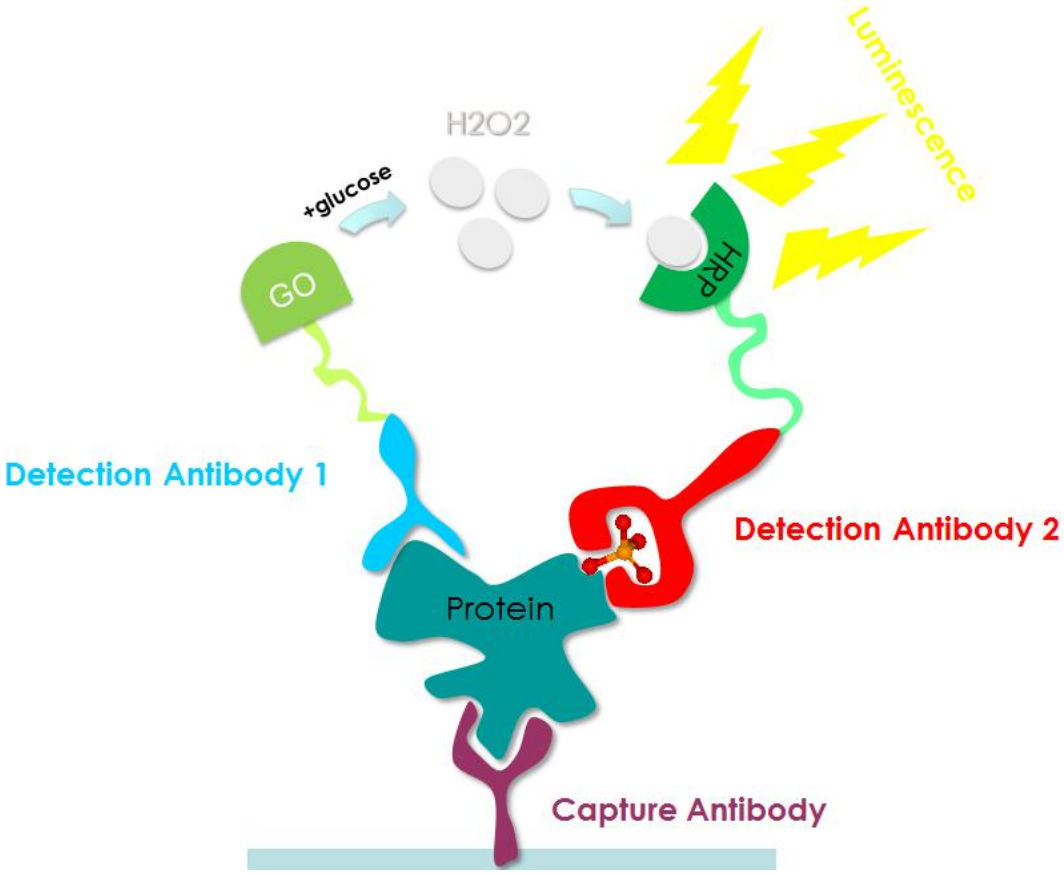
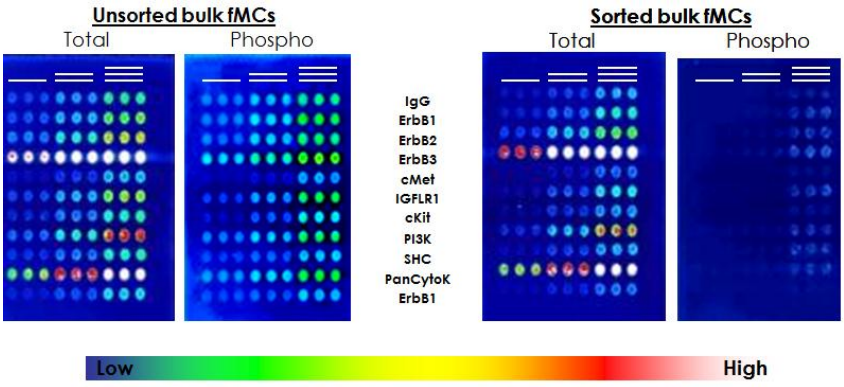
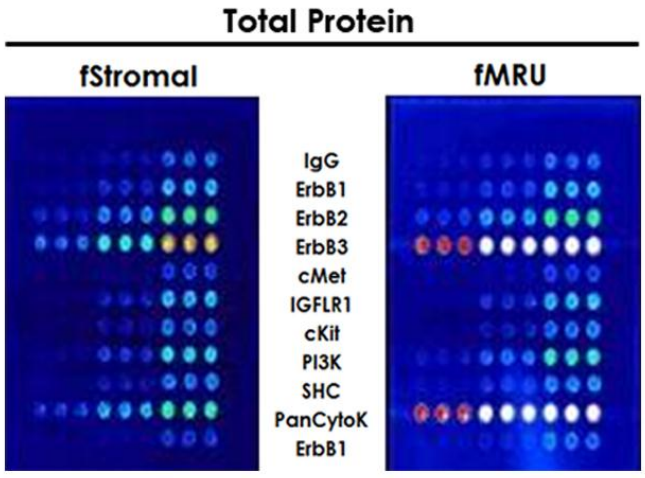
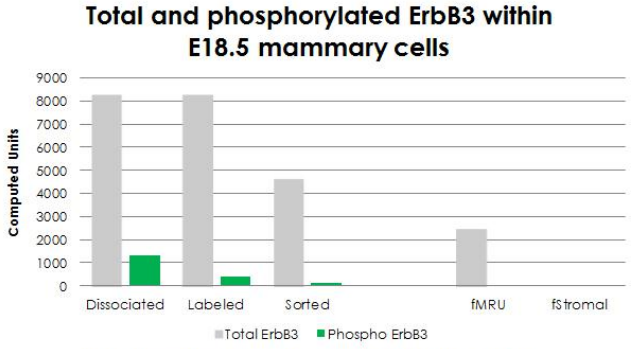


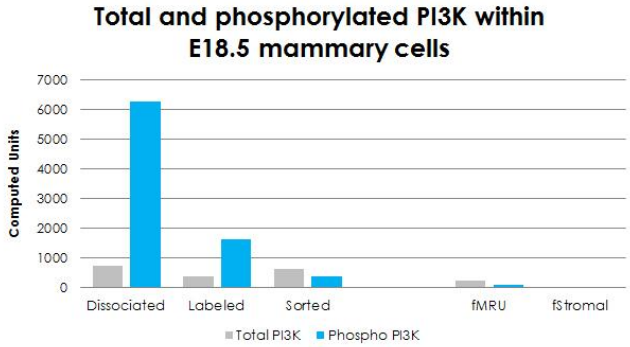
Figure 5-7. Schematic of the CEER assay from Prometheus for the detection of proteins from limited samples.

Figure 5-8. Top panel. Evaluation of protein levels in mammary cells from E18.5 mammary rudiments. Rudiments were dissociated, labeled, and then sorted as one population and samples were obtained at each step to evaluate the effects of processing on protein concentration and phosphorylation status. Shown are the dissociated, unsorted mammary cells and the same cells several hours later upon completion of the sorting processing step**. Middle Panel.** Evaluation of total protein levels in the fMRU-enriched and fStromal-enriched populations. **Bottom panel,** Quantification in computed units (CU) of the expression levels of ErbB3 and PI3K in the different cell populations.


Unsorted bulk fMCs
Total
Phospho
Sorted bulk fMCs
Total
Phospho
IgG
ErbB1
ErbB2
ErbB3
cMet
IGFLR1
cKit
PI3K
SHC
PanCytoK
ErbB1
Low
High


Total Protein
fStromal
fMRU
IgG
ErbB1
ErbB2
ErbB3
cMet
IGFLR1
cKit
PI3K
SHC
PanCytoK
ErbB1


Total and phosphorylated ErbB3 within E18.5 mammary cells
Computed Units
9000
8000
7000
6000
5000
4000
3000
2000
1000
0
Dissociated
Labeled
Sorted
fMRU
fStromal
Total ErbB3
Phospho ErbB3


Total and phosphorylated PI3K within E18.5 mammary cells
Computed Units
7000
6000
5000
4000
3000
2000
1000
0
Dissociated
Labeled
Sorted
fMRU
fStromal
Total PI3K
Phospho PI3K

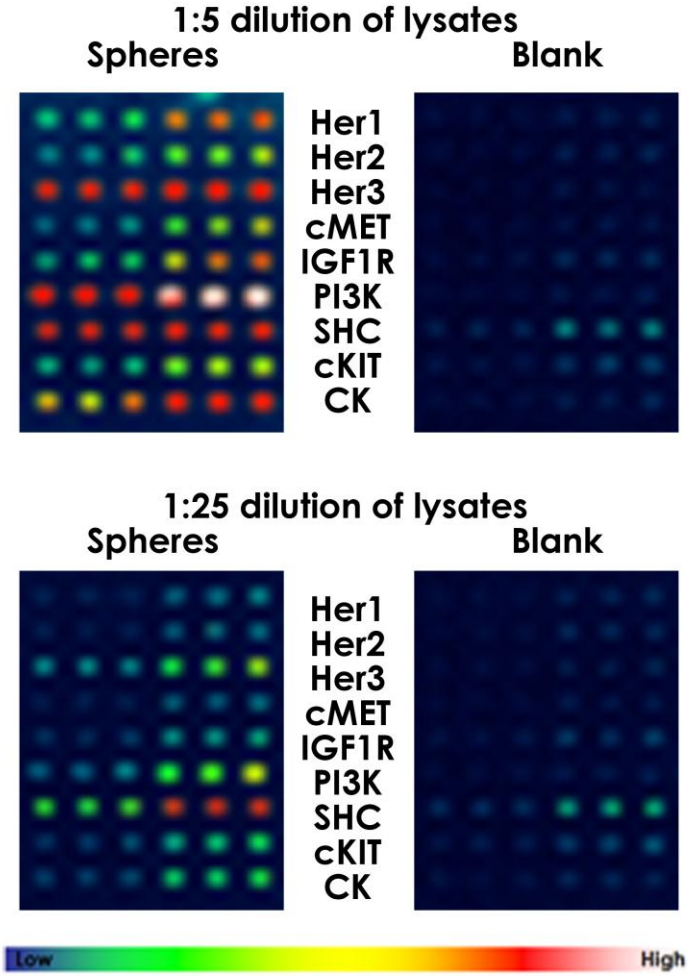


Figure 5-9. Evaluation of phosphorylation status of different proteins using the Prometheus CEER assay. Spheres were generated from the fMRU-enriched population and the phosphorylation status was determined relative to blank controls.

inhibited by the reversible ErbB1/2 dual tyrosine kinase inhibitor Lapatinib (Rusnak et al., 2001) and by the irreversible pan-ErbB kinase inhibitor Neratinib (Figure 5-6) (Rabindran et al., 2004). Sphere forming efficiency showed the same inhibition kinetics as a human cell line over-expressing Her2, which is known to be very sensitive to the effects of these inhibitors (inset, Figure 5-6) (Wang et al., 2006).

In order to determine the specificity of this effect, we collaborated with Prometheus and utilized their CEER assay to detect small quantities of protein with high specificity (Figure 5-7). This strategy utilizes a multiplexed, proximity-based, collaborative immune assay platform. Briefly, the protein lysate is washed over a microarray surface printed with a capture antibody to your protein of interest. Two additional antibodies that recognize distinct domains or phosphorylation sites are incubated. These antibodies are conjugated to two different enzymes, glucose oxidase and horseradish peroxidase. Glucose oxidase will produce hydrogen peroxide upon addition of glucose. The close proximity of the hydrogen peroxide enables generation of chemiluminescence by horseradish peroxidase. In these analyses, we determined that phosphorylation signals were rapidly lost during the processing steps required to isolate the fMRU-enriched population, despite numerous strategies to preserve these signals (Figure 5-8). However, we were able to discern that the fMRU-enriched population expresses high levels of Erbb3 and also has detectable levels of ErbB1 and ErbB3 phosphorylation (CU=18.4 and 12.5,

respectively). Given the loss of the phosphorylation signals in the fMRU-enriched population, we analyzed spheres generated by this population as an alternative. We found that the fMRU-enriched population derived spheres had high levels of ErbB1 and ErbB3 phosphorylation and high levels of PI3K activation (Figure 5-9). Altogether, these data demonstrate the importance of ErbB signaling for growth of fMRU-derived spheres *in vitro*, which is in stark contrast to the previously reported insensitivity of the aMRU-enriched population to ErbB kinase inhibitors *in vivo (*Asselin-Labat et al., 2010).

Fetal mammary-like gene expression in specific breast cancer subtypes.

The enrichment of cancer associated genes and pathways among our fetal gene signatures prompted us to search for additional molecular links between fetal mammary development and tumorigenesis. Breast cancers have been divided into ‘intrinsic subtypes’ correlating with diverse histo-pathologies and clinical outcomes based on differences in their gene expression profiles (Perou et al., 2000a). Therefore, we determined whether fetal gene expression signatures were enriched in molecularly classified human breast cancer subtypes using archival tumor micro-array data. We identified many tumors significantly enriched or depleted for the fMRU or fSTR signature genes (Figure 5-10), implying that gene expression patterns related to fetal development in both compartments correlate with tumor phenotypes. Significant enrichment for the fMRU signature was concentrated among

tumors designated as “basal-like”, which tend to be poorly differentiated and stem-like (Ben-Porath et al., 2008; Mizuno et al., 2010) and are frequently ‘triple negative’ by clinical criteria [i.e. ER negative, PR negative, ErbB2 negative (un-amplified but often expressing low levels of ErbB2) (Jacobs et al., 1999)]. In addition, Her2+ tumors frequently showed significant enrichment for the fMRU signature (Figure 5-10). In contrast to the fMRU enrichments, we noted that enrichment for the fSTR signature often correlated with tumor subtypes characterized by low proliferation and favorable prognoses (Figure 5-10) (Sorlie et al., 2001). We also noted that Claudin-low and Metaplastic-like tumors, which have been suggested to be stem-like (Hennessy et al., 2009), were generally enriched for the fSTR signature and depleted for the fMRU signature (Figure 5-10).

The $CD24^{high}CD49f^{high}$ fMRU-enriched population has a gene expression signature most similar to putative human MaSCs that exhibit high mammosphere forming efficiency and that also express high levels of CD24 and CD49f (Figure 5-11) (Pece et al., 2010). In contrast, the fSTR signature most closely resembles other adult mammary populations previously acknowledged to contain stromal and/or basal components including some populations designated as aMaSC populations (Figure 5-11) (Kendrick et al., 2008; Lim et al., 2009; Lim et al., 2010; Pece et al., 2010; Stingl et al., 2006b). The latter observation likely reflects the difficulty of effectively purifying the rare adult MaSCs from their associated stroma. It is noteworthy that breast

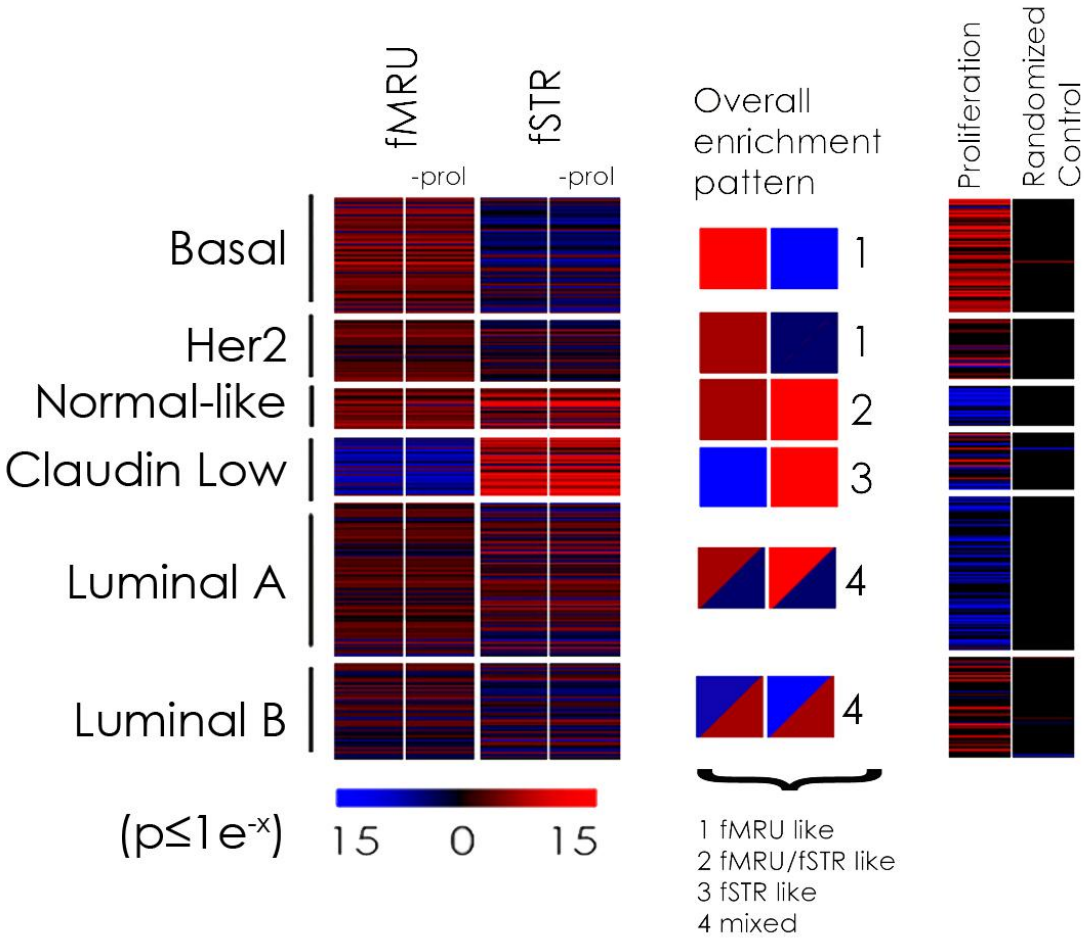


Figure 5-10. Significant correlation between fMRU and fSTR gene signatures and human breast cancers (Prat et al., 2010) are indicated by horizontal bars, each representing the gene expression profile from an individual tumor sample. Red bars indicate tumors enriched in fetal signature expression, blue bars indicate signature repression. Black bars indicate no significant correlation. Larger colored squares illustrate the trend for each intrinsic subtype. For comparison, a randomized signature of equivalent size and a proliferation signature (Ben-Porath et al., 2008) are shown.

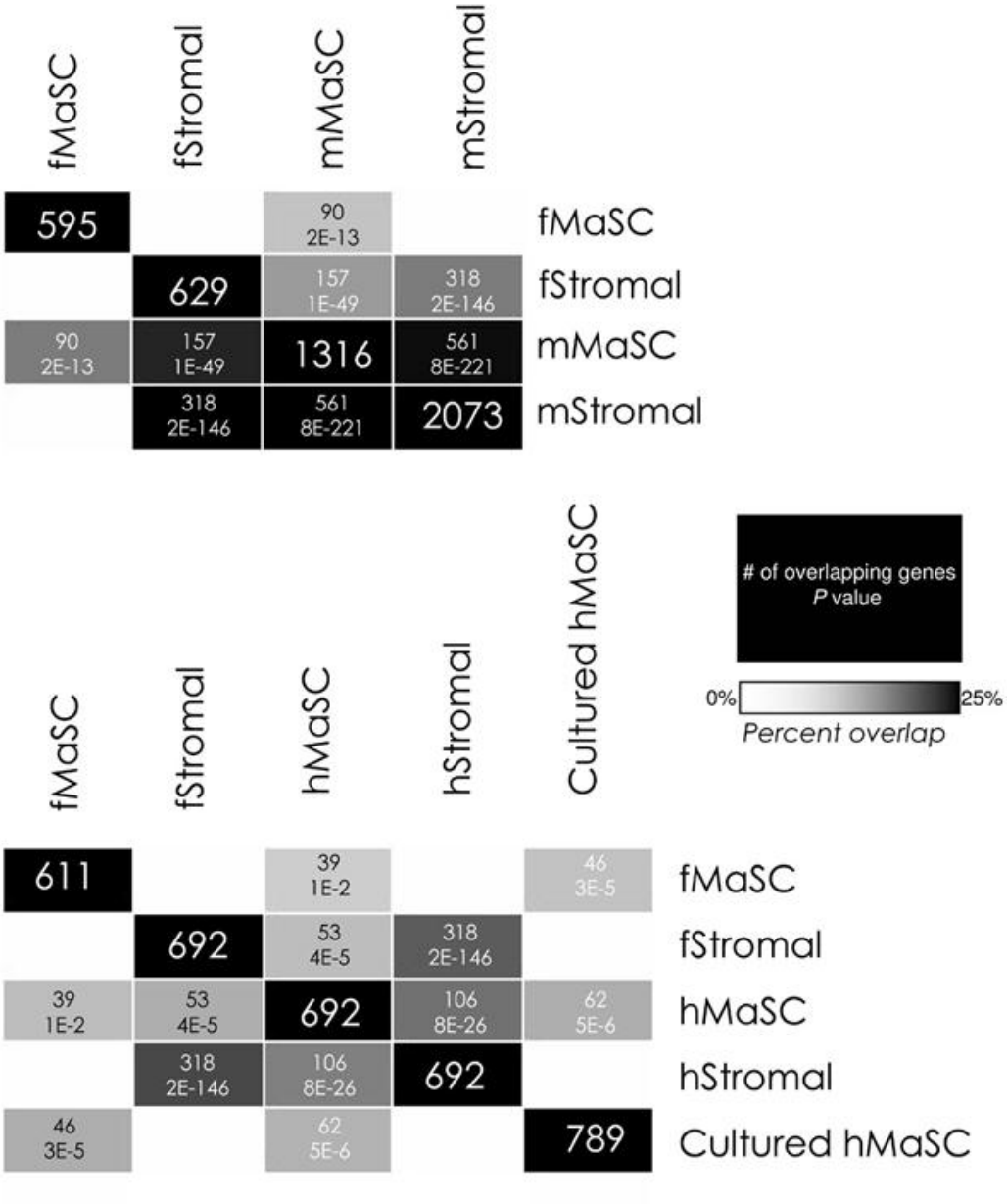


Figure 5-11. Overlap of fMRU signature genes and their orthologues with previously reported normal adult mammary signatures including stem cell enriched signatures. The upper panel shows mouse signatures (Lim et al., 2009; Pece et al., 2010)**.** The lower panel shows hMaSC and hStromal signatures are from Lim et al., (Lim et al., 2010). Cultured hMaSCs are from Pece et al., (Lim et al., 2009; Pece et al., 2010)**;** p-values represent the hypergeometric probability based on all 20,309 probes in the mouse array and 19,828 probes in the human arrays.

