Stony Brook University

MicroRNA Regulation in Mammary Progenitor and Tumor-Initiating Cells

A Dissertation Presented

By

Ingrid Ibarra

To

The Graduate School

In Partial Fulfillment of the

Requirements

For the Degree of

Doctor of Philosophy

In

Genetics

Stony Brook University

August 2009

Stony Brook University

The Graduate School

Ingrid Ibarra

We, the dissertation committee for the above candidate for the

Doctor of Philosophy degree,

hereby recommend acceptance of this dissertation.

Dr. Gregory J Hannon, Dissertation Advisor
Professor, Cold Spring Harbor Laboratory, Genetics Program, Stony Brook University

Dr. Scott Lowe, Chairperson of Defense
Professor, Cold Spring Harbor Laboratory, Genetics Program, Stony Brook University

Dr. Rafaella Sordella
Assistant Professor, Cold Spring Harbor Laboratory
Genetics Program Stony Brook University

Dr. Howard Crawford
Associate Professor, Dept. Pharmacological Sciences,
Genetics Program, Stony Brook University

Dr. Anthony M.C. Brown
Associate Professor, Weill Cornell Medical College

This dissertation is accepted by the Graduate School

Lawrence Martin
Dean of the Graduate School

Abstract of the Dissertation

MicroRNA Regulation in Mammary Progenitor and Tumor-Initiating Cells

By

Ingrid Ibarra

Doctor of Philosophy

In

Genetics

Stony Brook University

2009

The mammary gland is a dynamic tissue that undergoes various morphological transitions during puberty, pregnancy, lactation and involution in response to ovarian-secreted hormones. While there is evidence for the existence of stem cells within the mammary epithelium, the location of these cells during mammary gland development remains elusive. Part of my thesis work has focused on the identification of stem/progenitor cell markers that would enable the detection of these populations from mixed cultures, as well as determining their spatial position during mammary development. Using an immortalized murine mammary cell line (called COMMA-Dβ) containing cells with reconstitution potential, we have characterized ALDH (aldehyde dehydrogenase) as a functional marker of self-renewing mammary progenitor cells.

In addition, my work has focused on the role of microRNAs in mammary epithelial stem cell maintenance. MicroRNAs are ~22 nucleotide RNA molecules that function in a dose-dependent and spatiotemporal manner. In general, miRNAs are

viewed as post-transcriptional regulators that fine tune transcriptional outputs as opposed to acting as on/off switches of gene expression. They have been implicated in cellular processes such as cell proliferation, cell death, and cell-fate determination. These small RNAs bind to the 3'UTR of messenger RNA in a sequence-dependent manner culminating in either mRNA degradation or translational repression. We cloned small RNAs from COMMA-Dβ mammary progenitors and their differentiated counter-parts sorted on the basis of ALDH activity and determined their miRNA signature. Several miRNAs, including miR-205, and miR-22 are highly expressed in mammary progenitor cells, while others, including let-7 and miR-93 are significantly down-regulated. Interestingly, let-7 expression is post-transcriptionally regulated in embryonic stem cells and neuronal precursors. Enforced let-7 expression results in the depletion of mammary progenitor cells in culture, suggesting a role for let-7 in differentiation programs.

We took advantage of these finding and generated let-7 and miR-93 miR-sensor constructs. Sensors were first used in plants to detect the expression patterns of miRNAs during development. They contain miRNA binding sites in the 3'UTR of a reporter transgene and take advantage of the sequence-complementarity required between miRNA:mRNA duplex. In the presence of a miRNA the reporter is silenced (by RNAi), whereas, in cells lacking the miRNA that the sensor is designed to detect, the transgene RNA is stable and allows reporter expression. We recently generated a let-7 sensor mouse that represses eGFP expression upon let-7 –mediated silencing. In various tissues, including the mammary gland, let-7 expression is down-regulated in primitive cell -types.