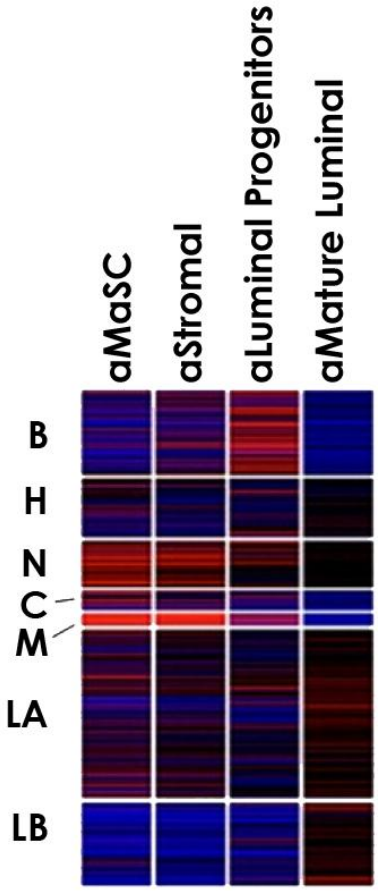


Figure 5-12. Adult “MaSC” and Stromal signatures show enrichment patterns in tumors that are very similar to those observed for the fetal non transplanting population (fStromal-enriched population) rather than the fMRU-enriched population. The subtraction of genes from the proliferation signature reported in Ben-Porath et al. (Ben-Porath et al., 2008) indicates a negligible effect of proliferation signature genes on fetal signature enrichment among breast tumors.

cancers showed nearly identical enrichment patterns for the stromal and stem cell-enriched signatures from these studies (Figure 5-12).

Several gene signatures showing enrichment for different mammary tumor types have been derived previously either directly from breast cancer array data, from specific biological contexts such as serum stimulation of fibroblasts (indicative of wound healing), or from the adult mammary cell hierarchy (Fan et al., 2011). The fMRU signature is novel as it has very little overlap between its gene constituents and those comprising several of the aforementioned signatures (Figure 5-13). Specifically, while many signatures have significant representation of proliferation-associated genes (Fan et al., 2006; Wirapati et al., 2008), these are poorly represented in the fMRU signature (Figure 5-13). This is expected since we derived our signature based on differential gene expression between two proliferating fetal populations. This led to the exclusion of many proliferation and cell cycle associated genes expressed in both populations. Furthermore, we confirmed that enrichment for fetal-specific signatures is not due to the influence of residual proliferation associated genes since removing them from the fSTR or fMRU signatures did not markedly alter overall tumor enrichments (Figure 5-10). This was consistent with the observed difference between the fMRU enrichment pattern and the pattern of enrichment for an independent proliferation signature presented as a control (Figure 5-10). Thus, the fMRU signature enabled us to

evaluate the representation of genes associated with other aspects of fetal mammary gland biology and stem cells in cancer as described below.

Unique fetal gene subsets correlate with breast cancers of diverse clinical prognosis

The stage of development with the highest fMRU activity correlates with invasion of the developing rudiment into the underlying fat pad. This process resembles aspects of metastasis and some of the cells within the fMRU-enriched population exhibit characteristics associated with aggressive breast cancers. Therefore, we determined whether the fMRU and/or fSTR signatures could be further sub-divided into modules correlating with breast cancer progression, subtype or prognosis. We used hierarchical clustering of the 96 most highly up-regulated genes within each fetal signature and tumor samples to obtain gene subsets that we used to correlate with tumor grade and subtype (Figure 5-14, 5-15, and 5-16). We subdivided the fMRU and fSTR signatures into five and four modules, respectively, that segregated patient samples in different ways such that certain modules were up or down regulated in specific subtypes and grades of breast cancer. In this way, we were able to ascertain prognostic relevance in univariate survival analyses as independent variables (Figure 5-17). As an example of how these modules operate in different intrinsic subtypes of breast cancer, enrichment in signature fMRUii or repression of signature fSTRiv correlates with Her2+ and basal-like tumors,

high grade, and reduced probability of patient survival (Figure 5-16). This is consistent with the predicted outcome of intrinsic subtypes (Sorlie et al., 2001).

As fMRU signatures are derived from cells with low levels of ER and PR expression (Figure 5-5), it is not surprising that the tumors they identify generally bear some relationship to these clinically relevant breast cancer stratifiers (Figure 5-13). Therefore, upon inclusion of these commonly used clinical parameters, some of the modules lost their prognostic value in multivariate analyses and do not have added benefit to the prediction of patient outcome (Figure 5-17). However, regardless of their prognostic value in multivariate analyses, these modules help explain the underlying biology behind these clinical designations. Importantly, some fetal gene subsets still have prognostic value beyond these commonly used clinical metrics (Figure 5-17). Of note, although enrichment for sub-signature fMRUiii correlates with high grade Her2+ and basal-like tumors, it also correlates with patients having a higher probability of survival in the compendium analyzed. The genes comprising these enriched subsets represent diverse biological processes previously implicated in cancer, including immune response, cell survival and wounding (Chang et al., 2004; Perou et al., 2000a), but also other processes including embryonic morphogenesis and adhesion (Figure 5-16). Many existing signatures have significant contribution from genes correlated to ER expression and proliferation (Fan et al., 2006; Wirapati et al., 2008). Our

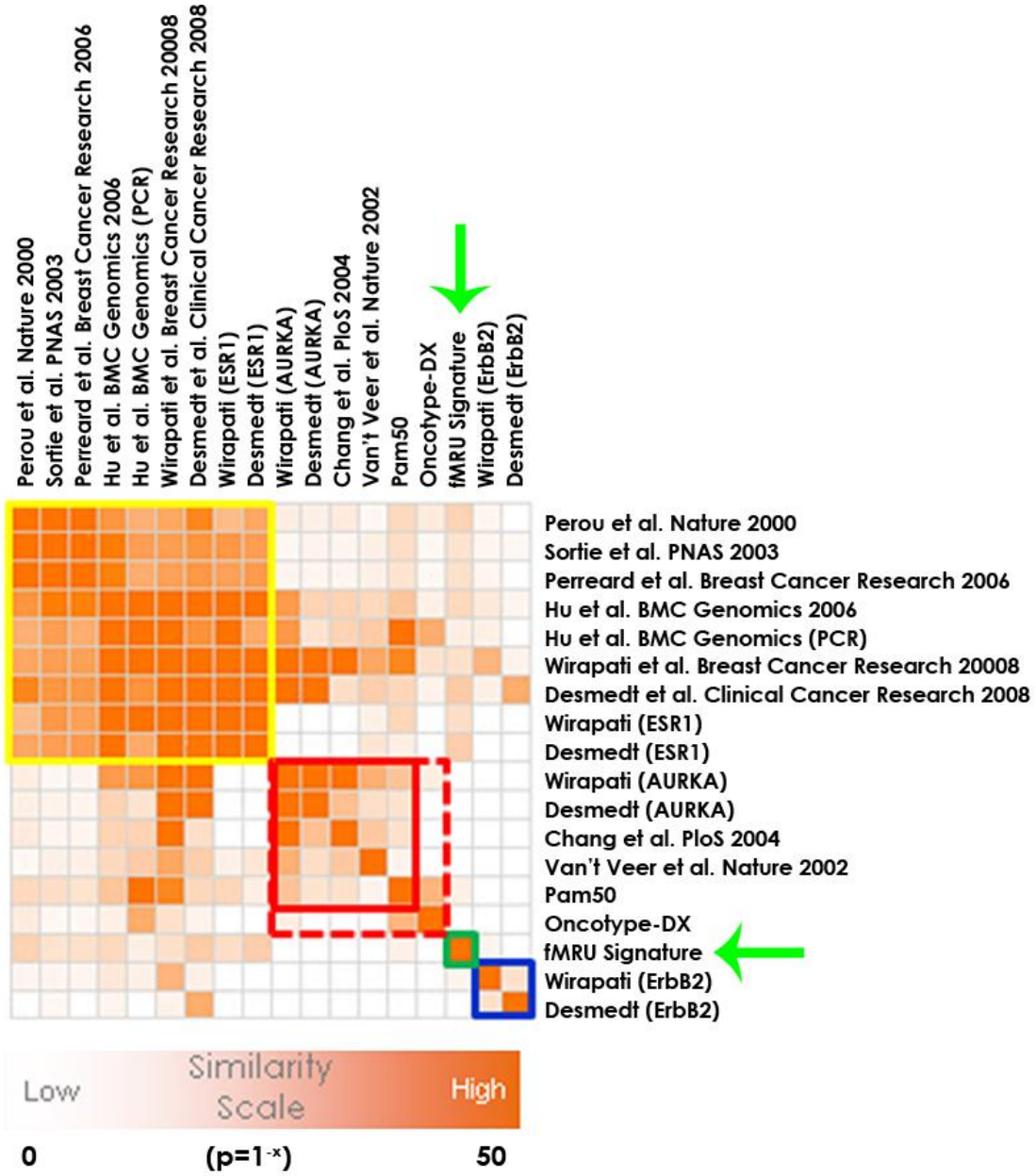


Figure 5-13. A comparison of several signatures and clinical metrics by significance of gene overlap. Most signatures are closely related and are significantly associated with ER (yellow box) or Proliferation (AURKA; red box) related signatures. The small OncotypeDX signature has marginal correlation with the proliferation group (dotted red line). The fMRU signature (green box and arrows) is relatively unique showing no significant overlap with proliferation or ErbB2 related signatures (blue) and relatively low association with ER related signatures.

Figure 5-14. Hierarchical clustering of the 96 most highly variable fMRU genes (upper heat map) shows fetal gene signature expression to be modular in human breast cancers. Intrinsic subtype designation and overall fMRU enrichment are indicated by colored vertical bars. The lower, fSTR heat map is ordered according to hierarchical clustering of samples in the upper fMRU heat map.

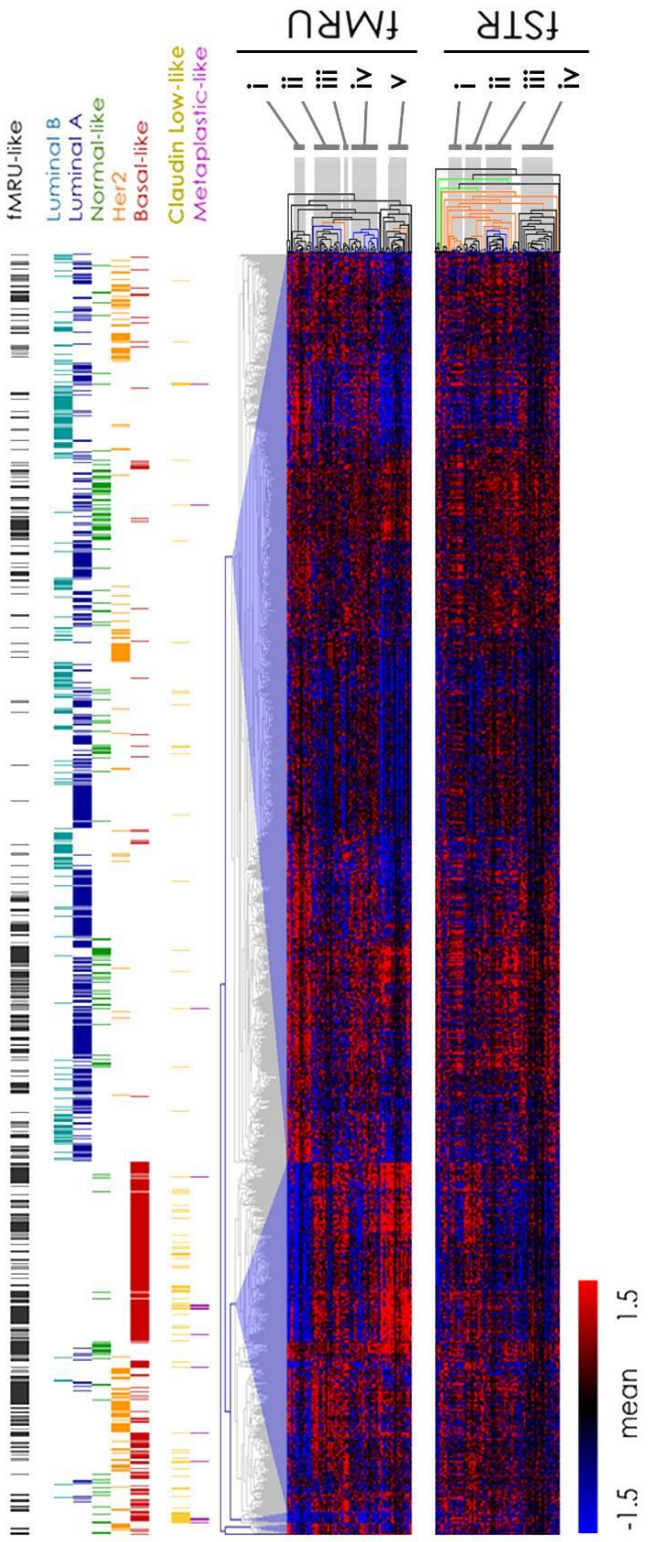

fMRU-like
Luminal B
Luminal A
Normal-like
Her2
Basal-like
Claudin Low-like
Metaplastic-like
fMRU
i
ii
iii
iv
v
fSTR
i
ii
iii
iv
-1.5
mean
1.5

Figure 5-15. Detail of hierarchical gene cluster analysis of the fMRU and fSTR signature genes. Hierarchical gene clusters from the 96 most variably expressed fMRU and fSTR signature genes were analyzed by Leave-One-Out (jackknifing) permutation analysis (100 permutations) to indicate relative stability of coordinate expression. The gene clusters analyzed in Figure 5 in the manuscript are indicated in black boxes and the Entrez ID's for genes corresponding to each branch in the dendrogram are indicated.

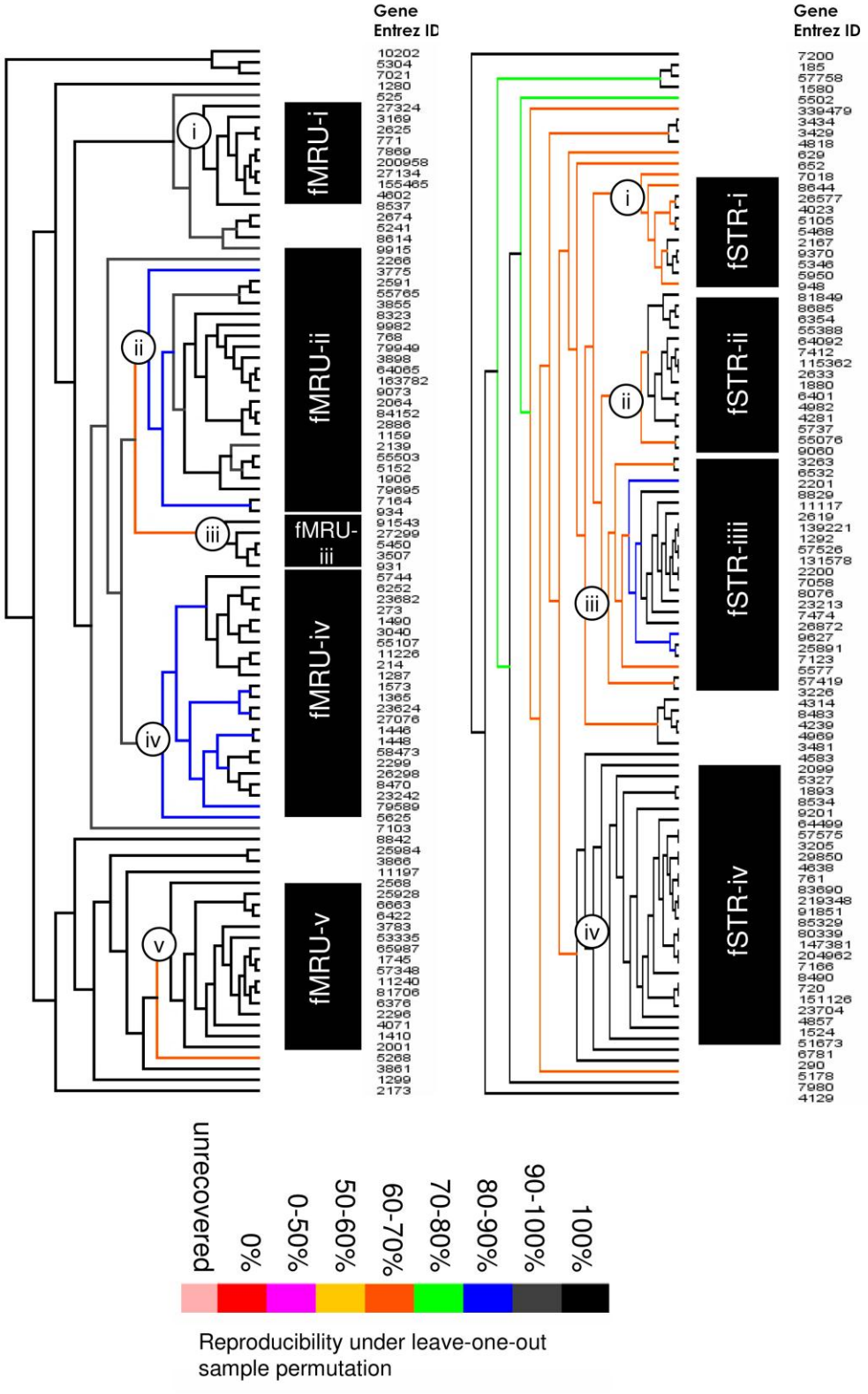
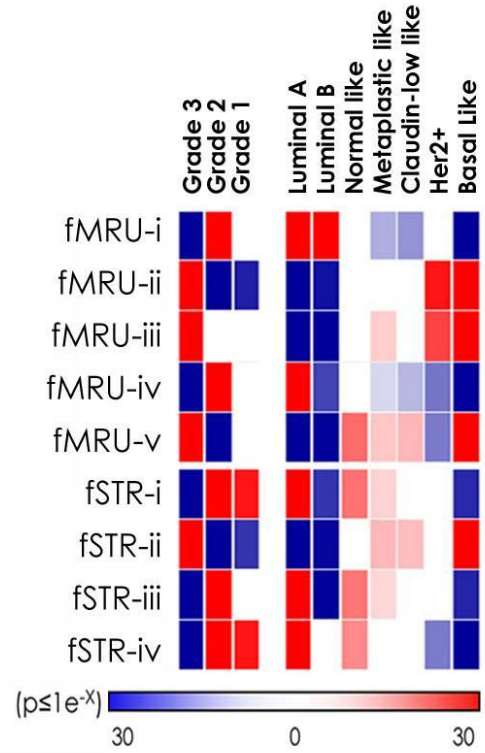

Gene Entrez ID
fMRU-i
fMRU-ii
fMRU-iii
fMRU-iv
fMRU-v
i
ii
iii
iv
v
Gene Entrez ID
fSTR-i
fSTR-ii
fSTR-iiii
fSTR-iv
unrecovered
0%
0-50%
50-60%
60-70%
70-80%
80-90%
90-100%
100%
Reproducibility under leave-one-out sample permutation

Figure 5-16. Significance of enrichment for sub-signatures among diverse breast cancers in a large microarray compendium (Ben-Porath et al., 2008).Enrichments according to subtype and grade are indicated by colored squares that represent probabilities for the percentage of tumors found to be enriched or repressed in each annotation group. Genes comprising each sub-signature are listed. Biological functions associated gene constituents of the sub-signatures (gene set enrichment $p<0.05$).



Tox3, FOXA1, GATA3, CA12, SEMA3B,
MUC20, TJP3, AGR3, MYB, BCAS1

KCNK1, GALNT3, C1orf106, Krt7, FZD6, FGFBP1, CA9, C10orf81, LAD1, PERP, ANKRD38, CLDN8,
ERBB2, PPP1R1B, GRB7, CKMT1B, EYA2, TRPV6, PDE9A, EDN1, GALNT12, TPD52L1, CD24

RSAD2, ADAMDEC1, POU2AF1, IGHM, MS4A1

PTHLH, RTN1, RAB38, AMPH, CTGF, HBA2, TMEM16A, GALNT6, ALCAM, COL4A5, CYP2J2, CLDN3,
CBLC, LYPD3, CSN1S1, CSN3, PLEKHB1, FOXI1, EHF, SORBS2, COBL, RNF128, PRODH

GABRP, SOSTDC1, SOX10, SFRP1, KCNN4, BCL11A, KCTD14, DLX1, TTYH1,
PADI2, PPP1R14C, CX3CL1, FOXC1, TM4SF1, CRYAB, ELF5, SERPINB5

TF, AKR1C3, PCOLECE2, LPL, PCK1, PPARG, FABP4, ADIPOQ, PLIN, RBP4, CD36

ST6GALNAC5, MARCO, CCL7, MCM10, SAMSN1, VCAM1, GBP5, GBP1,
EBI2, SELE, TNFRSF11B, MID1, PTGFR, TMEM45A, PAPSS2

HPX, SLC6A4, FBN2, NRP1, EMELIN1, GAS1, COL6A2, PCDH19, LRRC15, FBN1, THBS2,
MFAP5, SULF1, WNT5A, STEAP1, SNCAIP, DKFZP586H2123, CLEC3B, PRKAR2B, SLC24A3, HOXC10
PLAT, ECM1, CHST1, DCLK1, TPSB2, PCDH10, HOXA9, TRPM5, MULK, CA3, CRISPLD1, PLAC9, CHRDL1,
LGALS12, PNPLA3, CBLN2, SLC44A5, TPH1, RGS5, C4A, ZNF533, KCNE4, NOVA1, CX3CR1, CGI-38

fMRU i transcriptional regulation
ii embryonic morphogenesis
iii immune response
iv cell adhesion
v embryonic development/ anti-apoptosis

fSTR i glucose/lipid/carbohydrate metabolism
adipose/PPARγ signaling
ii immune response/wounding
iii embryopnic morphogenesis/ extracellular matrix
iv wound response

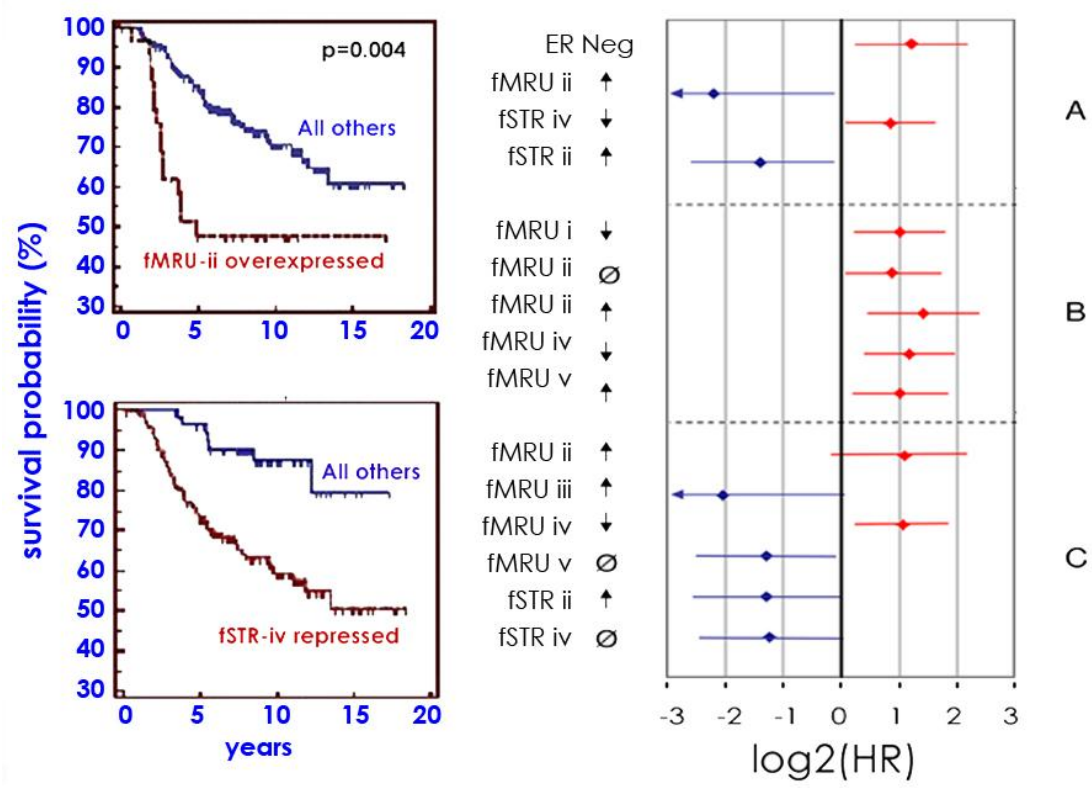


Figure 5-17. Left. Kaplan-Meier survival probability calculated for patients exhibiting fMRU-ii enrichment and fSTR-iv repression from the NKI-295 cohort (van de Vijver et al., 2002a). For multi-testing correction $P_{value} = P_{shown} \times 9_{gene\ sets}$. **Right.** Sub-signatures showing significance in multivariate analysis ($p<0.1$) are graphed for models including the following categorical clinical variables: A=ER status, Grade, Lymph-node status, Tumor size; B=Grade, Lymph-node status and Tumor size; C=ER status, Lymph-node status and size. A positive (or red) value indicates a poorer prognosis, while a negative (or blue) value indicates a better prognosis. The associated hazard for ER-negative tumors is shown in model A for comparison. ↑=sub-signature enrichment, ↓=sub-signature repression, Ø=no significant signature enrichment and/or depletion.

signatures are distinct from others with regard to gene content and the biology they encompass (Figure 5-3, 5-13, and 5-16). The genes within the fetal modules represent features that are likely masked by the signals from proliferation and other prominent cancer phenotypes when the deranged biology of the tumor is the starting point for signature identification. Furthermore, approximately 60% of the genes comprising these fetal gene modules are specifically up-regulated in the fMRU- relative to aMRU-enriched populations. These genes would not previously have been detected in expression profiles from homeostatic adult mammary and represent novel candidates for therapeutic and prognostic strategies that would have been overlooked in previous studies.

Discussion

As stated earlier, breast cancers are a very heterogeneous disease between individual patients, but also within the same tumor. The signatures that are acquired from breast tumors represent an average of the particular region that was selected for analysis. This amalgamation of information is useful for gross observations, but if for example the contribution of a rare cancer stem cell population is to be evaluated in this population, whether this signature will be detectable among the various other cell type contributions is unclear. Furthermore, all primary tissue stem cell populations isolated to date are heterogeneous, suffering from contributions of other cell types to the

signature. This leaves the question of what purity is sufficient to begin making biologic and molecular conclusions regarding your population of interest. The ultimate and most clinically relevant test is to determine if your population associated signature adds any benefit to existing patient stratification strategies. The detection of dangerous cell types will most likely rely on the current trajectory towards small population or even single cell transcriptomic analyses. However, in the interim, by creating signatures associated with a functional readout or phenotype, it may be possible to glean relevant information. For instance, if there is a discernable difference between a population harboring stem cell activity and the remaining cell types, it is possible to create a signature associated with this activity with the assumption that non-relevant genes have been excluded by this type of comparative analysis.

The fMRU-enriched population has many important distinctions from their adult counterparts. The fMRU-enriched population exhibits higher ErbB2 expression, and includes individual cells co-expressing luminal, myoepithelial and mesenchymal markers. The sensitivity of the fMRU-enriched population to ErbB kinase inhibition *in vitro* contrasts with the insensitivity reported for the aMRU-enriched population *in vivo* (Asselin-Labat et al., 2010). Interestingly, human breast cancers with un-amplified, but often measurable, ErbB2 expression are clinically designated Her2 negative' (ASCO/Herceptest (Jacobs et al., 1999)), though they could conceivably also rely upon ErbB signaling in a

manner similar to the fMRUs. Indeed, recent and ongoing clinical trials have indicated that some verified Her2 negative patients benefit from ErbB2 targeted therapies (NSABP 31, NSABP 47). Given the unique properties of the fMRU-enriched population, it follows that many novel genes and programs were identified in this population. These novel programs offer prognostic value in multivariate analysis and furthermore, offer new targets for therapeutic approaches. In light of our and others' findings, these results suggest a change in how treatment strategies are allocated to Her2 negative breast cancer patients.

Cancers resemble organogenesis gone awry and result from disruptions in normal tissue homeostasis and development. However, to date, most prognostic signatures derive from retrospective analyses of breast cancer microarray datasets rather than from developmental contexts. Although such approaches have proven generally useful for predicting patient outcome, their reliance on factors such as ER status or proliferation may not sufficiently access the biological context that could provide further insight into disease pathogenesis or enable identification of new prognostic metrics or therapeutic targets. Consistent with this expectation, we identify fetal mammary genes correlating with the elaboration of tissue-specific stem cells during development that present additional molecular links to tumors arising in the same tissue. We expect that refinement of the fMRU and fSTR gene signatures, and further functional analyses of individual cells within the fMRU-

enriched population, will provide markers useful for prospective identification of these cells during normal development, and for more precise identification of stem-like cells and expression signatures within breast cancers.

Links between embryogenesis and tumorigenesis range from Conheim's 1867 "embryonal rest" model proposing that undifferentiated embryonic cells persisting into adulthood are the origin of cancer (Brewer et al., 2009) to mid-20th century identification of fetal proteins expressed in diverse tumors (i.e. onco-fetal proteins (Brewer et al., 2009)). Since then, specific embryonic proteins, splice isoforms, miRNA and expressed genes supporting such links have been reported (Ben-Porath et al., 2008; Christofk et al., 2008; Powers and Mu, 2008). However, the cells and gene networks responsible for the proposed linkages have remained obscure. Here, we overcome the technical challenges associated with obtaining and profiling the relevant cell populations that have contributed to this gap in our understanding. Our findings advance the model (Brewer et al., 2009) that pathways governing fetal tissue stem cells and normal development are indeed resurrected by cells that fuel tumorigenesis in the adult.

Acknowledgements

This chapter, in full, has been submitted for publication as it may appear in Cancer Cell or Cell Stem Cell, 2011, Lin, Jennifer*; Engle, Dannielle*; Spike, Benjamin*; Cheung, Samantha; La, Justin; Wahl, Geoff. *

indicates equal contribution. The dissertation author was one of the primary authors of this paper.

Chapter 6. Prospective tracking of mammary stem cells *in vivo*

Background

Wnt signaling is essential for mammary gland development. The expression of a Wnt inhibitor, Dickkopf 1 (Dkk1), prevents the formation of all mammary buds (Chu et al., 2004). Dickkopf acts by binding the Wnt co-receptor, LRP, preventing it from co-localizing with the Wnt receptor, Frizzled, upon Wnt binding. This leads to a general antagonism of canonical Wnt signaling. In other studies, knock out of the Wnt transcription factor, Lef1, leads to the transient formation of only bud four while the rest of the buds are absent (van Genderen et al., 1994). Wnt signaling has been demonstrated to also function in stem cell self-renewal of both embryonic stem cells and tissue stem cells (Sato et al., 2004). Addition of Wnt ligands to mammosphere culture media extended the ability of these spheres to be serially passaged as well as transplanted *in vivo* (Zeng and Nusse, 2010).

Many different genetic Wnt reporters have been created, however, consistency between them is generally lacking. These Wnt reporters are largely based on a system originally designed in the laboratory of Dr. Hans Clevers called TOPFLASH/TOPCAT (Korinek et al., 1997). The original TOP reporter system contains three optimized TCF binding motifs. These TCF sites drive expression from a minimal promoter upon transfection into cell lines. In the case of TOPFLASH, the *c-Fos* minimal promoter will drive expression of

luciferase upon Wnt activation of the TOP promoter (Korinek et al., 1997). The alternative system uses the herpes virus thymidine kinase minimal promoter to drive expression of chloramphenical acetyltransferase (TOPCAT) (Korinek et al., 1997). DasGupta and Fuchs applied this system to make a Wnt reporter mouse, TOPGAL. This transgenic line utilizes the same three TCF binding sites upstream of the *c-Fos* minimal promoter and *LacZ* (Figure 6-1, (DasGupta and Fuchs, 1999)). Independently, the laboratory of Dr. Alman developed another TOPGAL reporter line using the same three TCF binding motifs driving expression of *LacZ* from the *c-Fos* minimal promoter (Cheon et al., 2002). However, although both of these transgenic lines label the mammary buds during development, the Fuchs TOPGAL line (TOPGAL-F) labels the epithelium while the Alman TOPGAL line (TOPGAL-A) labels the mesenchyme (Boras-Granic et al., 2006; Chu et al., 2004). Another Wnt reporter line, BAT-gal, contains seven Tcf binding sites that utilize a different TCF binding sequence (Maretto et al., 2003). This line uses the minimal promoter of *siamois*, a Wnt responsive gene present in *Xenopus*, but not mammals (Maretto et al., 2003). This line is reportedly more sensitive; however, the expression of this reporter is restricted to a small population of epithelial cells within the mammary bud unlike the TOPGAL-F line (Lindvall et al., 2006). These discrepancies are most likely due to copy number and position effect variegation between transgenic lines, as each independent line has a different insertion site and multimerization level. The Behrens laboratory

created another Wnt reporter line in which they homologously recombined a *LacZ* transgene into the second exon of *Axin2* (Lustig et al., 2002). However, evaluation of *LacZ* expression has not yet been pursued in the embryonic mammary bud. The identification of Lgr5 as a Wnt target gene in colon cancer cell lines and its expression in the crypt based columnar cells of the intestine led to the creation of a line with eGFP and CRE^{ER} inserted into this locus to create another endogenous Wnt target gene reporter line (Barker et al., 2007). The utility of this line for identifying intestinal and hair follicular stem cells was demonstrated, however its expression pattern in the mammary gland remains unknown (Barker et al., 2007; Jaks et al., 2008).

Previous attempts at identifying tissue stem cells have utilized the putative stem cell property of infrequent cell division, and several tissue stem cell populations have been reported to be quiescent. The reporter component H2BGFP can be used to preferentially label infrequently dividing cells. Originally designed and generated by Dr. Tehru Kanda, this fusion between the histone, H2B, and EGFP, and was shown to stably and evenly incorporate into chromosomes without disturbing the mitotic index or growth rate of HeLa cells (Kanda et al., 1998). To evaluate the ability of this label to track cell divisions, Dr. EeTsin Wong in our laboratory used CHO and HeLa cells with single copy insertions of TRE-H2BGFP under the control of the Tet-On system (Wong et al., 2005). The Tet-On system utilizes rtTA, which is a tetracycline transactivator dependent on the addition of doxycycline (Gossen et al., 1995).

By maximally labeling the nuclei of synchronized cells with H2BGFP, washing out doxycycline, then monitoring fluorescence with each population doubling, Dr. Wong established the cell division dependent dilution of the H2BGFP signal ((Wong et al., 2005). Inhibition of new expression diluted H2BGFP fluorescence by half with each cell division such that non-dividing cells retain a bright fluorescent signal (Wong et al., 2005). The retention of H2BGFP fluorescence as an indication of infrequent cycling has been previously applied to the identification of stem cells in the hair follicle, which are postulated to enter the cell cycle rarely (Tumbar et al., 2004). It was found that H2BGFP positive cells resided in the bulge, while the H2BGFP signal was diluted to an undetectable level in surrounding progenitors and differentiated cells (Tumbar et al., 2004). Upon wounding, division of these cells was visualized and migration of a cell out of the bulge to the site of injury was observed, during which fluorescence was gradually lost upon proliferation (Tumbar et al., 2004).

Results

Establishing a genetic system in mice to identify Wnt responsive, label retaining cells *in vivo*.

This system consists of activator and reporter components (Figure 6-1). The expression of the activator component is dependent on Wnt signaling, which is necessary for mammary development. In addition, this system is regulated by doxycycline to enable the system to be shut off for pulse-chase

analysis. Using this Wnt dependent pulse-chase analysis, we will determine whether Wnt signaling and infrequent entry into the cell cycle detected by retention of H2BGFP labeling are properties of MaSCs. Finally, to avoid the copy number and position effect variegation of traditional transgenesis, site-specific recombination into a pre-determined open locus, Rosa26, will be utilized for generation of the transgenic reporter line.

Wnt signaling will be used to functionally label the mammary buds in early embryonic development. SuperTOP, the Wnt responsive promoter that has been utilized in this system, contains seven TCF binding motifs driving expression from a herpes virus thymidine kinase minimal promoter. By using this promoter to drive expression of a tetracycline transactivator, tTA, its expression will be limited to Wnt responsive tissues. This promoter has been shown to achieve hundred-fold activation at maximal Wnt induction in HEK cells (Dr. Randy Moon, personal communication). Although this version of the TOP promoter has not been previously characterized in mice, several of its relatives have been shown to give reliable canonical Wnt pathway readout. The same TCF binding motifs were used in the TOPGAL-F and TOPGAL-A mice, while the TK minimal promoter has been used in TOP-EGFP and TOP-GAL mice (Moriyama et al., 2007). To characterize the different possible Wnt reporter expression patterns, we characterized two different reporters that were knocked into the endogenous locus of the Wnt target genes, Axin2 and Lgr5 (Figure 6-2 and 6-3) (Barker et al., 2007; Lustig et al., 2002). The Axin2-

lacZ reporter line labels Wnt responsive tissues, including the embryonic mammary gland, developing hair follicles, and vibrissae. However, expression of this reporter was not observed in the E18.5 mammary rudiments. Lgr5-eGFP mice exhibited expression in the mammary rudiments as well, but its expression extended to the luminal compartment of the E18.5 mammary rudiment. Interestingly, expression of Lgr5-eGFP was observed in the vibrissae, but not the hair follicles. These two lines will be useful in validating the Wnt dependent expression pattern of the genetic system once it comes online.

The Tet-Off component enables this system to be toggled off by the addition of the antibiotic, doxycycline. Under doxycycline-free conditions, tTA binds to tetracycline response elements in the TRE promoter, activating expression of the reporter. Doxycycline binds to the tTA and prevents it from binding the TRE promoter, ablating further expression of the reporter. tTA has been used in multiple transgenic lines, however it has recently been shown that this transactivator contain portions of VP16 that are toxic at high levels (Baron et al., 1997). To circumvent this toxicity, our system utilizes tTA2. This version of the transactivator has the same activation level as the original tTA, but is tolerated at three fold higher concentrations (Baron et al., 1997). We expect that the reduced tTA toxicity will prevent any VP16 related lethality in the putative mammary stem cells. In addition, we obtained a transgenic line carrying tTA driven by the cytokeratin 14 promoter to demonstrate the

Figure 6-1. Schematic representation of the different components of the genetic reporter system. The Wnt signaling panel demonstrates the localization of Wnt signaling to the mammary rudiments during development. The Tet-Off System panel illustrates how the label retention component will operate. The Modular Activator/Reporter Cassette panel shows the different components included in this system, which will be integrated into a universal acceptor site homologously recombined into the Rosa26 locus (bottom panel).

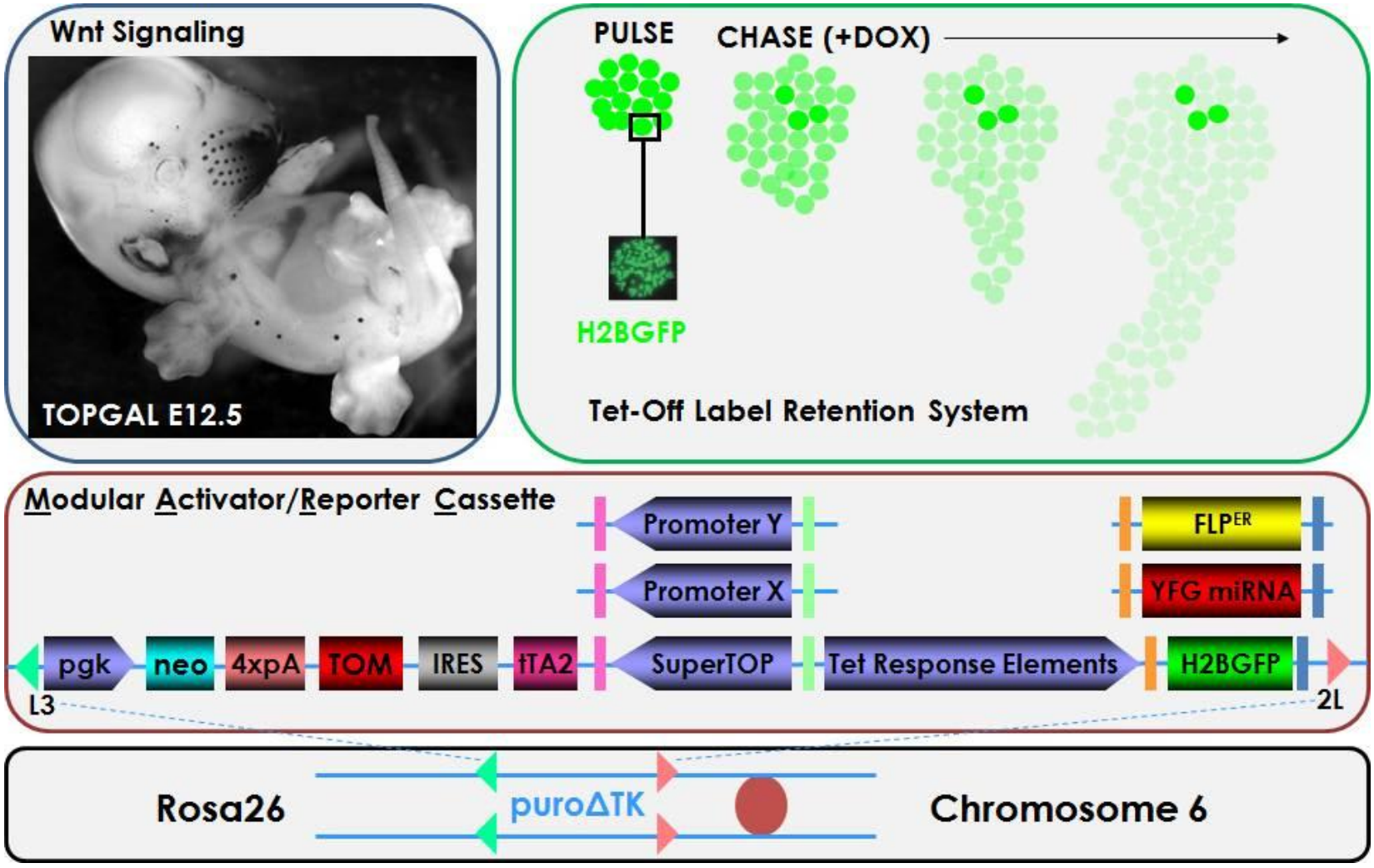

Wnt Signaling
TOPGAL E12.5
PULSE
CHASE (+DOX)
H2BGFP
Tet-Off Label Retention System
Modular Activator/Reporter Cassette
Promoter Y
Promoter X
FLPER
YFG miRNA
pgk
neo
4xpA
TOM
IRES
tTA2
SuperTOP
Tet Response Elements
H2BGFP
L3
2L
Rosa26
puroΔTK
Chromosome 6

Figure 6 – 2. Characterization of Axin2-lacZ expression during embryonic mammogenesis. Embryos were evaluated as whole mounts by X-gal staining at E11.5, E12.5, E13.5, and E15.5, and at E13.5, sections of the embryonic mammary buds were characterized. E18.5 embryos did not exhibit any lacZ positivity in the embryonic mammary glands and therefore those images are not shown.