Many cancers seem to depend on a small population of cells (referred to as 'cancer stem cell' or 'tumor-initiating cells') for their continued growth, propagation and

metastatic potential. We have shown that enforced let-7 and miR-93 expression decreases the ALDH compartment (tumor-initiating cells) of HCC-1954 and Sum-159 cells. Enforced expression of let-7 and miR-93 in these cell lines perturbs tumor growth. We have also shown that sensor technology can be used for the enrichment of tumor-initiating cells in HCC-1954 and Sum-159 (personal communication Christophe Ginestier). Overall, we show that miR signatures can be used not only to define and track rare cell populations *in vitro* and *in vivo* but that manipulation of these signatures might be used to expand and deplete stem cell and tumor-initiating cell populations for therapeutic benefit.

TABLE OF CONTENTS

List of Abbreviations

MESC Mammary Epithelial Stem Cells
H2B Histone 2B
TEB Terminal End Bud
FGF-2 Fibroblast Growth Factor-2
TGFβ Transforming Factor Beta
FACS Flow Activated Cell Sorting
Sca-1 Stem Cell Antigen 1
ABC transporter ATP-binding Cassette
ALDH Aldehyde Dehydrogenase
miRNA MicroRNA
RNAi RNA Interference
3'UTR 3' Untranslated region
Let-7 lethal-7
AML Acute Myeloid Leukemia
CML Chronic Myeloid Leukemia
MAF Mafosfamide
DOX Doxycycline
GFP Green Florescent Protein
ESC Embryonic Stem Cell
EB Embroid Body
LIF Leukemia Inhibitory Factor
TETOP Tetracycline Responsive Operator
tTA Tetracycline-controlled Transactivator Protein
CFU-E Colony Forming Unit Erythroid
CFU-M Colony Forming Unit Macrophage
CFU-GEMM Colony Forming Granulocyte -Erythroid-Macrophage- Megakaryocyte
CFU-GM Colony Forming Unit Granulocyte Macrophage
LNA Locked Nucleic Acid
MSCV Murine Stem Cell Virus
HSC Hematopoietic Stem Cell
SP Side Population
RNA Ribonucleic Acid
shRNA Short-Hairpin RNA
MET Mesenchymal Epithelial Transition
EMT Epithelial Mesenchymal Transition
RFP Red Fluorescent Protein
NOD/SCID Nonobese Diabetes/Severe Combined Immunodeficiency
CMV Cytomegalovirus
ZEB1 Zinc finger E-box binding homeobox 1
Lgr-5 Leucine-rich repeat-containing G protein-coupled receptor 5,
DEAB Diethylamino-benzaldehyde

List of Figures:

List of Tables

Acknowledgements

The work described in this thesis would have been impossible without the enthusiasm and encouragement provided by several people that I have been so fortunate to have in my life.

First of all, when I think of Greg Hannon the following Chinese proverb comes to mind, "Give a man a fish and you feed him for a day. Teach a man to fish and you feed him for a lifetime." I have learned that if you're going to do something you should do it right. Of course, this entails numerous attempts, most of which fail, but if you don't love it than science is not for you.

I would like to recognize my thesis committee Scott Lowe, Rafaella Sordella, Howard Crawford and Anthony Brown. Being the only woman on my committee I am very grateful to have had Rafaella as a member. She is truly a role model for women in science. Aside from being beautiful she is extremely critical and never short of ideas. I'd like to thank Howard and Anthony for agreeing to participate on such a short notice and for their input and advice. Grisha Enikolopov and Senthil Muthuswamy provided insights and advice on mammary gland development and stem cell biology in general during my thesis committee meetings. I'd like to thank Scott Lowe for agreeing to be my chairperson despite his busy schedule, even though I've had to pay a heavy price.

I'd like to thank our collaborators in the Wicha Laboratory, Suling Liu and Christophe Ginestier for fruitful discussions; Maria Deugnier for sharing unpublished data on Comma-D cells; and Susan Lindeman for advice with using mafosfamide.

Our lab is really like a family, partly because we spent a lot of time together. Arguments are not uncommon and I think they are actually quite healthy. I am thankful to past Hannon Lab members: Elizebeth Murchison, Lin He, Jose Silva, Michelle Carmell, Fabiola Rivas, Monica Dus, Ahmet Denli, Julius Brennecke, Yijun Qi, and Despina Siolas. To current members: Katalin Fejes Toth, Alexi Aravin, Sihem Cheloufi, Astrid Hasse, Ben Czech, Colin Malone, Vasily Vagin, Ikuko Hotta, Fedor Karginovb, Xingue He, Camila Dos Santos, Emily Hodges, Michelle Rooks, Sabrina Boettcher, Ralph Burgess, Andres Canela, Yaniv Erlich, Assaf Gordon, Paloma Guzzardo Tamargo, Anna Jankowska, Marek Kudla, Ravi Sachidanandam, Krista Marran, Hana Mizuno, Antoine Molaro, Maria Mosquera, Felix Muerdter, Clare Rebbeck, , Frederick Rollins, Nikolay Rozhkov, Vihra Sotirova, Amy Valentine, and Maria Mead, I value their friendship. Greg has a gift for attracting the brightest of minds.

I have to thank Oliver Tam for taking the time to read the first draft of my thesis, which I know were rough. Without his patience and support I'm certain I would have gone mad. I have to thank my lab sisters, Despina, Kata, Sihem Astrid, Camila, and Krista for healthy distractions and never saying no to my wine/tequila ideas. I have to thank Kenneth Chang for sharing his wisdom, and Jo Leonardo, Kate Bell and Jerry Thomsen for helping me plan my thesis defense.

I'd like to thank Lisa Bianco for teaching me fat pad transplantation, Stephen Hearn for confocal microscopy assistance and Pam Moody for teaching me flow cytometry.

I would like to thank Peter Gergen for academic guidance during my first year in graduate school.
Lastly, I am very grateful to have parents, brothers and friends that spoil me with love.

CHAPTER 1

Introduction

1.1 Identification of Multipotent Cells in Normal Tissues and Cancer

Breast cancer has long been noted as a highly heterogeneous class of tumors, with a high level of diversity within patient specimens in terms of genetic alterations and genomic abnormalities. An attractive theory to explain this heterogeneity is that breast tumors are the end results of mutations in the stem or progenitor cells of the mammary gland. This cellular origin of breast cancer would explain, in part, the heterogeneity of a given tumor mass (having arisen from multiple tumorigenic stem or progenitor cells), resistance to current chemotherapy (which might not eliminate the cancerous stem cell population), and the metastatic potential of a subclass of tumor cells (containing the cancer "stem cells"). The elucidation of the stem and progenitor cell compartments in the mammary epithelium, and uncovering their potential role in tumorigenic processes, has been the focus of my thesis research.

A multipotent stem cell can be defined by two main properties; the ability to self-renew and maintain a multipotent population, and the ability to differentiate into unipotent progenies when given the appropriate cellular or environment signals. In particular, the potential for self-renewal is essential to maintain their numbers after every cellular division, thus ensuring the future progenies differentiate into tissue-specific cells and maintain tissue homeostasis over the lifetime of the animal. Both of these attributes are exemplified in cells located at the crypt base of intestinal epithelium, cells within the hair-bulge of skin, and hematopoietic stem cells in the bone marrow (Bach et al. 2000), Tumbar et al. 2004, (Morrison et al. 1995). In general various stem cells share common features such as an enormous proliferative capacity, anchorage-independent growth (in

culture), and resistance to chemical toxins. Interestingly, gene expression profiles of embryonic stem cells and neuronal precursors have provided evidence for a universal 'stemness' signature illustrated in figure 1 (Ivanova et al. 2002). It is likely that hallmark properties shared by stem cells, such as the ability to balance self-renewal and differentiation will be governed by similar molecular mechanisms.