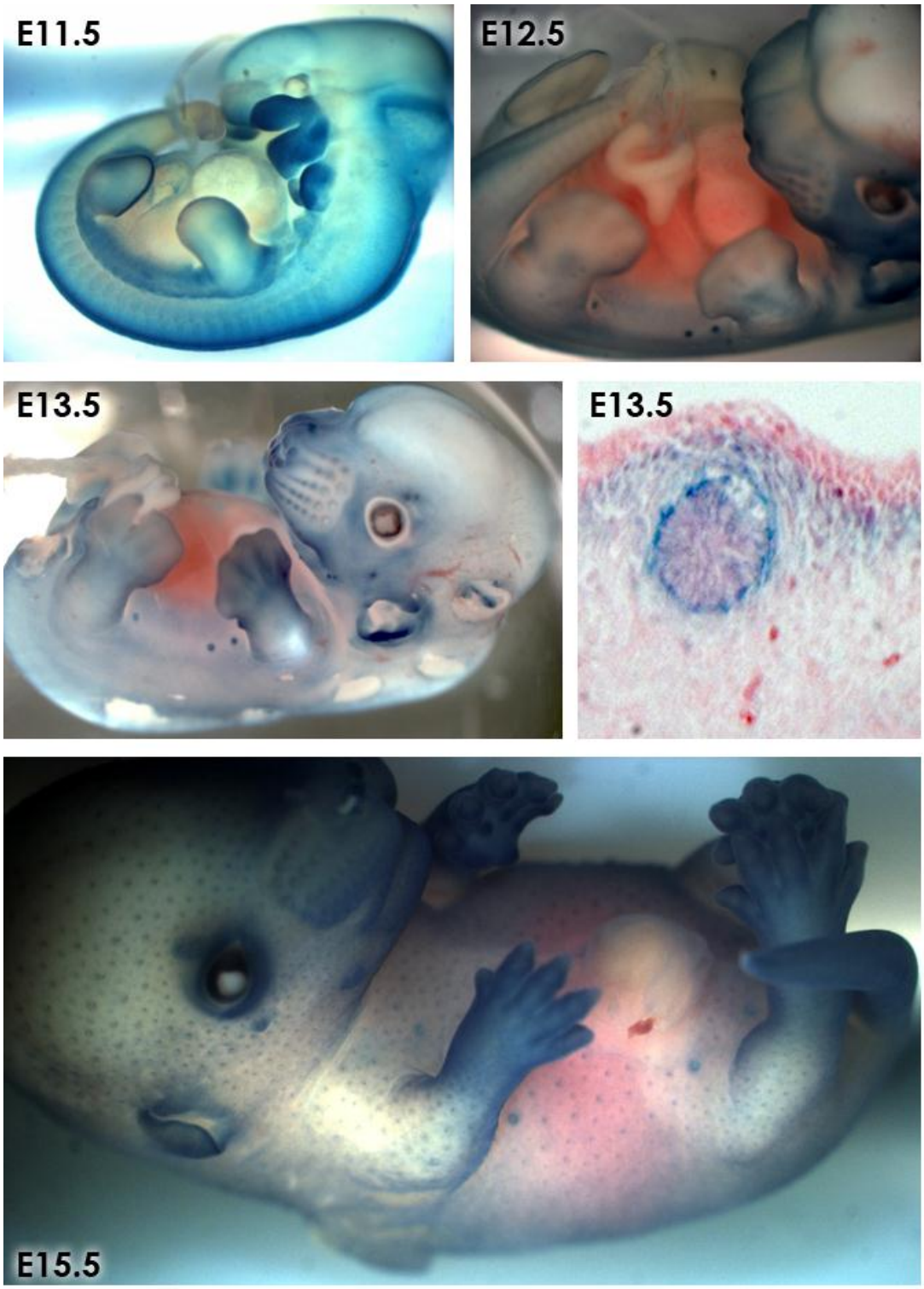
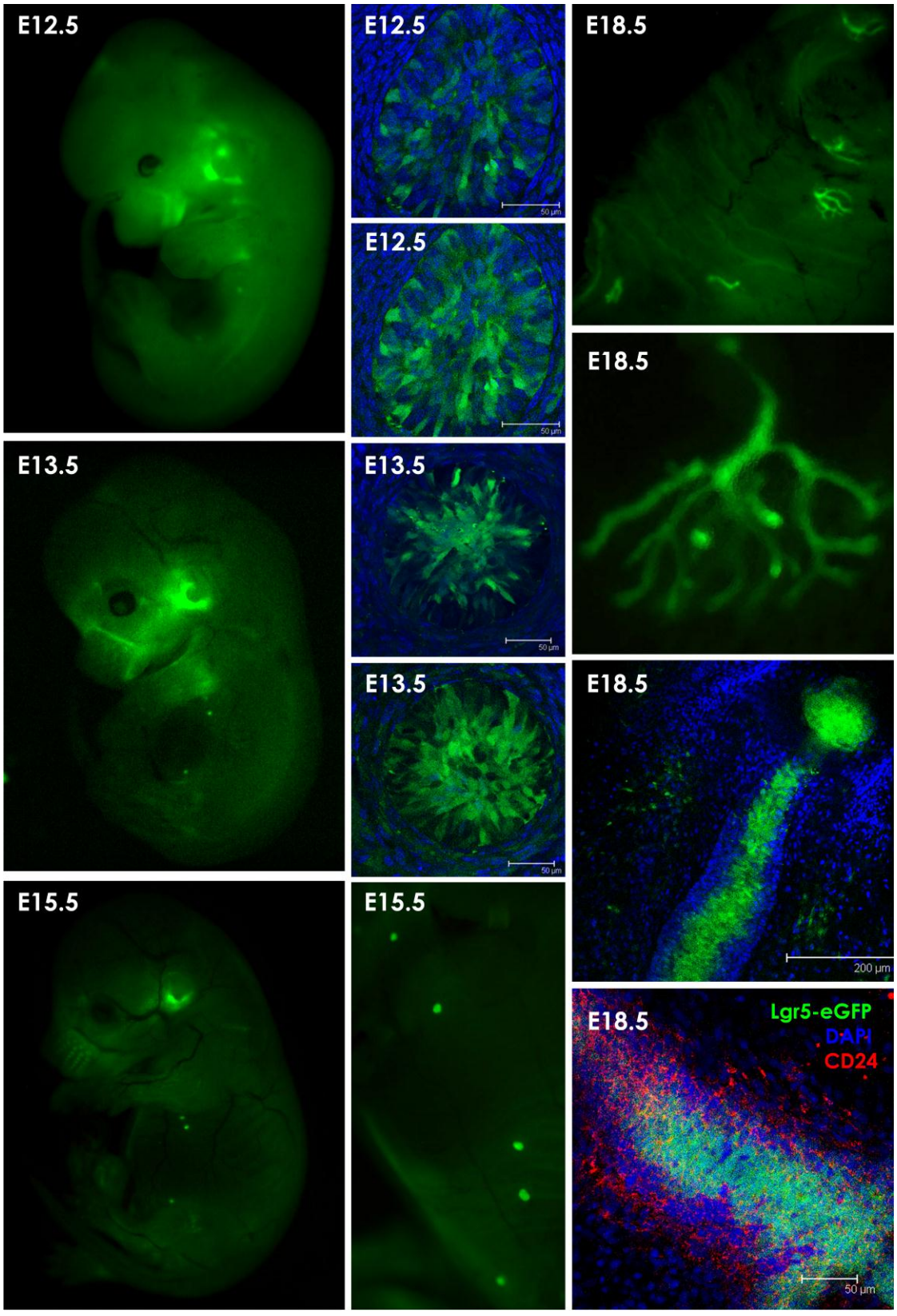

E11.5
E12.5
E13.5
E13.5
E15.5

Figure 6 – 3. Evaluation of Lgr5-eGFP during embryonic mammogenesis. Lgr5eGFP expression is present in mammary rudiments evaluated as early as E12.5. Embryos were evaluated by fluorescence microscopy before dissection for confocal microscopy and immunofluorescence where indicated.


E12.5
E12.5
E18.5
E12.5
E18.5
E13.5
E13.5
E13.5
E18.5
E15.5
E15.5
E18.5
Lgr5-eGFP
DAPI
CD24
50 µm
200 µm

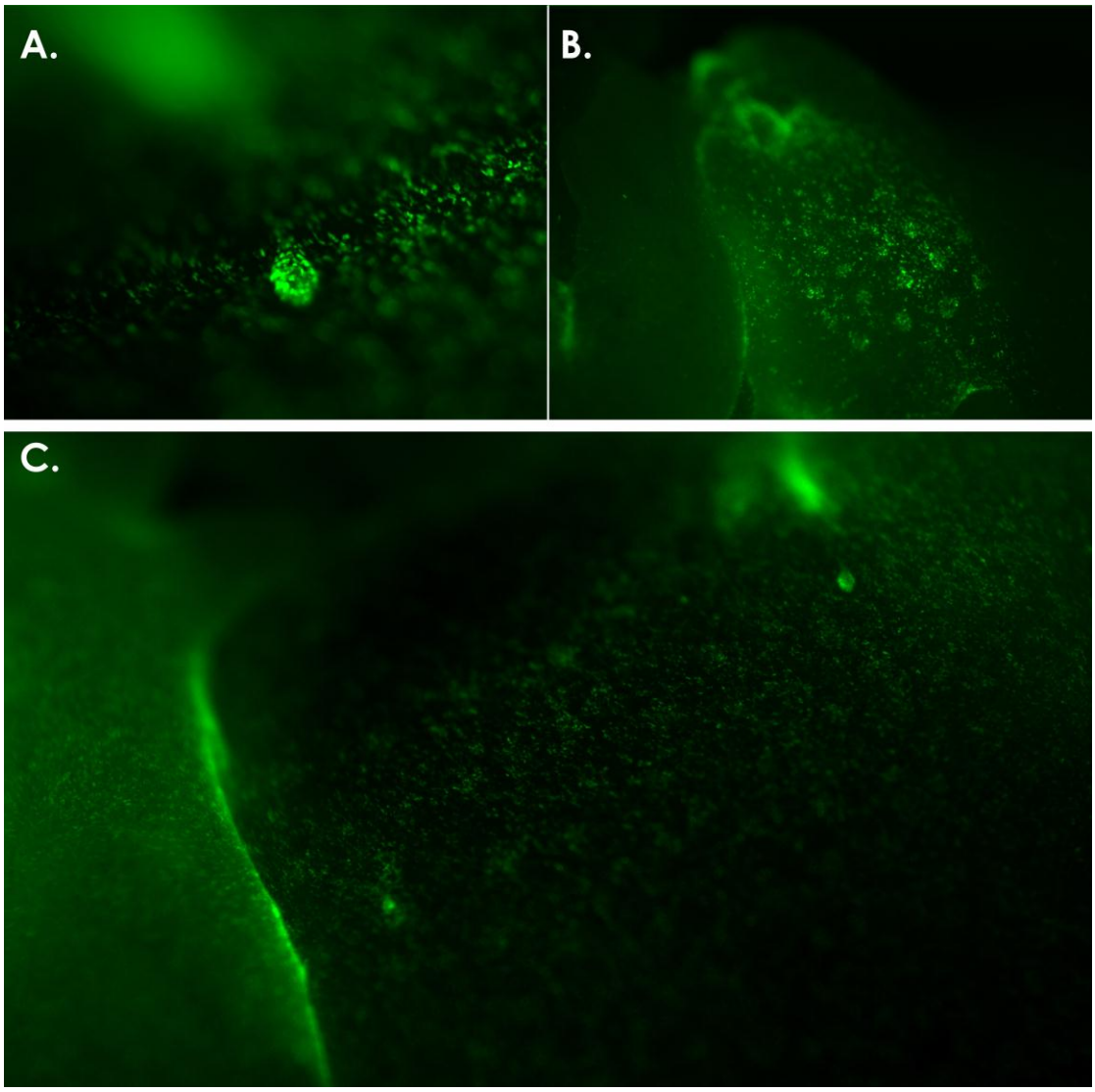


Figure 6-4. K14-tTA; TRE-H2BGFP Embryonic Day 14.5 (E14.5) embryonic tissues. **A.** Mammary bud number four. **B.** Vibrissae (primitive, developing whiskers. **C.** Mammary buds three and four in between the two developing limbs.

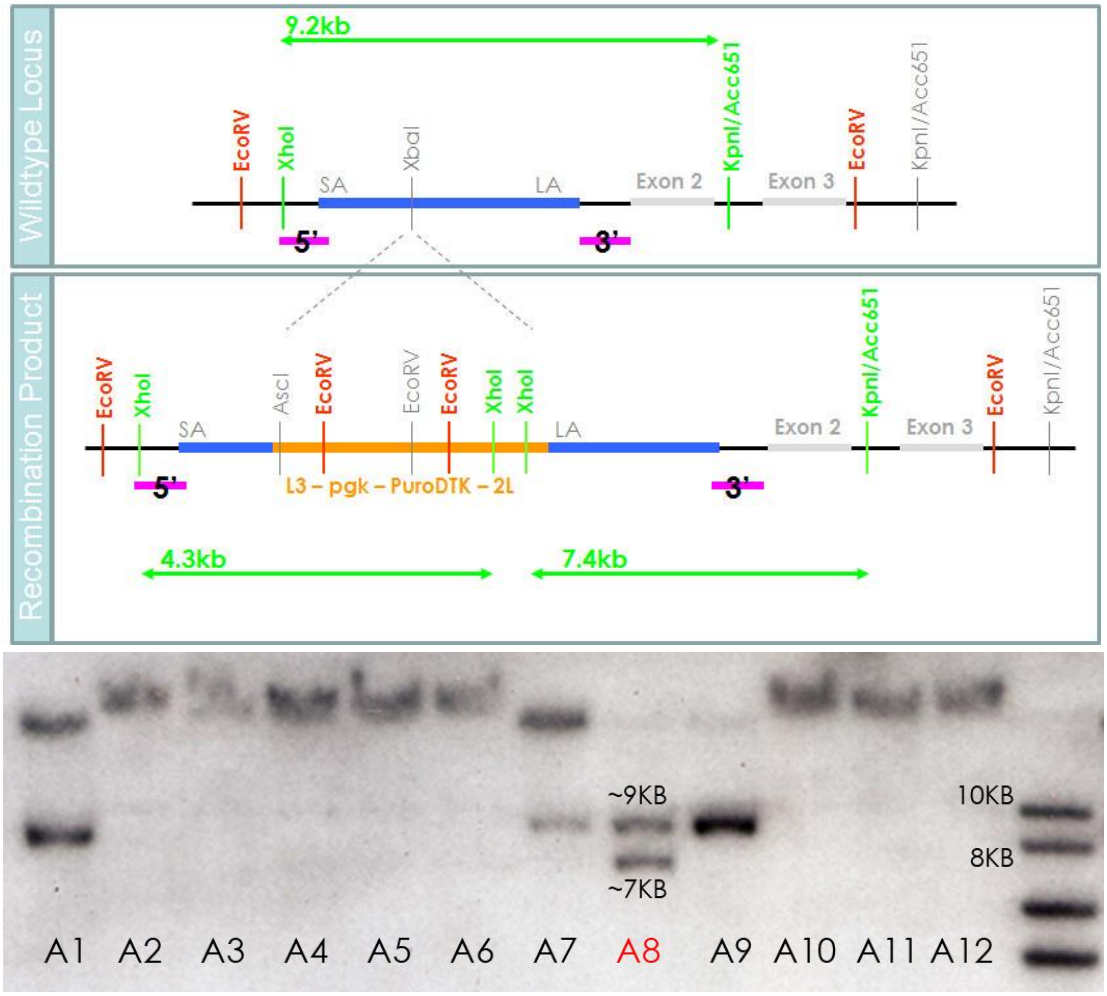


Figure 6-5. Top panel. Southern detection strategy. **Bottom panel.** Southern analysis of the three prime junction of the homologous recombination product. The positive clone, A8, is in red.

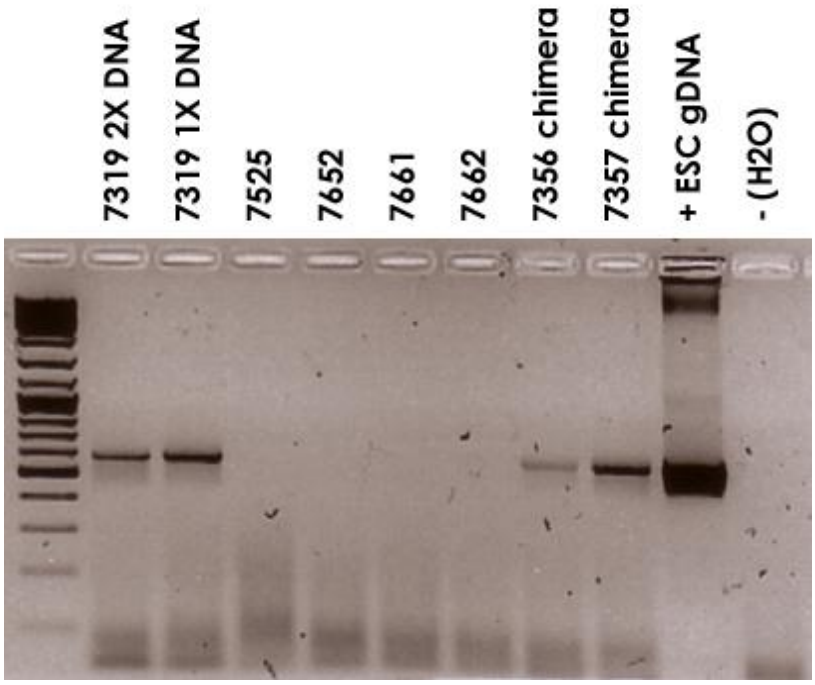


Figure 6-6. PCR genotyping of chimeras and their progeny and the detection of germline transmission (7319).

feasibility of our system and to examine any toxicity results from high expression of tTA and H2BGFP in the developing mammary gland (Dunbar et al., 2001). When crossed with a TRE-LacZ reporter line, this promoter results in labeling of the embryonic mammary bud epithelium and the myoepithelial cells of adult virgin mammary glands (Dunbar et al., 2001). Preliminary studies were conducted in our lab crossing the K14-tTA line to TRE-H2BGFP mice to obtain K14-tTA; TRE-H2BGFP embryos. This cross yielded H2BGFP labeling of the mammary bud epithelial cells in E14.5 embryos with no signs of toxicity from expression of tTA (Figure 6-4). These results demonstrate the feasibility of labeling the mammary bud using the Tet-Off system and H2BGFP.

To utilize this system for pulse chase analysis *in vivo*, doxycycline will be administered to mothers by adding 2mg/mL doxycycline to the drinking water. It has been shown that delivery at this concentration results in ablation of tTA transactivation within one week (Kistner et al., 1996). Furthermore, doxycycline induction of Tet-On expression occurs maximally by 24 hours after administration (Kistner et al., 1996). Therefore, tight regulation of this system is possible within our genetic system.

Upon completion, this system in its entirety will be SuperTOP-tTA2-IRES-Tomato; TRE-H2BGFP (Figure 6-1). This transgene will be integrated into the Rosa26 locus using Recombinase Mediated Cassette Exchange (RMCE). A positive and negative selection cassette containing a puromycin resistance gene and a ganciclovir/FIAU sensitivity gene (TK) flanked by two

different heterospecific loxP recombination sequences was homologously recombined into the Rosa26 locus in a strategy developed by Soriano (Soriano, 1999). ES cells in which the correct targeting was achieved were detected by Southern analysis of both five and three prime junctions (data not shown and Figure 6-5). ES cells from correct clones were injected into blastocysts. These blastocysts were implanted into the uteri of pseudo-pregnant females. Several chimeras were generated with contribution to coat color ranging from 60% to 100%. From these chimeras, founder lines were established upon detection of germline transmission (Figure 6-6). While this colony is expanding, the second round of targeting using these ES cells has proceeded to introduce the modular activator/reporter cassette (MARC) into this predefined locus. In addition, a previous targeting of the MARC transgene yielded positive clones that unfortunately did not contribute to the germline. However, it was possible to carry out preliminary characterization of the transgene to determine whether it was functioning as intended.

To evaluate the inducibility of MARC in ES cells, two clones were further characterized *in vitro*. Wnt signaling was induced in these clones by using a GSK3 inhibitor, Factor XV. The two ES cells clones, M2 and M16, were analyzed by qRT-PCR and confocal microscopy (Figure 6-7). The induction of Axin2, a Wnt target gene, was observed upon addition of both 50nM and 200nM of Factor XV, demonstrating that a Wnt response was being propagated. Accordingly, expression of Tomato and H2BGFP were observed,

but at higher levels of Factor XV, H2BGFP expression was reduced. This may relate to the toxicity associated with over stimulation of the Wnt signaling pathway and off target effects of Factor XV. Due to this depression of H2BGFP expression upon stimulation with 200nM of Factor XV, repression of H2BGFP expression by Doxycycline addition was not observed. H2BGFP was detectable by fluorescence, though dim, after 3 days of Wnt induction. By confocal analyses, Tomato was expressed in ES cells under all conditions while H2BGFP was not detected in No LIF and LIF controls. Undifferentiated ES cells maintain a basal level of Wnt signaling, which may explain the low-level expression of Tomato in the absence of Wnt induction. These data demonstrate that MARC operates as intended within the context of the two ES cell clones evaluated.

As the system utilizes the stem property of infrequent entry into the cell cycle, it was an initial concern that since the mammary bud is already postulated to undergo little proliferation, H2BGFP incorporation would proceed at different rates depending on the frequency of division and new histone incorporation. This would make evaluation of brightly labeled cells problematic as the stem cells may incorporate H2BGFP at a slower rate and therefore fluoresce at comparable levels as progenitors or differentiated cells. To evaluate this *in vitro,* a line of CHO cells containing a TRE-H2BGFP transgene adjacent to *DHFR* under control of the Tet-On system was used as a

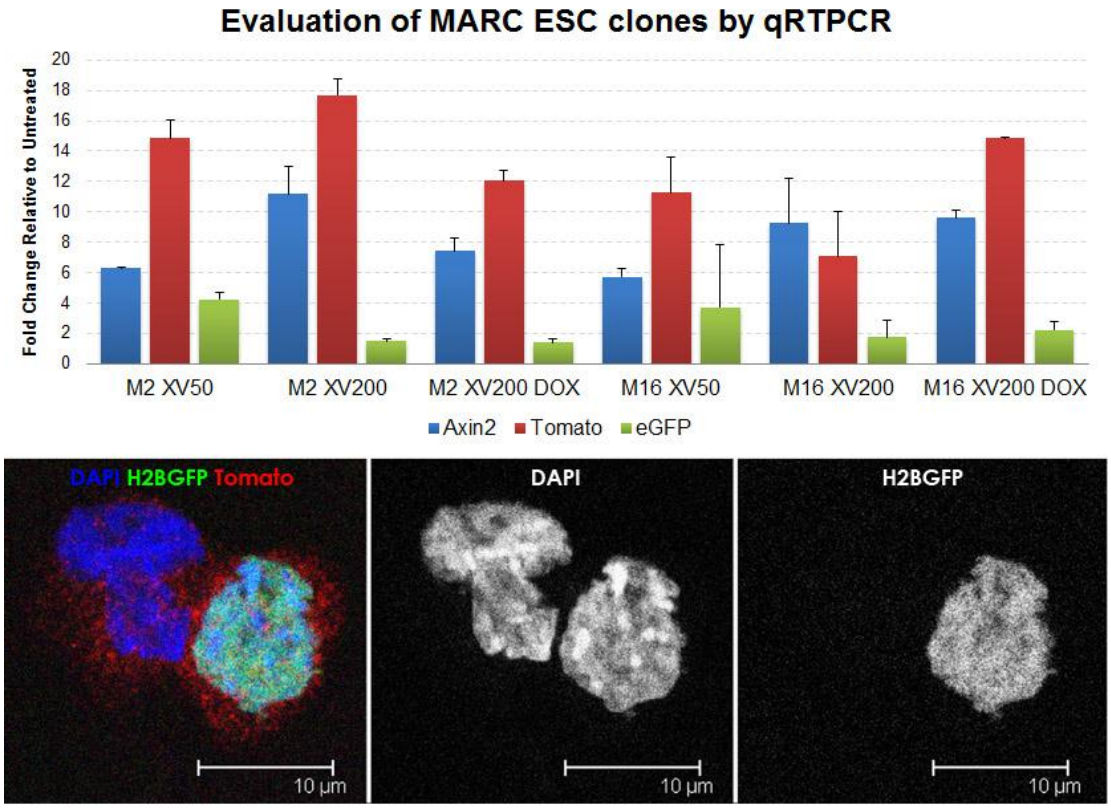


Figure 6-7. Top panel. Characterization of two MARC ES cell clones by qRT-PCR for the expression of Axin2, Tomato, and eGFP upon treatment with the Wnt activator, Factor XV (XV) at different concentrations in the presence and absence of Doxycycline. **Bottom panel.** Representative confocal image of a Wnt responsive ES cells expressing H2BGFP.

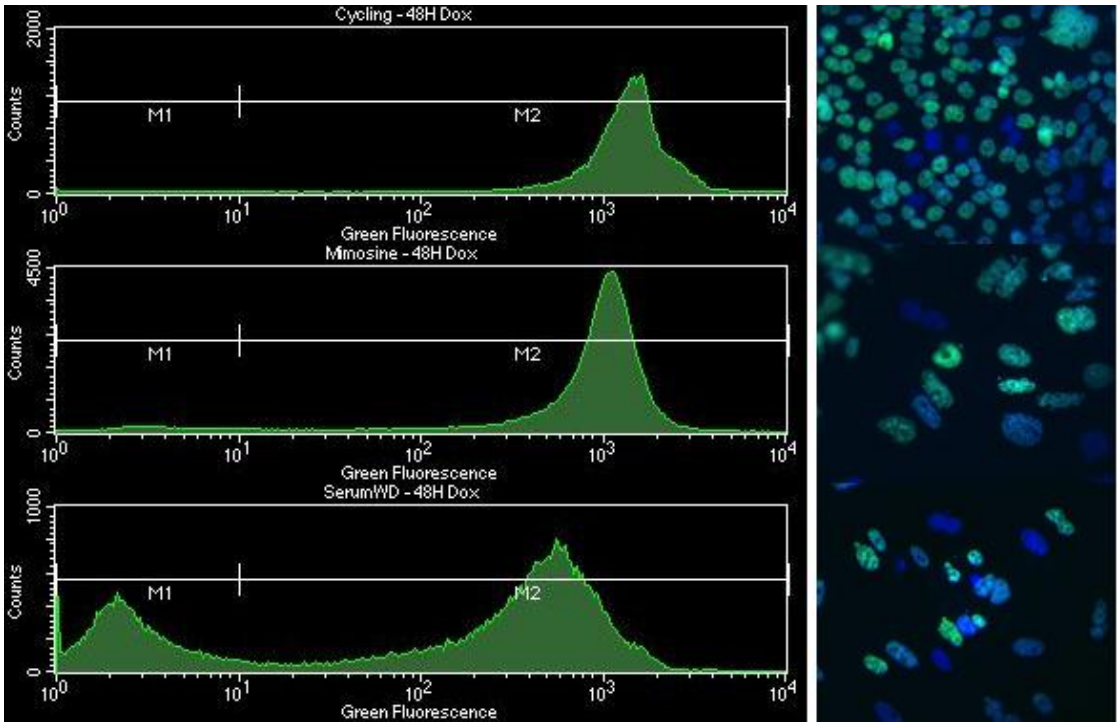


Figure 6-8.Left panels. Flow cytometric histograms depicting the number of cells exhibiting different ranges of fluorescence in cycling **(Top)**, Mimosine arrested **(Middle)**, and serum starved **(Bottom)** Chinese Hamster Ovary (CHO) cells. **Right panels.** Corresponding fluorescent microscopy images of H2BGFP expression in the nucleus of CHO cells.

model system (Wong et al., 2005). *DHFR* is translated into the enzyme, Dihydrofolate reductase, and is critical for DNA synthesis (Urlaub and Chasin, 1980). The CHO cell line that was used to generate the TRE-H2BGFP clone is deficient for this enzyme, carrying only one complete copy of this gene (Urlaub and Chasin, 1980). The TRE-H2BGFP was inserted adjacent to the remaining copy of the *DHFR* gene, discouraging its loss from the genome.

The CHO line expressing the TRE-H2BGFP transgene in the *DHFR* locus and the Tet-On system was arrested using either mimosine or serum starvation, inhibiting DNA replication by preventing DNA synthesis and elongation or activating the G1 restriction point, respectively (Gilbert et al., 1995). After level of arrest was confirmed by measuring the percentage of cycling cells incorporating BrdU (data not shown), H2BGFP expression was induced by the addition of doxycycline, and the fluorescence of cycling cells compared to that of arrested cells. When arrested with mimosine, these CHO cells had peak fluorescence very similar to that of cycling cells (Figure 6-8). The serum deprivation arrest lead to a reduced peak fluorescence with a larger population of non-fluorescent cells. This was most likely due to nutrient deprivation of the serum-starved cells as by the end of the assay, serum deprivation had continued for five days. Fluorescence microscopy also revealed that this fluorescence originated from the nucleus of these cells, and that the brightness level was visually comparable between arrested and

cycling cells (Figure 6-8). These results suggest that H2BGFP incorporates into both cycling and non-cycling cells at similar rates.

Discussion

Purification of hematopoietic stem cells to a 50% level of purity has taken two decades. The MaSC field has attempted to follow in the foot-steps of the blood system, however progress has been hampered and purity levels remain low. One of the issues that may play a role in the limited MaSC purity levels is the irrelevance of the enrichment markers for mammary gland development and their expression on multiple differentiated cell types. Identification of a marker that is functionally relevant for stem cell activity instead of solely correlative will greatly advance the MaSC field. Unfortunately, definitive MaSC characteristics are difficult to discern from the adult MaSC-enriched populations due to their high level of stromal and differentiated myoepithelial cell contribution.

Given the relevance of Wnt signaling for both mammary gland development and stem cell self-renewal, this pathway may be implicated in MaSC identity. In order to ascertain the functional contribution of Wnt signaling to MaSC activity, we created a genetic system to report canonical Wnt signaling activation. This reporter was designed to be a ubiquitous Wnt reporter that is more encompassing than evaluating individual Wnt target genes with different spatiotemporal expression patterns. Using this genetic

reporter system, we will be able to prospectively isolate Wnt responding cells and determine whether Wnt signaling correlates with MaSC activity. However, given the wealth of information revealed by the whole genome, microarray analyses carried out in previous chapters, it is likely that additional potential properties of MaSCs will require further evaluation. To facilitate evaluation of these putative characteristics, unique restriction sites have been included in the design of this genetic system to enable rapid transgenesis and functional characterization of mammary cells exhibiting different genetic reporter phenotypes.

Materials and Methods

Mice and embryos. Actin-eGFP mice (with C57BL/6 and CD-1 mixed background) were bred and maintained in the Salk Institute animal facility according to institutional guidelines. CD-1 nulliparous and pregnant mice and CB17-SCID recipient female mice were purchased from Charles River. Timed pregnancies were setup by crossing either homozygous or heterozygous adult actin-eGFP mice with adult CD-1 mice. The morning on the day a plug was found was designated as E0.5 and the plugged female was individually housed upon discovery of the plug. At the beginning of each experiment, embryos were staged according to morphological criteria to confirm gestation period. All experiments were approved by IACUC/AAALAC.

Intact mammary gland isolation. Aseptic technique was used during adult and fetal mammary gland dissections for cell preparations. Stereoscopes (Leica) with oblique bottom illumination were used to visualize fetal mammary glands during dissection. The isolated intact fetal mammary rudiments include both the epithelium and their surrounding stroma. Fetal and adult mammary glands intended for enzymatic digestions were placed directly in Epicult-B media containing collagenase and hyauronidase (see Cell preparation). Intact fetal and adult mammary fragments (~1mm in size) dissected for transplants were placed in DMEM (Mediatech) supplemented with 10% fetal bovine serum

(FBS) (Chemicon). A fluorescent stereoscope (Leica) was utilized to visualize Actin-eGFP+ mammary glands and Actin-eGFP+ positive mammary repopulated fat pads.

Cell preparation. *Adult mammary glands* were obtained from 12-16-week-old nulliparous Actin-eGFP or CD-1 female mice. Minced adult mammary glands were digested for 8-12 hours at 37^{o}C in supplemented EpiCult-B media containing collagenase and hyaluronidase. All reagents were purchased from Stem Cell Technologies (SCT), unless otherwise specified. Mammary organoids resulting from overnight digestion were treated with Trypsin followed by Dispase, and then washed and resuspended with Hank's Balanced Salt Solution (HBSS) supplemented with 2% FBS. Adult single cells were collected by passing the suspension through a 40µm nylon filter (BD Biosciences). *CD-1 or Actin-eGFP fetal mammary glands* were obtained from various embryonic stages. On average 150 mammary rudiments were dissected on each experimental day. Dissected fetal mammary glands were digested for 90 minutes at 37^{o}C in supplemented EpiCult-B media containing collagenase and hyaluronidase. Fetal mammary glands were treated with Dispase, washed and resuspended with Hank's Balanced Salt Solution supplemented with 2% FBS. The collection of ~150 E13.5-E15.5 rudiments generally yielded >200,000 bulk mammary cells, which were then used for subsequent flow cytometric analyses and/or transplant studies (see below).

Cell labeling, flow cytometry and sorting. Single cell suspensions were incubated on ice with Fc receptor antibody (2.4G, SCT) for 15 minutes, then incubated with biotinylated CD31 (SCT), biotinylated CD45 (SCT) and biotinylated TER119 (SCT) for 10 minutes, and then followed by streptavidin-PerCPcy5.5 (BD Biosciences), CD24-PE (M1/69, SCT) and CD49f-FITC (GoH3, SCT) for 20 minutes. Antibody-labeled cells were resuspended and incubated in HBSS with 2% FBS containing DAPI (Roche) for live/dead discrimination. In cases where the cells came from Actin-eGFP+ donor female mice, CD49f-FITC was replaced with biotinylated CD49f (GoH3 BD Biosciences) and streptavidin-APCCy7 (BD Biosciences). In that situation, cells were incubated with biotinylated CD31, biotinylated CD45 and biotinylatedTER119, followed by streptavidin-PerCPcy5.5 and washed twice before adding biotinylated CD49f and streptavidin-APCcy7 sequentially. For detection of PSA-NCAM, an APC conjugated antibody (2-2B, MACS) and mouse IgM isotype control (IS5-20C4, MACS) was utilized with CD24-PE and CD49f-FITC. For the detection of un-polysialylated NCAM, a biotin conjugated antibody (MEM-188, Novus Biologicals) and mouse IgG2a isotype control (MOPC-173, Novus Biologicals) was utilized followed by incubation with streptavidin-APCcy7. Mammary cells were lineage depleted by gating out PerCPCy5.5 positive cells or by using magnetic bead mediated lineage depletion (SCT) in addition to live/dead discrimination using either DAPI or a viability dye conjugated to eFluor450 (65-0863, eBioscience). Cell sorting was

carried out on either a FACSDiVa or FACSAria II cell sorter (Becton Dickinson) and the purity of sorted population was routinely >95%. Data were processed using FlowJo software.

In order to directly determine how fetal populations correspond to existing MRU sorting strategies, we processed fetal and adult cells together for subsequent analysis. The fetal cells were isolated from eGFP positive embryos and mixed with adult non-eGFP expressing tissue for processing, labeling, and flow cytometry. We use nomenclature to describe the fMRU-enriched population that is consistent with both the Stingl and Smalley groups, referring to the fMRU-enriched population as $CD24^{high}$ (≥500 units in our plots), $CD49f^{high}$. The fStromal-enriched population contains cells with a range of intermediate staining for CD24, from CD24 medium (between 100 and 500 units) to low (between 10 and 100 units) to negative (<10 units, a level of staining equivalent to the IgG control).

***In vivo* mammary transplantation and analysis.** The surgical technique to clear the mammary fat pads and to conduct mammary transplantation has been well established(Deome et al., 1959). Briefly, the inguinal #4 and #9 mammary glands of 21-26 day old CB17-SCID recipients were surgically removed and the cleared (epithelial-free) mammary fat pads left behind served as the transplantation sites. *Mammary transplantation of single cell suspensions*: cells (freshly dissociated and/or sorted) in 2-5ul of Hank's

balanced salt solution supplemented with 2% FBS, mixed with Matrigel 1:1 (BD Biosciences) and 0.05% Trypan blue (Sigma), were injected into each cleared fat pads by mouth pipette using fire polished glass capillaries (Fisher). In some cases, the cells were not co-injected with Matrigel. *Mammary transplantation of intact mammary pieces*: a small pocket was made in the fat pad using sharp forceps (Fine Science Tool, Dumont no. 5), and then one intact mammary piece was inserted into the pocket in the fat pad. Transplanted glands were removed from the recipients 6-12 weeks post-surgery for mammary repopulation evaluation. The excised glands were laid onto slides as whole mounts and visualized for mammary regeneration either by direct fluorescence microscopy or by staining the whole mounts with carmine aluminum sulfate (Sigma). In cases where specified, some of the recipients were mated with males to evaluate alveolar development in the regenerated mammary. We scored a positive for mammary outgrowths in repopulated fat pads when two of the following three criteria were met; a) connected branching network, b) visible tertiary branches and c) terminal end buds.

Immunofluorescence staining and microscopic imaging. Intact mammary rudiments for whole mount staining were prepared by dissecting and then fixing rudiments for ~4 hours in 4% PFA at room temperature, washing 3 times in PBST (PBS-/0.2% Triton X-100), and then blocking and permeablizing in

10% goat serum, 0.1% Triton X-100 for 1 hour. Whole rudiments were then incubated in primary antibody overnight at 4^{o}C, washed in PBST, incubated for 1 hour in secondary antibodies, washed in PBST, and then mounted in Vectashield. Paraffin embedded sections were made by fixing mammary tissues in 4% PFA overnight, then dehydrating the tissues through 70%, 95%, and 100% ethanol, treated with xylene, washed in paraffin 4 – 5 times at 55ºC, and then incubated in paraffin overnight. 10-20um sections were generated and then deparaffinized and hydrated before retrieving antigens in citrate buffer by microwaving for 6 minutes on high power. The sections were then washed in PBS and rinsed in water. Sections were blocked for 3 hours in 10% goat serum, incubated in primary antibody overnight at 4ºC, incubated in secondary antibodies for 1 hour at room temperature, and then mounted in Vectashield. For detection of Axin2-LacZ, β-galactosidase was visualized as previously detailed (Hogan et al., 1994). Briefly, embryos were dissected and puncture with forceps were made to allow access of the fixative, followed by fixation in 4% PFA for 2 hours. The fixed embryos were washed for 15 minutes, three times in detergent rinse containing 0.1M phosphate buffer (pH 7.3), 2mM $MgCl_2$, 0.01% sodium deoxycholate, and 0.02% Nonidet P-40 (NP-40). The embryos were then stained in the dark at 37ºC for 2 – 5 hours in staining solution containing 0.1M phosphate buffer (pH 7.3), 2mM $MgCl_2$, 0.01% sodium deoxychlolate, 0.02% NP-40, 5mM potassium ferricyanide, 5mM potassium ferrocyanide, 20mM Tris (pH 7.3), and 1mg/mL of X-gal.

Embryos were stored in 70% ethanol at 4ºC. For sectioning, embryos were dehydrated through alcohol to xylene and then embedded in paraffin as described above. For staining of dissociated cells, cells were cytospun at 800RPM onto Superfrost/Plus slides (Fisher), allowed to air dry, rinsed in PBS+, and then fixed for 15min in 4% PFA (Electron Microscopy Sciences), blocked for 1 hour in 10% goat serum/PBS- at room temperature, and then incubated in primary antibodies overnight at 4ºC. Cytospins were then washed and incubated in secondary antibodies for 1 hour at room temperature, washed and mounted in Prolong Gold. Cytospun cells were quantified for their different subtypes using OpenLab 3.1.7 software to create density slices and run Boolean operations to count DAPI positive cells with or without different marker expression. Antibodies used were Cytokeratin 14 (AF-64, Covance, 1:1000), Cytokeratin 8 (Troma-1, DSHB, 1:100), CD24 (M1/69, BD Biosciences, 1:1000), CD49f (GoH3, BD Biosciences, 1:1000), Casein (a gift from Gilbert Smith and Daniel Medina, 1:25), pan-cytokeratin (ab9377, Abcam), PSA-NCAM (2-2B, MACS, 1:15), Vimentin (AB5733, Chemicon, 1:1000) and ErbB2 (29D8, Cell Signaling, 1:500). Secondary antibodies were purchased from Molecular Probes/Invitrogen: anti-rat Alexafluor 488 (1:1000), anti-rabbit Alexafluor 568 and 660 (1:1000), anti-chicken Alexafluor 488 (1:1000). All sections were imaged using a Leica TCS SP2 AOBS confocal microscope or a Zeiss LSM 710 Laser Scanning confocal microscope.

In vitro* three-dimensional sphere forming cultures for fetal and adult mammary cells.** All reagents were purchased from SCT except where indicated. *Mammosphere culture*: freshly sorted cells were plated on ultra-low adherence plates (Corning) at a density of 1000 cells/cm^2 in supplemented Epicult-B mouse media containing rhEGF, rhbFGF, Heparin and Penicillin/Streptomycin (Chemicon). A similar culturing condition replacing supplemented Epicult-B basal media with DMEM/F12 media (Cellgro) and B27 (Invitrogen) was also used and yielded similar results. *2% Matrigel culture:** freshly sorted cells from the fMRU-enriched population were plated on ultra-low adherence plates (Corning) at a density of 1000cells/cm^2 in Epicult-B mouse media containing 2% Matrigel (growth factor reduced, BD Biosciences), B supplement, rhEGF, rhbFGF, Heparin and Penicillin/Streptomycin (Chemicon). Mammary cells from E15.5 were plated at a density of 20,000 cells per well in the same media. Cells were treated with carbachol (C4382, Sigma) at the indicated doses. *100% Matrigel culture***:** freshly sorted cells were seeded on top of 35□l bed of Matrigel (growth factor reduced) per well of an 8-well chamber slide (BD Falcon 354108) in Epicult-B mouse media containing supplement, 5% FBS, 2% Matrigel (growth factor reduced), rhEGF, rhbFGF, Heparin and Penicillin/Streptomycin (Chemicon). Either vehicle (DMSO), Lapatinib (LC Laboratories) or Neratinib (HKI-272, Pfizer) were added at the indicated doses to the media. Media was changed every 3 days. After 8-10 days, cultures were photographed and spheres were

scored. *Clonal sphere culture.* Two different cell plating strategies were employed to demonstrate that the fMRU-enriched population derived spheres originated from single cells rather than from aggregates. First, single cells were sorted into 96well plates at a single cell per well density. The cells were cultured in Epicult-B mouse media supplemented with 2% Matrigel (growth factor reduced), rhEGF, rhbFGF, Heparin and Penicillin/Streptomycin. New media was added every 3 days. After 8-10 days, cultures were photographed and spheres were scored. We detected single spheres emerging from 10% of the wells (the sphere forming efficiency was 10.7 ± 1.5%, n=3) (Figure S8). In parallel, we also isolated the fMRU-enriched populations from wild-type CD1 and Actin-eGFP, mixed the cells in either 1:1 or 1:4 ratios, and seeded them at low density on top of Matrigel (growth factor reduced). The cells were also cultured in Epicult-B mouse media supplemented with 2% Matrigel (growth factor reduced), rhEGF, rhbFGF, Heparin and Penicillin/Streptomycin (Chemicon). New media was added every 3 days. After 8-10 days, cultures were photographed and spheres were scored. All of the spheres examined were either eGFP+ or non eGFP with the exception of one sphere (out of 208), which consisted of a mixture of eGFP+ and eGFP- cells.