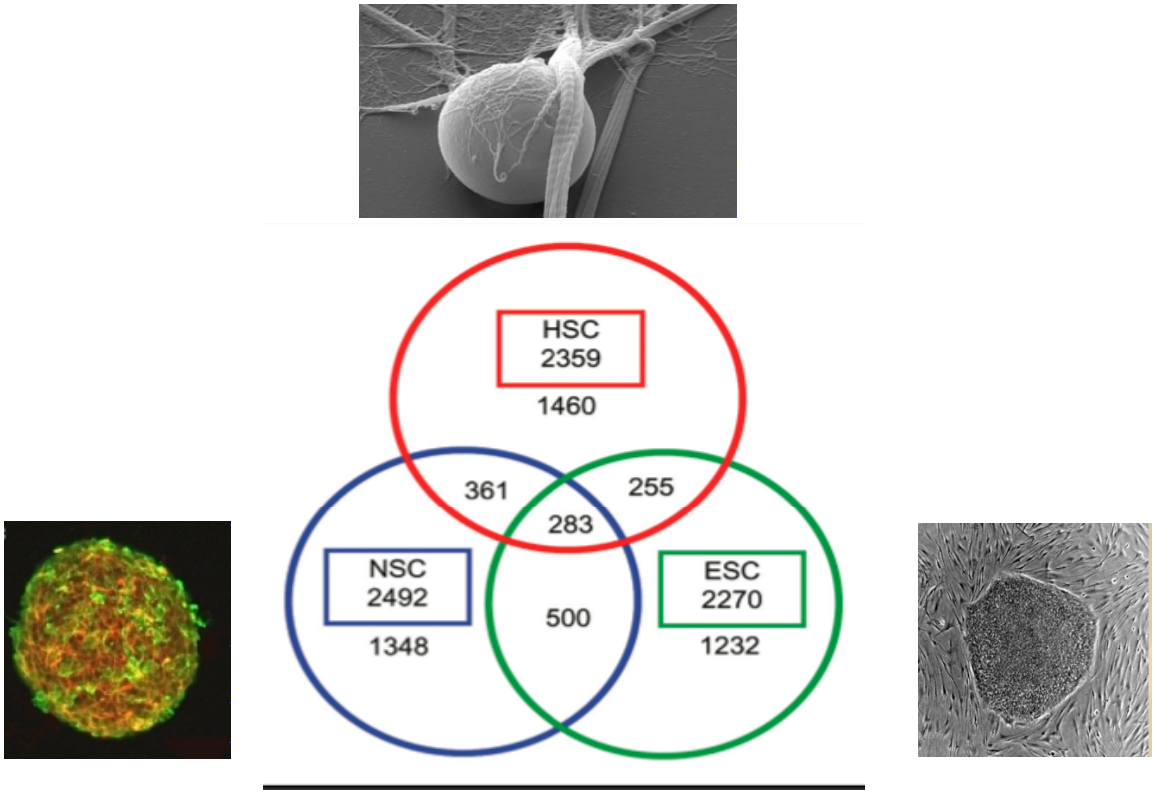


Figure 1.1 Overlapping gene expression in diverse murine stem cells.
Venn diagram detailing shared and distinct gene expression among hematopoietic stem cells (HSC), neuronal stem cells (NSC) and embryonic stem cells (ESC). Wnt and Notch signaling molecules are common to all. Adapted from Ivanova et al. 2002.

While somatic stem cells and their 'niche' have been well defined in tissues such as skin, colon, intestine and bone marrow (Tumbar et al. 2004, Kiel et al. 2005, Barker et al. 2007), the stem cell population within the mammary epithelium remains elusive

(Lindeman et al. 2008). As Figure 2 illustrates, there are various methods that have been employed in skin and intestine to identify these cells. These take advantage of the differential cell cycle length between the stem cells and the differentiated progeny. The slower cell cycle of the stem cell population in these tissues has allowed the retention of an H2B-GFP protein in the cell (histone 2B), whereas fast dividing cells rapidly turnover the label. Genes previously identified to be specific to stem cell compartments, such as Wnt target gene, lrg-5, have also been exploited to mark cells at the base of the intestinal crypt. Work in this thesis has attempted to identify mammary epithelial stem cell (MESC) markers to uncover the stem and progenitor compartments in the mammary epithelium.