Two-dimensional control cultures. 10,000 MCF10A/HER2 cells (Wang et al., 2006) or BT549 (ATCC) cells were plated in triplicate wells of a 96 well dish in RPMI supplemented with 10% FBS and insulin. MCF10A/HER2 cells

were additionally supplemented with 5% horse serum, rhEGF, hydrocortisone, cholera toxin, insulin and Penicillin/Streptomycin (Chemicon). Cells were allowed to adhere and recover for 24 hours. Cells were then treated with Lapatinib or Neratinib at the indicated doses diluted in DMSO or with Vehicle (DMSO) alone for 48 hours and live cells were counted using Trypan blue exclusion.

Three-dimensional control cultures.2000 MCF10A/HER2 cells or BT549 cells were suspended in 30ul Matrigel (growth factor reduced) and spread to cover the surface of one well of an 8 well chamber slide (BD Falcon 354108). Matrigel was allowed to congeal in a tissue culture incubator at 37ºC for 25 minutes. Semisolid Matrigel/cell mixtures were overlaid with 300ul of MCF10A media containing B27 supplement (Invitrogen) and containing Lapatinib or Neratinib (or vehicle) at the indicated doses. Colonies were allowed to grow at 37ºC and 5% CO2 in a humidified tissue culture incubator for 10 days with media changes on the 4^{th} and 7^{th} days. Sphere sizes were measured and compiled from a representative 50X microscope field on a Leica DMIL LED microscope from each of the duplicate wells for all treatments.

Quantitative real-time PCR analyses on Estrogen Receptor, Progesterone Receptor, and ErbB family members. RNA was isolated, reverse transcribed, and pre-amplified from 40,000 cells using the CellsDirect

kit from Invitrogen. Pre-amplification and QPCR was performed using TaqMan® Gene Expression Assays (Applied Biosystems, see below). The qRT-PCR was carried out according to the manufacturer's recommendations on an ABI 7900HT Fast Real-Time PCR system. Data was analyzed using Applied Biosystems Signal Detection System 2.3 and Microsoft Excel. Cts were normalized between samples using the HPRT gene and relative fold differences between means was calculated using the formula $2^{\Delta CT}$.

Single Cell qRT-PCR analysis. For single cell gene expression analyses, we used qRT-PCR 96.96 dynamic arrays (Fluidigm). Single cells were sorted by FACS into a total of four 96-well plates containing a reaction mix for reverse transcription (CellsDirect, Invitrogen) and pre-amplification (0.2X concentration of TaqMan Assays, Applied Biosystems, see below for list). In addition, 10, 100 and 1000 cells were sorted in triplicate. The cDNA samples were amplified for 18 PCR cycles and then diluted fivefold in TE. All samples were evaluated for expression of GAPDH on an ABI 7900HT. The 10, 100, and 1000 cell dilutions were used to create a standard curve such that the expected Ct value for a single cell could be extrapolated. We used single cell samples that had detectable GAPDH levels roughly equivalent to single cells. The dynamic array chip was primed using control line fluid (Fluidigm) in a HX IFC controller (Fluidigm) and then 2.25µL of the cDNA samples were transferred into the chip sample inlets after mixing with TaqMan universal PCR

master mix (2.5µL, Applied Biosystems) and sample loading reagent (0.25µL, Fluidigm). The individual TaqMan assays (Applied Biosystems) were mixed with equal volumes of assay loading reagent (Fluidigm) and then transferred into the chip assay inlets. The chip was then loaded by the chip loader and then transferred to the reader (Biomark, Fluidigm) for thermocycling and fluorescent quantification.

The fMRU-enriched population was analyzed by running duplicate chips. All sample Ct values were normalized to GAPDH; cells with no detectable GAPDH expression were dropped from analysis. GAPDH normalization included a 10 – 20 fold expression correction to visualize relative expression within positive ranges. HPRT and/or GusB were also included as housekeeping genes, and normalization to these genes gave similar results but reduced the number of processed samples. ΔCt values were subjected to a log2 redistribution and then clustered using the Pearson correlation coefficient.

Southern Blot Analysis. gDNA was isolated from 96 well plates and digested with Acc651 and Xho1 for 24 hours at 37ºC. The digested gDNA and bromophenol blue dye were loaded onto a 0.7% agarose gel approximately 0.8cm thick. The gel was run overnight at 30 – 40V. The next day, fresh bromophenol blue dye was loaded into the outside and middle wells as a color indicator for acid and base treatment. The dye was allowed to run until fully

within the wells. The gel was then visualized with ethidium bromide followed by acid treatment in 0.25N HCl until the dye turned yellow. The gel was then treated in 0.5N NaOH, 1.5N NaCl until the dye returned to its original blue color. The gel was then neutralized in 0.5M Tris pH7.4, 1.5M NaCl with two 15 minute washes, followed by a rinse in transfer buffer. The membrane was soaked in water for 5 minutes followed by storage in 10X SSC. A sandstone sponge was placed in a reservoir of 10X SSC. Upon flooding the surface, three pieces of Whatman filter paper were placed on the sponge and wetted with 10X SSC. The gel was then transferred onto this apparatus and then framed by 4 strips of plastic wrap. The nylon membrane was placed on top of the framed gel and any bubbles were removed. Five additional pieces of Whatman paper were placed on top of the membrane. Paper towels were stacked on top of the transfer apparatus, followed by a Plexiglas plate. The transfer was allowed to proceed overnight.

The next day, the membrane is rinsed in wash buffer consisting of 0.1X SSPE, 0.1% SDS followed by UV crosslinking. The membrane is then transferred in wash buffer into a hybridization tube. The wash buffer is decanted and then blocking buffer was added. Blocking buffer consist of 7.5mL of Formamide, 150µL of 10% SDS, 300µL of Denhardt's, 3.75mL of 20X SSPE, 3mL of Dextran Sulfate, and 200µL of boiled and snap cooled ssDNA. Blocking was allowed to proceed for 4 hours in a rotating oven at 42ºC. During this incubation time, probe was generated from fresh [α-P^{32}]

dCTP using the Prime-It random primer labeling kit (Strategene) followed by purification using the Purelink PCR purification kit (Invitrogen). At least 20 million cpm of probe was mixed with 100µL of ssDNA, followed by incubation at 95ºC for 5 minutes and snap cooling on ice. This mixture was then added to the blocking buffer in the hybridization tube and allowed to incubate at 42ºC in the rotating oven overnight.

The next day, the membrane was washed in wash buffer pre-warmed to 58ºC for 20min followed by another wash at 60ºC. The membrane was then sealed into Mylar for exposure on a phosphor-imager screen. After exposure was allowed to continue for 24 hours, it was scanned on a Typhoon 8600 (Molecular Dynamics).

Prometheus Analysis. Samples for evaluation by Prometheus were lysed on ice using Protein Later (Prometheus) in the presence of a cocktail of protease and phosphatase inhibitors. Whole rudiment samples were minced, then vortexed and sonicated in Protein Later. Cell suspensions and spheres were pelleted and then resuspended in Protein Later on ice and then transported to Prometheus for analysis.

Recombinase mediated cassette exchange. Approximately 8 x 10^5 ES cells were resuspended in electroporation buffer. 15µg of a plasmid containing CRE was pre-mixed with 200µg of plasmid containing the exchange cassette with a

final volume less than 100µL. The DNA mixture was transferred to the ES cells and mixed, followed by transfer to an electroporation cuvette. A single pule was applied at 230V, 500µF. The cells were allowed to incubate at room temperature for 10 minutes before plating at a density of less than 10^5 cells per 10mL dish already containing arrested, triple resistance MEF feeders. Additional dilutions of 1:40 and 1:80 were included as bystander killing can contribute to loss of correctly targeted clones. After 1 – 2 days, selection with 0.2µM FIAU and 200µg/mL of G418 was applied for 7 – 10 days, or until distinct colonies were apparent. Colonies were picked into 96 well plates already containing arrested triple resistant MEFs and allowed to expand through 48 and duplicate 24 well plates. One 24 well plate was allowed to reach full confluency for gDNA isolation and PCR genotyping while the other plate was frozen down.

Statistical Analysis. Limiting dilution analysis of mammary repopulating unit frequency was performed using the ELDA web interface (http://bioinf.wehi.edu.au/software/elda/index.html) (Hu and Smyth, 2009), which includes a goodness of fit test based on the complementary loglog (cloglog) model originally developed by Bonnefoix and colleagues (Bonnefoix et al., 1996). ELDA was developed to account for limiting dilution analysis data where the assays are predominantly all positive or all negative. The active cell frequency estimate is based on the single-hit Poisson model, with frequency

range given by the usual two sided 95% Wald confidence intervals except in cases with 0% or 100% responses where a modified Clopper-Pearson 1-sided confidence interval was given (Hu and Smyth, 2009). In instances where the single-hit Poisson model hypothesis could not be assessed because the cloglog model did not hold, a situation which can result from insufficient number of repopulated fat pads, those frequencies (†) are noted as “rough MRU frequency estimate” (as recommended in personal communication by Thierry Bonnefoix at Institut National de la Sante de la Recherche Medicale, INSERM U823, Institut Bonniot, Grenoble, France).

Microarray Analyses and Statistics. To obtain sufficient material for microarray analysis on rare biological samples (E15.5 (data not shown), fMRU-enriched (E18.5 $CD24^{high}CD49f^{high}$) and fStromal-enriched (E18.5 $CD24^{med/low/neg}$)), independent replicate pools were created. Each pool consisted of RNA isolated and amplified from at least two independent dissection sessions. For each E18.5 dissection session, we isolated as much material as possible from embryos between 5:00AM and 8:00AM on the day of dissociation and sorting. This generally consisted of between 100 and 200 rudiments from 10-20 female embryos derived from 2-4 timed pregnant female mice and yielded up to 30,000 fMRU cells. The E15.5 array data (which is not analyzed in this manuscript, but comprises a part of the normalized array data submitted to the public archive (see below)) was derived from a single

dissection session per sample to obtain at least 100 E15.5 mammary buds per sample. Adult samples were derived from a single dissection session each using 4 adult inguinal (#4) mammary glands from two mice as the starting material. Dissociated unsorted cells were gently centrifuged and resuspended in cold lysis buffer, and total RNA was extracted using the RNAqueous kit (Ambion). For sorted cells, we sorted directly into ice cold lysis buffer. RNA integrity was analyzed using an Agilent 2100 Bioanalyzer and RNA quality measures (RIN scores) were greater than 9.0 when sufficient RNA to yield a reading was present. DNase treated RNA samples were subsequently reverse transcribed using Superscript III reverse transcriptase (Invitrogen) and a custom T7 oligo-dT primer, 5'-GGCACGTGAATTGTAATACGACTCA CTATAGGGAGGCGG$(T)_{20}$-3'(pre-amplification) or random hexamers (post-amplification). Samples were amplified according to the linear procedure described by Van Gelder et. al. (Van Gelder et al., 1990) using T7 RNA Polymerase (Ambion) and a 14 hour IVT incubation time. Yield was determined using a NanoDrop ND8000 small scale spectrophotometer (Thermo Scientific).

Complimentary DNA libraries were combined into independent replicate pools as described above and were labeled using the Nimblegen One-Color DNA Labeling Kit, hybridized to a Nimblegen 12x135k MM9 mouse microarray at the UCSD-GeneChip Microarray Core (Roche Nimblegen) and scanned on a Genepix4000B, all according to manufacturer's recommendations. Raw

intensities were RMA normalized (Irizarry et al., 2003). The MIAME compliant array data are available from the gene expression omnibus at http://www.ncbi.nlm.nih.gov/ under accession GSE27027.

We removed probes with max intensity for any sample lower than the 10^{th} percentile of mean probewise intensities from further analysis to reduce noise. This minor noise reduction corresponds roughly to the minimum intensity yielding fold differences of >2-fold between any sample type mean and the overall mean across all samples. Intensities were Log2 transformed and median centered for each probe. Differentially expressed genes were subsequently identified using the Significance Analysis of Microarrays (SAM) (Tusher et al., 2001) plugin for Microsoft Excel in pairwise comparisons between fMRU-enriched (E18.5 $CD24^{high}CD49f^{high}$) fStromal-enriched (E18.5 $CD24^{low\text{-}med}$) and adult MaSC–enriched (aMRU; Lin-$CD24^{med}CD49f^{high}$) (Stingl et al., 2006b) samples with FDR manually adjusted to <0.1. The fold difference limit was set at 1.5 fold, however the minimum fold difference identified under these criteria for any sample was 2.3 fold. The mean and median fold differences across all samples were 7.6 fold and 4.7 fold respectively. SAM results including individual fold differences are listed. These analyses resulted in the identification of 1,308 probes up-regulated in the fMRU-enriched population relative to fStromal-enriched population and 1,047 probes down-regulated in the fMRU-enriched population relative to the fStromal-enriched population. The 1,047 probes can conversely be regarded as up-regulated in

the fStromal relative to fMRU-enriched population and were therefore designated as the fSTR signature.

The fMRU/fSTR reciprocal interaction model was adapted from GeneGo interaction networks (http://genego.com) and was constructed using fMRU and fSTR gene sets with automated filters for “receptors and ligands” and manual filtering for direct interactions. GeneGo uses a curated interaction database to designate input genes having gene products that are thought or known to interact. We manually filtered the results to leave only those input genes exhibiting at least one interaction to another input gene. As such, the model presents a restricted set of candidate interactions.

Molecular Function, Biological Process and Pathway enrichment analyses were carried out using the BINGO plug-in for Cytoscape (Cline et al., 2007; Maere et al., 2005) or Gene Ontology and KEGG Pathway annotations in GeneGo and through the DAVID website (http://david.abcc.ncifcrf.gov/) with the background created using the full list of probes on the Nimblegen array. The annotations we report have $p<0.01$ for BINGO and for DAVID met the modified Fisher Exact P-Value, EASE criteria of $p<0.05$ for whole gene set enrichment and $p<0.1$ for small gene clusters. Small clusters utilized the whole database as background. Translation of probe IDs into Entrez IDs and human orthologues was carried out using the Clone Gene ID converter at CNIO (http://idconverter.bioinfo.cnio.es/IDconverter.php) followed by the Ensembl database at Biomart (http://biomart.org) and finally DAVID

(http://david.abcc.ncifcrf.gov) to fill in additional IDs not found by CNIO. All resources were based on NCBIM37 and GRCh37.p2 genome builds. These “translated” ID’s were subsequently mapped for human orthologues using the human mouse orthology list downloaded from the Jackson laboratory website on April 6, 2010, (http://jax.org). The step-wise, translated gene lists are given in Table S5along with the number of extant probes at each stage. Reduction in gene number during translation is a well-known phenomenon, since not all probes on the microarray correspond to known, unique genes. Therefore the number of genes identified as differentially expressed by microarray will be fewer than the number of differentially expressed probes identified. Similarly, not all mouse genes have known human orthologues. Therefore the number of genes translated into human orthologues will be fewer than the starting number of genes. Finally, not all orthologous human genes will be represented on human microarrays leading to a further reduction in genes tested for differential expression in human tumor microarrays.

QRT-PCR was performed on select targets for orthogonal confirmation of differential expression using TaqMan® Gene Expression Assays (Applied Biosystems, see TaqMan assay table). PCRs on material amplified as described above were carried out according to the manufacturer’s recommendations on an ABI 7900HT Fast Real-Time PCR system. Data was analyzed using Applied Biosystems Signal Detection System 2.3 and Microsoft Excel. Cts were set manually to a consensus phase of exponential

amplification for the entire population and samples showing erratic, grossly non-exponential or non-monophasic amplification was excluded. Cts were normalized between samples using the HPRT gene. Relative fold differences between means were calculated using the idealized formula $2^{\Delta CT}$ and the value for the cell type with the higher mean Ct for each gene was set to one. p values were determined by Students t-test (two-tailed, equal variance).

Signatures from published work representing differentially expressed adult mMaSC and mStromal gene sets were generated from normalized, Log2 transformed array data associated with Lim et al. (Lim et al., 2010) downloaded from the gene expression omnibus at http://www.ncbi.nlm.nih.gov/ (GSE19446) and were analyzed using SAM with median centering and FDR manually set to <0.1 as described above to generate population specific gene sets since the complete list of individual genes comprising the reported signatures were not publicly available at the time of our analyses. The hMaSC and hStromal signatures were as reported in Lim et al. (Lim et al., 2009). We also determined that 95% of the differentially expressed genes reported using LIMMA at FDR=0.05 (from the original report) were identified as differentially expressed using SAM at FDR<0.1 on the primary array data downloaded from GEO (GSE16997) (not shown). The “cultured hMaSC” gene set reported in Pece et al. (Pece et al., 2010) was taken directly from the published report.

Gene set similarities were determined by enumerating the overlap in the gene lists resulting from SAM analyses or as published in original reports (Lim et al., 2009; Pece et al., 2010). A cumulative hypergeometric probability was calculated for the overlap based on 20,309 and 19,828 probes in compendiated mouse and human arrays identified below, respectively. Probabilities calculated for signature overlaps used the 3338 genes represented in at least one signature as the set of possible genes.

To assess fetal gene expression profile enrichment in tumor samples, we obtained published, annotated microarray compendia from the public domain. Human tumor data was downloaded as annotated, pre-normalized and centered values from Ben-Porath et al. (http://jura.wi.mit.edu/bioc/benporath/) and represents ~1,200 tumor samples from 6 independent studies (Ben-Porath et al., 2008). Intrinsic subtype annotations were as determined in Ben-Porath et al. (Ben-Porath et al., 2008) except that a subset of tumors were re-annotated as “Claudin low-like” and “Metaplastic-like” based on their repression or enrichment for gene clusters and signatures described in Hershkowitz et al. (Herschkowitz et al., 2007) and Hennesey et al. (Hennessy et al., 2009). These enrichments as well as the gene set enrichments among tumor samples we observed were determined according to previously published methods (Segal et al., 2005) using Genomica software (http://genomica.weizmann.ac.il/). Briefly, genes expressed at least two fold above or below the compendium mean were considered differentially

expressed. We then assessed the fraction of over- or under expressed genes that belong to each tested gene set, calculating a *P* value according to the hypergeometric distribution. This was repeated for every sample, using a threshold of $P < 0.05$ for significant enrichment. Benjamini-Hochberg multiple hypothesis testing correction was employed with FDR <0.05. Random gene sets equivalent in size to population derived gene sets were constructed from random ranking of genes using the Excel random number generator. Five or more random gene sets were generated for each population derived gene set and never resulted in class-wise enrichments comparable to population derived gene sets. The same procedure was carried out independently on array data comprising the NKI295 (Parker et al., 2009) and UNC datasets (Prat et al., 2010) obtained from the UNC microarray database at https://genome.unc.edu/.

As an additional measure of similarity between fetal mammary populations and tumors we used Distance Weighted Discrimination and Single Sample Predictor as described previously (Benito et al., 2004; Hu et al., 2006). The software was downloaded from https://genome.unc.edu/pubsup/dwd/. The replicates representing the fMRU-enriched and the fSTR enriched samples and the set of differentially expressed genes as determined by SAM were used as the training data. We used standardized DWD centered at 0 and measured Euclidean distances for each sample relative to the mean euclidean distance for the dataset.

To compare enrichment patterns for different intrinsic subtype or grade designations, we calculated the fraction of samples of a given designation showing significant enrichment or repression for a particular gene set, and assigned a *P* value according to the hypergeometric distribution. In other experiments, we established that there exists general agreement between this enrichment approach (Segal et al., 2005) and an alternative statistical strategy, GSEA, (Subramanian et al., 2005) in subtype gene set enrichments (not shown).

The 96 genes in each fetal gene set with the greatest variance across the breast cancer arrays were subsequently selected for sample-class and gene-cluster discovery. We chose 96 high variance genes since higher variance can yield greater statistical power in delineating gene and sample classes and in order to facilitate ongoing studies and potential clinical applications in standard 12 by 8 formats.

To select coordinately expressed gene subsets, we took a clustering approach similar to that originally used to identify ‘intrinsic’ breast cancer subtypes, although care must be taken in interpreting hierarchical clusters here as in the original studies (Perou et al., 1999) since these can be subject to initial condition effects and the specific gene set employed. Clustering and graphic representations were generated in the TM4 suite MeV software (Saeed et al., 2003). Hierarchical clustering used euclidean distance and average linkage as metrics. Nine clusters of 5 or more genes per cluster were

selected for further characterization following assessment of the stability of their coordinate expression by measuring their reproducibility under permutation analysis using the Support Trees algorithm in MeV.

Kaplan-Meier analysis was carried out on the, Chin et al. (Chin et al., 2006) and Miller et al. (Miller et al., 2005) datasets using MedCalc™ following classification of samples according to the enrichment methods described above. The same analysis was carried out on the NKI-295 data set (van de Vijver et al., 2002b) as annotated in Parker et al. (Parker et al., 2009). *P* values represent log rank test results and scores can be corrected for multiple testing of 9 gene clusters by the formula, P_{shown}x9=$P_{corrected}$. Multivariate analysis also used Medcalc and the Parker-annotated, NKI-295 dataset. Significant enrichment or depletion for individual subsignatures was determined as above and tumor subsignature annotations were tested in models containing various combinations of categorical lymph node positivity, tumor size, ER status and grade as indicated.