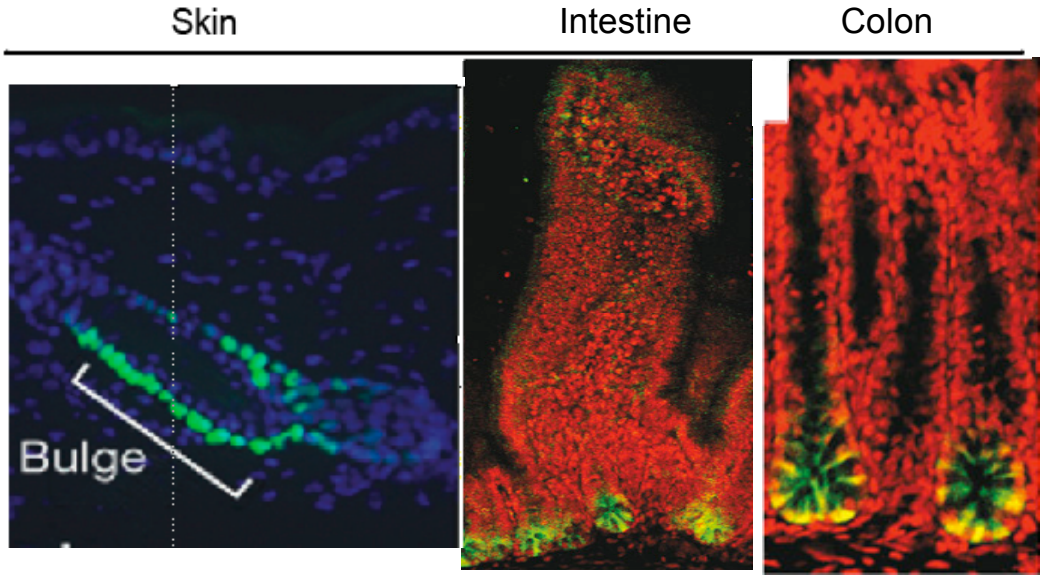


Figure 1.2 The Stem Cell Niche in Skin and Intestine
(a) Doxycycline- inducible H2B-GFP mice are used in pulse-chase experiments to identify cells that retain-label after 28 days. Lrg-5 is a Wnt-target gene and in (b) the intestine, Lgr5 is exclusively expressed in cycling crypt base columnar cells. (c) the crypt base of colon cells are also labeled with Lgr-5. Adapted from Elaine Fuchs 2009 and Barker et al. 2007

1.1.1 Evidence for Stem cells in Murine Mammary Epithelium

The mammary gland is a dynamic tissue that undergoes massive remodeling in response to ovarian-secreted hormones. The four morphologically distinct developmental stages of the mammary gland are puberty, pregnancy, lactation and involution. During puberty, structures that protrude through the fat pad and are responsible for branching morphogenesis called terminal end buds (TEB), characterize a virgin gland. Pregnancy is marked by alveolar differentiation, a process that is critical for the efficient production of milk proteins. Further tissue remodeling occurs during lactation in response to prolactin. During lactation, the epithelium expands contributing to the majority of cells within the mammary gland, fat cells are no longer a dominant population. The last stage in mammary development is involution and it is characterized by massive cell death as the epithelium reverts to a 'virgin-like state'. The precise location of stem and progenitor cells throughout these developmental stages remains unknown.

The evidence for the existence of stem cells within the mammary epithelium comes from fat pad transplantation and retroviral- tagging experiments. A study in the late 1960s revealed that primary mammary cells could reconstitute a mammary tree when injected into the cleared fat pad of a syngenetic female (Daniel et al. 1968), Faulkin et al. 1969). It was later shown that these cells could be found throughout the mammary gland and in all developmental stages (Smith and Medina 1988). In more recent experiments using retroviral integration as a marker, the entire mammary epithelial outgrowth (after a series of transplantation) can be traced back to a single cell/clone (Kordon and Smith 1998). In similar experiments, Smith et al. provided evidence for a hierarchical organization of stem cells and progenitors similar to that of the hematopoietic system (Smith 1996). Using limiting dilution experiments with retrovirally-tagged cells, they identified 3 cell populations with varying proliferative potential. They concluded that these represented ductal progenitor cells, alveolar progenitors, and multipotent stem cells (able to generate all cells types in the mammary epithelium). A parity-induced population of cells (marked by beta-galactosidase expression) were shown to generate both luminal and myoepithelial cells upon serial transplantation (Boulanger et al. 2005), suggesting the presence of multipotent cells during pregnancy. In human breast tissue, *in situ* studies of X-chromosome inactivation patterns indicated a common cell of origin in adjacent lobules and ducts (Tsai et al. 1996). In vitro colony-forming experiments in which cells are sorted at clonogenic density and morphologically assessed have also provided evidence for the existence of cells with multi-lineage potential within the human breast epithelium (Stingl et al. 1998).