TaqMan Assays Utilized

Gene	Assay ID	Gene	Assay ID	Gene	Assay ID
Abcg2	Mm00496364_m1	Fgf10	Mm00433275_m1	Ngfr	Mm01309635_m1
Arhgap5	Mm00501557_m1	FZD9	Mm01206511_s1	Nr2f1	Mm00657937_m1
Axin2	Mm00443610_m1	GAPDH	Mm99999915_g1	Nrg1	Mm01212130_m1
Bmi1	Mm00776122_gH	gata3	Mm00484683_m1	Nrg2	Mm01158088_m1
CD10 (Mme)	Mm01285052_m1	Gli3	Mm00492333_m1	NTRK3	Mm00456222_m1
CD24a	Mm0078258_sH	Gnaq	Mm00492381_m1	Olfr620	Mm00748427_s1
Cd24a	Mm00782538_sH	Gnas	Mm00507037_m1	p63	Mm00495788_m1
Cd34	Mm00519283_m1	GusB	Mm03003537_s1	PDGFc	Mm00480205_m1
Cd44	Mm01277163_m1	HPRT	Mm01324427_m1	PDGFra	Mm01211694_m1
Cdkn1a	Mm00432448_m1	Hprt1	Mm01545399_m1	Pgr	Mm00435628_m1
Cldn6	Mm00490040_s1	Itga4	Mm00439770_m1	Pkm2	Mm00834102_gH
Cldn9	Mm00517434_s1	Itga6	Mm00434375_m1	Plp1	Mm01297210_m1
Cmtm5	Mm00509113_m1	Itgb1	Mm01253227_m1	Reln	Mm00465200_m1
Cnp	Mm01306640_m1	K05	Mm00503549_m1	RunX1	Mm01213405_m1
Cxcr4	Mm01292123_m1	K08	Mm00835759_m1	Sema3b	Mm00436477_m1
Cyclin D1	Mm00432360_m1	K14	Mm00516876_m1	Serpinb5	Mm00436763_m1
Ddr1	Mm01273494_g1	K18	Mm01601702_g1	Sfrp1	Mm00489161_m1
Dkk1	Mm00438422_m1	K19	Mm00492980_m1	Snai2	Mm00441531_m1
EDNRB	Mm00432989_m1	L1cam	Mm00493049_m1	Sox10	Mm01300162_m1
EGFR	Mm0043023_m1	Lef1	Mm00550265_m1	Sox2	Mm00488369_s1
EGFR	Mm0043023_m1	Lgr5	Mm00438890_m1	Tbx3	Mm00809779_s1
EpCam (Tacstd1)	Mm00493214_m1	Lmo4	Mm00495373_m1	Tcf3	Mm00493456_m1
Erbb2	Mm00658541_m1	Ly6a	Mm00726565_s1	Tnc	Mm00495662_m1
Erbb3	Mm01159990_g1	Mag	Mm00487538_m1	Tlr3	Mm01207403_m1
Erbb4	Mm01256793_m1	Msi2	Mm00475180_m1	Tspan8	Mm00524563_m1
Esr1	Mm01191130_m1	Muc1	Mm00449604_m1	Twist1	Mm00442036_m1
Esr1	Mm00433149_m1	Myc	Mm00487803_m1	Vim	Mm00449208_m1
Fgf07	Mm00433291_m1	Ncam1	Mm03053534_s1		

References

Al-Hajj, M., Wicha, M.S., Benito-Hernandez, A., Morrison, S.J., and Clarke, M.F. (2003). Prospective identification of tumorigenic breast cancer cells. Proc Natl Acad Sci U S A *100*, 3983-3988.

Asch, B.B., Burstein, N.A., Vidrich, A., and Sun, T.T. (1981). Identification of mouse mammary epithelial cells by immunofluorescence with rabbit and guinea pig antikeratin antisera. Proc Natl Acad Sci U S A *78*, 5643-5647.

Asselin-Labat, M.L., Vaillant, F., Sheridan, J.M., Pal, B., Wu, D., Simpson, E.R., Yasuda, H., Smyth, G.K., Martin, T.J., Lindeman, G.J.*, et al.* (2010). Control of mammary stem cell function by steroid hormone signalling. Nature *465*, 798-802.

Barker, N., Ridgway, R.A., van Es, J.H., van de Wetering, M., Begthel, H., van den Born, M., Danenberg, E., Clarke, A.R., Sansom, O.J., and Clevers, H. (2009). Crypt stem cells as the cells-of-origin of intestinal cancer. Nature *457*, 608-611.

Barker, N., van Es, J.H., Kuipers, J., Kujala, P., van den Born, M., Cozijnsen, M., Haegebarth, A., Korving, J., Begthel, H., Peters, P.J.*, et al.* (2007). Identification of stem cells in small intestine and colon by marker gene Lgr5. Nature *449*, 1003-1007.

Baron, U., Gossen, M., and Bujard, H. (1997). Tetracycline-controlled transcription in eukaryotes: novel transactivators with graded transactivation potential. Nucleic acids research *25*, 2723-2729.

Beachy, P.A., Karhadkar, S.S., and Berman, D.M. (2004). Tissue repair and stem cell renewal in carcinogenesis. Nature *432*, 324-331.

Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. (2008). An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. Nat Genet *40*, 499-507.

Benito, M., Parker, J., Du, Q., Wu, J., Xiang, D., Perou, C.M., and Marron, J.S. (2004). Adjustment of systematic microarray data biases. Bioinformatics *20*, 105-114.

Bonavia, R., Inda, M.D., Cavenee, W.K., and Furnari, F.B. (2011). Heterogeneity maintenance in glioblastoma: a social network. Cancer Res *71*, 4055-4060.

Bonfanti, L. (2006). PSA-NCAM in mammalian structural plasticity and neurogenesis. Prog Neurobiol *80*, 129-164.

Bonnefoix, T., Bonnefoix, P., Verdiel, P., and Sotto, J.J. (1996). Fitting limiting dilution experiments with generalized linear models results in a test of the single-hit Poisson assumption. J Immunol Methods *194*, 113-119.

Bonnefoix, T., and Callanan, M. (2009). Reassessing the human mammary stem cell concept by modeling limiting dilution transplantation assays. Nat Med *15*, 602-604; author reply 604-605.

Bonnet, D., and Dick, J.E. (1997). Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med *3*, 730-737.

Boras-Granic, K., Chang, H., Grosschedl, R., and Hamel, P.A. (2006). Lef1 is required for the transition of Wnt signaling from mesenchymal to epithelial cells in the mouse embryonic mammary gland. Dev Biol *295*, 219-231.

Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A.*, et al.* (2009). Genes that mediate breast cancer metastasis to the brain. Nature *459*, 1005-1009.

Bourguignon, L.Y., Peyrollier, K., Xia, W., and Gilad, E. (2008). Hyaluronan-CD44 interaction activates stem cell marker Nanog, Stat-3-mediated MDR1 gene expression, and ankyrin-regulated multidrug efflux in breast and ovarian tumor cells. J Biol Chem *283*, 17635-17651.

Brewer, B.G., Mitchell, R.A., Harandi, A., and Eaton, J.W. (2009). Embryonic vaccines against cancer: an early history. Exp Mol Pathol *86*, 192-197.

Briegel, K.J. (2006). Embryonic transcription factors in human breast cancer. IUBMB Life *58*, 123-132.

Calbo, J., van Montfort, E., Proost, N., van Drunen, E., Beverloo, H.B., Meuwissen, R., and Berns, A. (2011). A functional role for tumor cell

heterogeneity in a mouse model of small cell lung cancer. Cancer Cell *19*, 244-256.

Chang, H.Y., Sneddon, J.B., Alizadeh, A.A., Sood, R., West, R.B., Montgomery, K., Chi, J.-T., Rijn, M.v.d., Botstein, D., and Brown, P.O. (2004). Gene Expression Signature of Fibroblast Serum Response Predicts Human Cancer Progression: Similarities between Tumors and Wounds. PLoS Biol *2*, e7.

Cheon, S.S., Cheah, A.Y., Turley, S., Nadesan, P., Poon, R., Clevers, H., and Alman, B.A. (2002). beta-Catenin stabilization dysregulates mesenchymal cell proliferation, motility, and invasiveness and causes aggressive fibromatosis and hyperplastic cutaneous wounds. Proceedings of the National Academy of Sciences of the United States of America *99*, 6973-6978.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.-L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T.*, et al.* (2006). Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *10*, 529-541.

Christofk, H.R., Vander Heiden, M.G., Harris, M.H., Ramanathan, A., Gerszten, R.E., Wei, R., Fleming, M.D., Schreiber, S.L., and Cantley, L.C. (2008). The M2 splice isoform of pyruvate kinase is important for cancer metabolism and tumour growth. Nature *452*, 230-233.

Christofori, G. (2003). Changing neighbours, changing behaviour: cell adhesion molecule-mediated signalling during tumour progression. EMBO J *22*, 2318-2323.

Chu, E.Y., Hens, J., Andl, T., Kairo, A., Yamaguchi, T.P., Brisken, C., Glick, A., Wysolmerski, J.J., and Millar, S.E. (2004). Canonical WNT signaling promotes mammary placode development and is essential for initiation of mammary gland morphogenesis. Development *131*, 4819-4829.

Cicalese, A., Bonizzi, G., Pasi, C.E., Faretta, M., Ronzoni, S., Giulini, B., Brisken, C., Minucci, S., Di Fiore, P.P., and Pelicci, P.G. (2009a). The Tumor Suppressor p53 Regulates Polarity of Self-Renewing Divisions in Mammary Stem Cells. *138*, 1083-1095.

Cicalese, A., Bonizzi, G., Pasi, C.E., Faretta, M., Ronzoni, S., Giulini, B., Brisken, C., Minucci, S., Di Fiore, P.P., and Pelicci, P.G. (2009b). The tumor

suppressor p53 regulates polarity of self-renewing divisions in mammary stem cells. Cell *138*, 1083-1095.

Clarke, M.F., Dick, J.E., Dirks, P.B., Eaves, C.J., Jamieson, C.H., Jones, D.L., Visvader, J., Weissman, I.L., and Wahl, G.M. (2006). Cancer Stem Cells--Perspectives on Current Status and Future Directions: AACR Workshop on Cancer Stem Cells. Cancer Res *66*, 9339-9344.

Cline, M.S., Smoot, M., Cerami, E., Kuchinsky, A., Landys, N., Workman, C., Christmas, R., Avila-Campilo, I., Creech, M., Gross, B.*, et al.* (2007). Integration of biological networks and gene expression data using Cytoscape. Nat Protoc *2*, 2366-2382.

Creighton, C.J., Li, X., Landis, M., Dixon, J.M., Neumeister, V.M., Sjolund, A., Rimm, D.L., Wong, H., Rodriguez, A., Herschkowitz, J.I.*, et al.* (2009). Residual breast cancers after conventional therapy display mesenchymal as well as tumor-initiating features. Proc Natl Acad Sci U S A *106*, 13820-13825.

Cremer, H., Chazal, G., Goridis, C., and Represa, A. (1997). NCAM is essential for axonal growth and fasciculation in the hippocampus. Mol Cell Neurosci *8*, 323-335.

Cremers, N., Deugnier, M.A., and Sleeman, J. (2010). Loss of CD24 expression promotes ductal branching in the murine mammary gland. Cell Mol Life Sci *67*, 2311-2322.

Croker, A.K., Goodale, D., Chu, J., Postenka, C., Hedley, B.D., Hess, D.A., and Allan, A.L. (2009). High aldehyde dehydrogenase and expression of cancer stem cell markers selects for breast cancer cells with enhanced malignant and metastatic ability. J Cell Mol Med *13*, 2236-2252.

Daniel, C.W., De Ome, K.B., Young, J.T., Blair, P.B., and Faulkin, L.J. (1968). The in vivo life span of normal and preneoplastic mouse mammary glands: a serial transplantation study. Proceedings of the National Academy of Sciences *61*, 53-60.

DasGupta, R., and Fuchs, E. (1999). Multiple roles for activated LEF/TCF transcription complexes during hair follicle development and differentiation. Development (Cambridge, England) *126*, 4557-4568.

Davies, J.A. (1996). Mesenchyme to epithelium transition during development of the mammalian kidney tubule. Acta Anat (Basel) *156*, 187-201.

Deng, S., Yang, X., Lassus, H., Liang, S., Kaur, S., Ye, Q., Li, C., Wang, L.P., Roby, K.F., Orsulic, S.*, et al.* (2010). Distinct expression levels and patterns of stem cell marker, aldehyde dehydrogenase isoform 1 (ALDH1), in human epithelial cancers. PLoS One *5*, e10277.

Deome, K.B., Faulkin, L.J., Jr., Bern, H.A., and Blair, P.B. (1959). Development of mammary tumors from hyperplastic alveolar nodules transplanted into gland-free mammary fat pads of female C3H mice. Cancer Res *19*, 515-520.

Desmedt, C., Haibe-Kains, B., Wirapati, P., Buyse, M., Larsimont, D., Bontempi, G., Delorenzi, M., Piccart, M., and Sotiriou, C. (2008). Biological processes associated with breast cancer clinical outcome depend on the molecular subtypes. Clin Cancer Res *14*, 5158-5165.

Dick, J.E. (2008). Stem cell concepts renew cancer research. Blood *112*, 4793-4807.

Diehn, M., Cho, R.W., Lobo, N.A., Kalisky, T., Dorie, M.J., Kulp, A.N., Qian, D., Lam, J.S., Ailles, L.E., Wong, M.*, et al.* (2009). Association of reactive oxygen species levels and radioresistance in cancer stem cells. Nature *458*, 780-783.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003a). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes Dev *17*, 1253-1270.

Dontu, G., Abdallah, W.M., Foley, J.M., Jackson, K.W., Clarke, M.F., Kawamura, M.J., and Wicha, M.S. (2003b). In vitro propagation and transcriptional profiling of human mammary stem/progenitor cells. Genes & Development *17*, 1253-1270.

Du, L., Wang, H., He, L., Zhang, J., Ni, B., Wang, X., Jin, H., Cahuzac, N., Mehrpour, M., Lu, Y.*, et al.* (2008). CD44 is of functional importance for colorectal cancer stem cells. Clin Cancer Res *14*, 6751-6760.

Dunbar, M.E., Dann, P., Brown, C.W., Van Houton, J., Dreyer, B., Philbrick, W.P., and Wysolmerski, J.J. (2001). Temporally regulated overexpression of parathyroid hormone-related protein in the mammary gland reveals distinct fetal and pubertal phenotypes. The Journal of endocrinology *171*, 403-416.

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S., Nobel, A.B., van't Veer, L.J., and Perou, C.M. (2006). Concordance among gene-expression-based predictors for breast cancer. N Engl J Med *355*, 560-569.

Fan, C., Prat, A., Parker, J.S., Liu, Y., Carey, L.A., Troester, M.A., and Perou, C.M. (2011). Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. BMC Med Genomics *4*, 3.

Gershengorn, M.C., Hardikar, A.A., Wei, C., Geras-Raaka, E., Marcus-Samuels, B., and Raaka, B.M. (2004). Epithelial-to-mesenchymal transition generates proliferative human islet precursor cells. Science *306*, 2261-2264.

Gilbert, D.M., Neilson, A., Miyazawa, H., DePamphilis, M.L., and Burhans, W.C. (1995). Mimosine Arrests DNA Synthesis at Replication Forks by Inhibiting Deoxyribonucleotide Metabolism. J Biol Chem *270*, 9597-9606.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S.*, et al.* (2007). ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. Cell Stem Cell *1*, 555-567.

Godar, S., Ince, T.A., Bell, G.W., Feldser, D., Donaher, J.L., Bergh, J., Liu, A., Miu, K., Watnick, R.S., Reinhardt, F.*, et al.* (2008). Growth-inhibitory and tumor- suppressive functions of p53 depend on its repression of CD44 expression. Cell *134*, 62-73.

Goridis, C., and Brunet, J.F. (1992). NCAM: structural diversity, function and regulation of expression. Semin Cell Biol *3*, 189-197.

Gossen, M., Freundlieb, S., Bender, G., Muller, G., Hillen, W., and Bujard, H. (1995). Transcriptional activation by tetracyclines in mammalian cells. Science (New York, NY *268*, 1766-1769.

Guo, G., Huss, M., Tong, G.Q., Wang, C., Li Sun, L., Clarke, N.D., and Robson, P. (2010). Resolution of cell fate decisions revealed by single-cell gene expression analysis from zygote to blastocyst. Dev Cell *18*, 675-685.

Haibe-Kains, B., Desmedt, C., Piette, F., Buyse, M., Cardoso, F., Van't Veer, L., Piccart, M., Bontempi, G., and Sotiriou, C. (2008). Comparison of prognostic gene expression signatures for breast cancer. BMC Genomics *9*, 394.

Hay, E.D. (1995). An overview of epithelio-mesenchymal transformation. Acta Anat (Basel) *154*, 8-20.

Hendrix, M.J., Seftor, E.A., Seftor, R.E., and Trevor, K.T. (1997). Experimental co-expression of vimentin and keratin intermediate filaments in human breast cancer cells results in phenotypic interconversion and increased invasive behavior. Am J Pathol *150*, 483-495.

Hennessy, B.T., Gonzalez-Angulo, A.M., Stemke-Hale, K., Gilcrease, M.Z., Krishnamurthy, S., Lee, J.S., Fridlyand, J., Sahin, A., Agarwal, R., Joy, C.*, et al.* (2009). Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. Cancer Res *69*, 4116-4124.

Herschkowitz, J., Simin, K., Weigman, V., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K., Jones, L., Assefnia, S., Chandrasekharan, S.*, et al.* (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome Biology *8*, R76.

Hogan, B., Beddington, R., Costantini, F., and Lacy, E. (1994). Manipulating the mouse embryo.

Hu, Y., and Smyth, G.K. (2009). ELDA: extreme limiting dilution analysis for comparing depleted and enriched populations in stem cell and other assays. J Immunol Methods *347*, 70-78.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L.*, et al.* (2006). The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics *7*, 96.

Huntly, B.J., and Gilliland, D.G. (2005a). Cancer biology: summing up cancer stem cells. Nature *435*, 1169-1170.

Huntly, B.J., and Gilliland, D.G. (2005b). Leukaemia stem cells and the evolution of cancer-stem-cell research. Nat Rev Cancer *5*, 311-321.

Hynes, N.E., and Lane, H.A. (2005). ERBB receptors and cancer: the complexity of targeted inhibitors. Nat Rev Cancer *5*, 341-354.

Irizarry, R.A., Hobbs, B., Collin, F., Beazer-Barclay, Y.D., Antonellis, K.J., Scherf, U., and Speed, T.P. (2003). Exploration, normalization, and summaries of high density oligonucleotide array probe level data. Biostat *4*, 249-264.

Jacobs, T.W., Gown, A.M., Yaziji, H., Barnes, M.J., and Schnitt, S.J. (1999). Specificity of HercepTest in determining HER-2/neu status of breast cancers using the United States Food and Drug Administration-approved scoring system. J Clin Oncol *17*, 1983-1987.

Jaks, V., Barker, N., Kasper, M., van Es, J.H., Snippert, H.J., Clevers, H., and Toftgard, R. (2008). Lgr5 marks cycling, yet long-lived, hair follicle stem cells. Nat Genet *40*, 1291-1299.

Jamieson, C.H., Ailles, L.E., Dylla, S.J., Muijtjens, M., Jones, C., Zehnder, J.L., Gotlib, J., Li, K., Manz, M.G., Keating, A.*, et al.* (2004). Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. N Engl J Med *351*, 657-667.

Jemal, A., Siegel, R., Ward, E., Hao, Y., Xu, J., and Thun, M.J. (2009). Cancer statistics, 2009. CA Cancer J Clin *59*, 225-249.

Kanda, T., Sullivan, K.F., and Wahl, G.M. (1998). Histone-GFP fusion protein enables sensitive analysis of chromosome dynamics in living mammalian cells. Curr Biol *8*, 377-385.

Kendrick, H., Regan, J.L., Magnay, F.A., Grigoriadis, A., Mitsopoulos, C., Zvelebil, M., and Smalley, M.J. (2008). Transcriptome analysis of mammary epithelial subpopulations identifies novel determinants of lineage commitment and cell fate. BMC Genomics *9*, 591.

Kistner, A., Gossen, M., Zimmermann, F., Jerecic, J., Ullmer, C., Lubbert, H., and Bujard, H. (1996). Doxycycline-mediated quantitative and tissue-specific control of gene expression in transgenic mice. Proceedings of the National Academy of Sciences of the United States of America *93*, 10933-10938.

Kleinsmith, L.J., and Pierce, G.B., Jr. (1964). Multipotentiality of Single Embryonal Carcinoma Cells. Cancer Res *24*, 1544-1551.

Knox, S.M., Lombaert, I.M., Reed, X., Vitale-Cross, L., Gutkind, J.S., and Hoffman, M.P. (2010). Parasympathetic innervation maintains epithelial progenitor cells during salivary organogenesis. Science *329*, 1645-1647.

Korinek, V., Barker, N., Morin, P.J., van Wichen, D., de Weger, R., Kinzler, K.W., Vogelstein, B., and Clevers, H. (1997). Constitutive transcriptional activation by a beta-catenin-Tcf complex in APC-/- colon carcinoma. Science (New York, NY *275*, 1784-1787.

Korkaya, H., Paulson, A., Iovino, F., and Wicha, M.S. (2008). HER2 regulates the mammary stem/progenitor cell population driving tumorigenesis and invasion. Oncogene *27*, 6120-6130.

Lehembre, F., Yilmaz, M., Wicki, A., Schomber, T., Strittmatter, K., Ziegler, D., Kren, A., Went, P., Derksen, P.W., Berns, A.*, et al.* (2008). NCAM-induced focal adhesion assembly: a functional switch upon loss of E-cadherin. EMBO J *27*, 2603-2615.

Lenferink, A.E.G., Simpson, J.F., Shawver, L.K., Coffey, R.J., Forbes, J.T., and Arteaga, C.L. (2000). Blockade of the epidermal growth factor receptor tyrosine kinase suppresses tumorigenesis in MMTV/Neu + MMTV/TGF-beta bigenic mice. Proceedings of the National Academy of Sciences of the United States of America *97*, 9609-9614.

Li, X., Lewis, M.T., Huang, J., Gutierrez, C., Osborne, C.K., Wu, M.F., Hilsenbeck, S.G., Pavlick, A., Zhang, X., Chamness, G.C.*, et al.* (2008). Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. J Natl Cancer Inst *100*, 672-679.

Li, Y., and Rosen, J.M. (2005). Stem/progenitor cells in mouse mammary gland development and breast cancer. J Mammary Gland Biol Neoplasia *10*, 17-24.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A.*, et al.* (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med *15*, 907-913.

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

Lindvall, C., Evans, N.C., Zylstra, C.R., Li, Y., Alexander, C.M., and Williams, B.O. (2006). The Wnt signaling receptor Lrp5 is required for mammary ductal

stem cell activity and Wnt1-induced tumorigenesis. The Journal of biological chemistry *281*, 35081-35087.

Livasy, C.A., Karaca, G., Nanda, R., Tretiakova, M.S., Olopade, O.I., Moore, D.T., and Perou, C.M. (2006). Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol *19*, 264-271.

Luker, K.E., and Luker, G.D. (2006). Functions of CXCL12 and CXCR4 in breast cancer. Cancer Lett *238*, 30-41.

Lustig, B., Jerchow, B., Sachs, M., Weiler, S., Pietsch, T., Karsten, U., van de Wetering, M., Clevers, H., Schlag, P.M., Birchmeier, W*., et al.* (2002). Negative feedback loop of Wnt signaling through upregulation of conductin/axin2 in colorectal and liver tumors. Molecular and cellular biology *22*, 1184-1193.

Lyons, J.G., Lobo, E., Martorana, A.M., and Myerscough, M.R. (2008). Clonal diversity in carcinomas: its implications for tumour progression and the contribution made to it by epithelial-mesenchymal transitions. Clin Exp Metastasis *25*, 665-677.

Maere, S., Heymans, K., and Kuiper, M. (2005). BiNGO: a Cytoscape plugin to assess overrepresentation of gene ontology categories in biological networks. Bioinformatics *21*, 3448-3449.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M*., et al.* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715.

Maretto, S., Cordenonsi, M., Dupont, S., Braghetta, P., Broccoli, V., Hassan, A.B., Volpin, D., Bressan, G.M., and Piccolo, S. (2003). Mapping Wnt/beta-catenin signaling during mouse development and in colorectal tumors. Proceedings of the National Academy of Sciences of the United States of America *100*, 3299-3304.

Meyer, M.J., Fleming, J.M., Lin, A.F., Hussnain, S.A., Ginsburg, E., and Vonderhaar, B.K. (2010). CD44posCD49fhiCD133/2hi defines xenograft-initiating cells in estrogen receptor-negative breast cancer. Cancer Res *70*, 4624-4633.

Miller, L.D., Smeds, J., George, J., Vega, V.B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E.T.*, et al.* (2005). An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. Proc Natl Acad Sci U S A *102*, 13550-13555.

Mizuno, H., Spike, B.T., Wahl, G.M., and Levine, A.J. (2010). Inactivation of p53 in breast cancers correlates with stem cell transcriptional signatures. Proc Natl Acad Sci U S A *107*, 22745-22750.

Moraes, R.C., Zhang, X., Harrington, N., Fung, J.Y., Wu, M.F., Hilsenbeck, S.G., Allred, D.C., and Lewis, M.T. (2007). Constitutive activation of smoothened (SMO) in mammary glands of transgenic mice leads to increased proliferation, altered differentiation and ductal dysplasia. Development *134*, 1231-1242.

Morimoto, K., Kim, S.J., Tanei, T., Shimazu, K., Tanji, Y., Taguchi, T., Tamaki, Y., Terada, N., and Noguchi, S. (2009). Stem cell marker aldehyde dehydrogenase 1-positive breast cancers are characterized by negative estrogen receptor, positive human epidermal growth factor receptor type 2, and high Ki67 expression. Cancer Sci *100*, 1062-1068.

Moriyama, A., Kii, I., Sunabori, T., Kurihara, S., Takayama, I., Shimazaki, M., Tanabe, H., Oginuma, M., Fukayama, M., Matsuzaki, Y.*, et al.* (2007). GFP transgenic mice reveal active canonical Wnt signal in neonatal brain and in adult liver and spleen. Genesis *45*, 90-100.

Mork, C., van Deurs, B., and Petersen, O.W. (1990). Regulation of vimentin expression in cultured human mammary epithelial cells. Differentiation *43*, 146-156.

Morrison, S.J., and Kimble, J. (2006). Asymmetric and symmetric stem-cell divisions in development and cancer. Nature *441*, 1068-1074.

Navin, N., Krasnitz, A., Rodgers, L., Cook, K., Meth, J., Kendall, J., Riggs, M., Eberling, Y., Troge, J., Grubor, V.*, et al.* (2010). Inferring tumor progression from genomic heterogeneity. Genome Res *20*, 68-80.

Nielsen, P.J., Lorenz, B., Muller, A.M., Wenger, R.H., Brombacher, F., Simon, M., von der Weid, T., Langhorne, W.J., Mossmann, H., and Kohler, G. (1997). Altered erythrocytes and a leaky block in B-cell development in CD24/HSA-deficient mice. Blood *89*, 1058-1067.

Nimmo, R.A., and Slack, F.J. (2009). An elegant miRror: microRNAs in stem cells, developmental timing and cancer. Chromosoma *118*, 405-418.

Nusse, R., and Varmus, H.E. (1982). Many tumors induced by the mouse mammary tumor virus contain a provirus integrated in the same region of the host genome. Cell *31*, 99-109.

Paik, S., Kim, C., and Wolmark, N. (2008). HER2 status and benefit from adjuvant trastuzumab in breast cancer. N Engl J Med *358*, 1409-1411.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T.*, et al.* (2004). A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. N Engl J Med *351*, 2817-2826.

Parker, J.S., Mullins, M., Cheang, M.C.U., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z.*, et al.* (2009). Supervised Risk Predictor of Breast Cancer Based on Intrinsic Subtypes. J Clin Oncol *27*, 1160-1167.

Pece, S., Tosoni, D., Confalonieri, S., Mazzarol, G., Vecchi, M., Ronzoni, S., Bernard, L., Viale, G., Pelicci, P.G., and Di Fiore, P.P. (2010). Biological and molecular heterogeneity of breast cancers correlates with their cancer stem cell content. Cell *140*, 62-73.

Perez-Pomares, J.M., and Munoz-Chapuli, R. (2002). Epithelial-mesenchymal transitions: a mesodermal cell strategy for evolutive innovation in Metazoans. Anat Rec *268*, 343-351.

Perou, C., Jeffrey, S., van de Rijn, M., Rees, C., Eisen, M., Ross, D., Pergamenschikov, A., Williams, C., Zhu, S., and Lee, J. (1999). Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. Proc Natl Acad Sci USA *96*, 9212 - 9217.

Perou, C., Sorlie, T., Eisen, M., van de Rijn, M., Jeffrey, S., Rees, C., Pollack, J., Ross, D., Johnsen, H., and Akslen, L. (2000a). Molecular portraits of human breast tumours. Nature *406*, 747 - 752.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A.*, et al.* (2000b). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Petersen, O.W., and Polyak, K. (2010). Stem cells in the human breast. Cold Spring Harb Perspect Biol *2*, a003160.

Ponta, H., Wainwright, D., and Herrlich, P. (1998). The CD44 protein family. Int J Biochem Cell Biol *30*, 299-305.

Powers, S., and Mu, D. (2008). Genetic similarities Between Organogenesis and Tumorigenesis of the Lung. Cell Cycle *7*, 200-204.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., and Perou, C.M. (2010). Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res *12*, R68.

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

Quintana, E., Shackleton, M., Foster, H.R., Fullen, D.R., Sabel, M.S., Johnson, T.M., and Morrison, S.J. (2010). Phenotypic heterogeneity among tumorigenic melanoma cells from patients that is reversible and not hierarchically organized. Cancer Cell *18*, 510-523.

Quintana, E., Shackleton, M., Sabel, M.S., Fullen, D.R., Johnson, T.M., and Morrison, S.J. (2008). Efficient tumour formation by single human melanoma cells. Nature *456*, 593-598.

Rabindran, S.K., Discafani, C.M., Rosfjord, E.C., Baxter, M., Floyd, M.B., Golas, J., Hallett, W.A., Johnson, B.D., Nilakantan, R., Overbeek, E.*, et al.* (2004). Antitumor activity of HKI-272, an orally active, irreversible inhibitor of the HER-2 tyrosine kinase. Cancer Res *64*, 3958-3965.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. (2001). Stem cells, cancer, and cancer stem cells. Nature *414*, 105-111.

Ricardo, S., Vieira, A.F., Gerhard, R., Leitao, D., Pinto, R., Cameselle-Teijeiro, J.F., Milanezi, F., Schmitt, F., and Paredes, J. (2011). Breast cancer stem cell markers CD44, CD24 and ALDH1: expression distribution within intrinsic molecular subtype. J Clin Pathol.

Roesch, A., Fukunaga-Kalabis, M., Schmidt, E.C., Zabierowski, S.E., Brafford, P.A., Vultur, A., Basu, D., Gimotty, P., Vogt, T., and Herlyn, M. (2010). A

temporarily distinct subpopulation of slow-cycling melanoma cells is required for continuous tumor growth. Cell *141*, 583-594.

Ronn, L.C., Hartz, B.P., and Bock, E. (1998). The neural cell adhesion molecule (NCAM) in development and plasticity of the nervous system. Exp Gerontol *33*, 853-864.

Rusnak, D.W., Lackey, K., Affleck, K., Wood, E.R., Alligood, K.J., Rhodes, N., Keith, B.R., Murray, D.M., Knight, W.B., Mullin, R.J.*, et al.* (2001). The effects of the novel, reversible epidermal growth factor receptor/ErbB-2 tyrosine kinase inhibitor, GW2016, on the growth of human normal and tumor-derived cell lines in vitro and in vivo. Mol Cancer Ther *1*, 85-94.

Rutishauser, U., and Landmesser, L. (1996). Polysialic acid in the vertebrate nervous system: a promoter of plasticity in cell-cell interactions. Trends Neurosci *19*, 422-427.

Sadoul, R., Hirn, M., Deagostini-Bazin, H., Rougon, G., and Goridis, C. (1983). Adult and embryonic mouse neural cell adhesion molecules have different binding properties. Nature *304*, 347-349.

Saeed, A.I., Sharov, V., White, J., Li, J., Liang, W., Bhagabati, N., Braisted, J., Klapa, M., Currier, T., Thiagarajan, M.*, et al.* (2003). TM4: a free, open-source system for microarray data management and analysis. Biotechniques *34*, 374-378.

Sakakura, T., Nishizuka, Y., and Dawe, C.J. (1976). Mesenchyme-dependent morphogenesis and epithelium-specific cytodifferentiation in mouse mammary gland. Science *194*, 1439-1441.

Sakakura, T., Nishizuka, Y., and Dawe, C.J. (1979). Capacity of mammary fat pads of adult C3H/HeMs mice to interact morphogenetically with fetal mammary epithelium. J Natl Cancer Inst *63*, 733-736.

Sangiorgi, E., and Capecchi, M.R. (2008). Bmi1 is expressed in vivo in intestinal stem cells. Nat Genet *40*, 915-920.

Sato, N., Meijer, L., Skaltsounis, L., Greengard, P., and Brivanlou, A.H. (2004). Maintenance of pluripotency in human and mouse embryonic stem cells through activation of Wnt signaling by a pharmacological GSK-3-specific inhibitor. Nat Med *10*, 55-63.

Scadden, D.T. (2006). The stem-cell niche as an entity of action. Nature *441*, 1075-1079.

Sebens Muerkoster, S., Werbing, V., Sipos, B., Debus, M.A., Witt, M., Grobmann, M., Leisner, D., Kotteritzsch, J., Kappes, H., Kloppel, G.*, et al.* (2006). Drug-induced expression of the cellular adhesion molecule L1CAM confers anti-apoptotic protection and chemoresistance in pancreatic ductal adenocarcinoma cells. Oncogene *26*, 2759-2768.

Segal, E., Pe'er, D., Regev, A., Koller, D., and Friedman, N. (2005). Learning Module Networks. The Journal of Machine Learning Research *6*, 557-588.

Sell, S. (2010). On the stem cell origin of cancer. Am J Pathol *176*, 2584-2494.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., and Visvader, J.E. (2006). Generation of a functional mammary gland from a single stem cell. Nature *439*, 84-88.

Singh, S.K., Clarke, I.D., Hide, T., and Dirks, P.B. (2004). Cancer stem cells in nervous system tumors. Oncogene *23*, 7267-7273.

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

Smith, G.H., and Boulanger, C.A. (2003). Mammary epithelial stem cells: transplantation and self-renewal analysis. Cell Prolif *36 Suppl 1*, 3-15.

Smith, G.H., and Medina, D. (1988). A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. Journal of Cell Science *90*, 173-183.

Soriano, P. (1999). Generalized lacZ expression with the ROSA26 Cre reporter strain. Nat Genet *21*, 70-71.

Sorlie, T., Perou, C., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M., van de Rijn, M., and Jeffrey, S. (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci USA *98*, 10869 - 10874.

Sotiriou, C., and Piccart, M.J. (2007). Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? Nat Rev Cancer *7*, 545-553.

Southam, C.M., and Brunschwig, A. (1961). QUANTITATIVE STUDIES OF AUTOTRANSPLANTATION OF HUMAN CANCER - PRELIMINARY REPORT. Cancer *14*, 971-&.

Stern, D.F. (2003). ErbBs in mammary development. Exp Cell Res *284*, 89-98.

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

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

Subramaniam, V., Vincent, I.R., Gardner, H., Chan, E., Dhamko, H., and Jothy, S. (2007a). CD44 regulates cell migration in human colon cancer cells via Lyn kinase and AKT phosphorylation. Exp Mol Pathol *83*, 207-215.

Subramaniam, V., Vincent, I.R., Gilakjan, M., and Jothy, S. (2007b). Suppression of human colon cancer tumors in nude mice by siRNA CD44 gene therapy. Exp Mol Pathol *83*, 332-340.

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

Sun, P., Yuan, Y., Li, A., Li, B., and Dai, X. (2010). Cytokeratin expression during mouse embryonic and early postnatal mammary gland development. Histochem Cell Biol *133*, 213-221.

Thiery, J.P., and Sleeman, J.P. (2006). Complex networks orchestrate epithelial-mesenchymal transitions. Nat Rev Mol Cell Biol *7*, 131-142.

Thomas, P.A., Kirschmann, D.A., Cerhan, J.R., Folberg, R., Seftor, E.A., Sellers, T.A., and Hendrix, M.J. (1999). Association between keratin and

vimentin expression, malignant phenotype, and survival in postmenopausal breast cancer patients. Clin Cancer Res *5*, 2698-2703.

Tucker, A.S. (2007). Salivary gland development. Semin Cell Dev Biol *18*, 237-244.

Tumbar, T., Guasch, G., Greco, V., Blanpain, C., Lowry, W.E., Rendl, M., and Fuchs, E. (2004). Defining the epithelial stem cell niche in skin. Science *303*, 359-363.

Tusher, V., Tibshirani, R., and Chu, G. (2001). Significance analysis of microarrays applied to the ionizing radiation response. Proc Natl Acad Sci USA *98*, 5116 - 5121.

Urlaub, G., and Chasin, L.A. (1980). Isolation of Chinese Hamster Cell Mutants Deficient in Dihydrofolate Reductase Activity. PNAS *77*, 4216-4220.

van 't Veer, L., Dai, H., van de Vijver, M., He, Y., Hart, A., Mao, M., Peterse, H., van der Kooy, K., Marton, M., and Witteveen, A. (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530 - 536.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J.*, et al.* (2002a). A gene-expression signature as a predictor of survival in breast cancer. The New England journal of medicine *347*, 1999-2009.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., Marton, M.J.*, et al.* (2002b). A gene-expression signature as a predictor of survival in breast cancer. N Engl J Med *347*, 1999-2009.

Van Gelder, R.N., von Zastrow, M.E., Yool, A., Dement, W.C., Barchas, J.D., and Eberwine, J.H. (1990). Amplified RNA synthesized from limited quantities of heterogeneous cDNA. Proc Natl Acad Sci U S A *87*, 1663-1667.

van Genderen, C., Okamura, R.M., Farinas, I., Quo, R.G., Parslow, T.G., Bruhn, L., and Grosschedl, R. (1994). Development of several organs that require inductive epithelial-mesenchymal interactions is impaired in LEF-1-deficient mice. Genes & development *8*, 2691-2703.

Veltmaat, J.M., Mailleux, A.A., Thiery, J.P., and Bellusci, S. (2003). Mouse embryonic mammogenesis as a model for the molecular regulation of pattern formation. Differentiation *71*, 1-17.

Villadsen, R., Fridriksdottir, A.J., Ronnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M.A., Bissell, M.J., and Petersen, O.W. (2007). Evidence for a stem cell hierarchy in the adult human breast. J Cell Biol *177*, 87-101.

Vogel, C.L., Cobleigh, M.A., Tripathy, D., Gutheil, J.C., Harris, L.N., Fehrenbacher, L., Slamon, D.J., Murphy, M., Novotny, W.F., Burchmore, M.*, et al.* (2002). Efficacy and safety of trastuzumab as a single agent in first-line treatment of HER2-overexpressing metastatic breast cancer. J Clin Oncol *20*, 719-726.

Wang, S.E., Narasanna, A., Perez-Torres, M., Xiang, B., Wu, F.Y., Yang, S., Carpenter, G., Gazdar, A.F., Muthuswamy, S.K., and Arteaga, C.L. (2006). HER2 kinase domain mutation results in constitutive phosphorylation and activation of HER2 and EGFR and resistance to EGFR tyrosine kinase inhibitors. Cancer Cell *10*, 25-38.

Warburton, M.J., Hughes, C.M., Ferns, S.A., and Rudland, P.S. (1989). Localization of vimentin in myoepithelial cells of the rat mammary gland. Histochem J *21*, 679-685.

Warren, L., Bryder, D., Weissman, I.L., and Quake, S.R. (2006). Transcription factor profiling in individual hematopoietic progenitors by digital RT-PCR. Proc Natl Acad Sci U S A *103*, 17807-17812.

Weigelt, B., Lo, A.T., Park, C.C., Gray, J.W., and Bissell, M.J. (2010). HER2 signaling pathway activation and response of breast cancer cells to HER2-targeting agents is dependent strongly on the 3D microenvironment. Breast Cancer Res Treat *122*, 35-43.

White, D.E., Kurpios, N.A., Zuo, D., Hassell, J.A., Blaess, S., Mueller, U., and Muller, W.J. (2004). Targeted disruption of beta1-integrin in a transgenic mouse model of human breast cancer reveals an essential role in mammary tumor induction. Cancer Cell *6*, 159-170.

Wilson, A., Laurenti, E., Oser, G., van der Wath, R.C., Blanco-Bose, W., Jaworski, M., Offner, S., Dunant, C.F., Eshkind, L., Bockamp, E.*, et al.* (2008). Hematopoietic stem cells reversibly switch from dormancy to self-renewal during homeostasis and repair. Cell *135*, 1118-1129.

Wilson, A., Oser, G.M., Jaworski, M., Blanco-Bose, W.E., Laurenti, E., Adolphe, C., Essers, M.A., Macdonald, H.R., and Trumpp, A. (2007). Dormant and self-renewing hematopoietic stem cells and their niches. Ann N Y Acad Sci *1106*, 64-75.

Wirapati, P., Sotiriou, C., Kunkel, S., Farmer, P., Pradervand, S., Haibe-Kains, B., Desmedt, C., Ignatiadis, M., Sengstag, T., Schutz, F.*, et al.* (2008). Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures. Breast Cancer Res *10*, R65.

Wong, E.T., Kolman, J.L., Li, Y.C., Mesner, L.D., Hillen, W., Berens, C., and Wahl, G.M. (2005). Reproducible doxycycline-inducible transgene expression at specific loci generated by Cre-recombinase mediated cassette exchange. Nucleic acids research *33*, e147.

Yoshida, K., Rutishauser, U., Crandall, J.E., and Schwarting, G.A. (1999). Polysialic acid facilitates migration of luteinizing hormone-releasing hormone neurons on vomeronasal axons. J Neurosci *19*, 794-801.

Zecchini, S., and Cavallaro, U. (2010). Neural cell adhesion molecule in cancer: expression and mechanisms. Adv Exp Med Biol *663*, 319-333.

Zeng, Y.A., and Nusse, R. (2010). Wnt Proteins Are Self-Renewal Factors for Mammary Stem Cells and Promote Their Long-Term Expansion in Culture. Cell Stem Cell *6*, 568-577.

Zhang, M., Behbod, F., Atkinson, R.L., Landis, M.D., Kittrell, F., Edwards, D., Medina, D., Tsimelzon, A., Hilsenbeck, S., Green, J.E.*, et al.* (2008). Identification of Tumor-Initiating Cells in a p53-Null Mouse Model of Breast Cancer. Cancer Res *68*, 4674-4682.

Zhong, J.F., Chen, Y., Marcus, J.S., Scherer, A., Quake, S.R., Taylor, C.R., and Weiner, L.P. (2008). A microfluidic processor for gene expression profiling of single human embryonic stem cells. Lab Chip *8*, 68-74.