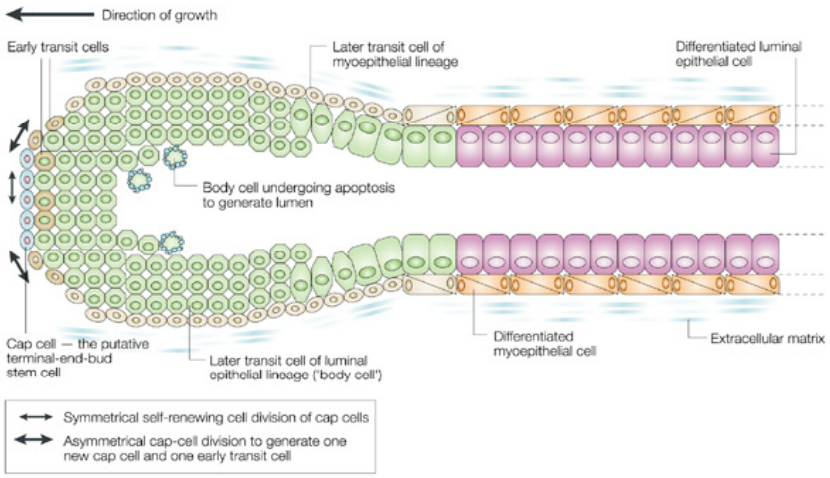


Figure 1.3 The ducts of the developing mammary gland, with their inner luminal epithelial cell layers and outer myoepithelial cell layers, are established as the terminal end buds (TEBs) move through the fat pad. It is thought that the cap cells at the tip of the TEB generate transit cells of a myoepithelial lineage on the outer side of the TEB and generate transit cells — known as 'body cells' — of a luminal epithelial lineage to form the central TEB mass. The ductal lumen is formed as central body cells apoptose and outer cells differentiate into luminal epithelial cells. Extracellular-matrix enzymes degrade the stroma in front of the TEB to enable it to move through the fat pad, but it is unclear how the structures actually 'move' through the gland. It might simply be that progressive cell division building up cell bulk at the front of the mass of body cells, coupled with progressive apoptosis degrading cell bulk at the back of the body cell mass, creates the illusion of forward movement. Adapted from Mathew Smalley and Alan Ashworth, 2003.

Despite functional experiments that show the existence of MESC, the presence of a mammary niche remains unclear. During puberty, terminal end buds are a major site of mitosis that facilitate ductal elongation and penetration of the fat pad to form the rudimentary mammary tree (Humphreys et al. 1996). TEB are 5-6 cell deep and are comprised of two positional distinct cell-types namely cap cells and body cells. Stem cells are hypothesized to reside within the cap layer of TEB of virgin mammary glands (fig.4) (Kenney et al. 2001), (Williams and Daniel 1983) (Smalley and Ashworth 2003). Caps cells are thought to divide asymmetrically to produce the outer myoepithelial layer while body cells are thought to give rise to luminal epithelial cells. TEB are intimately associated with stromal cells and mouse models of, FGF2 (fibroblast growth factor 2),

TGFß (transforming factor ß), and ERBB-3, have shown the critical importance of epithelial-stroma interactions for the maintenance of TEB structure and proper development of the mammary gland (Parsa et al. 2008), (Cheng et al. 2005b), (Jackson-Fisher et al. 2008). However, during pregnancy, lactation, and involution these stromal interactions are more difficult to assess and the niche remains unknown during these later developmental stages.

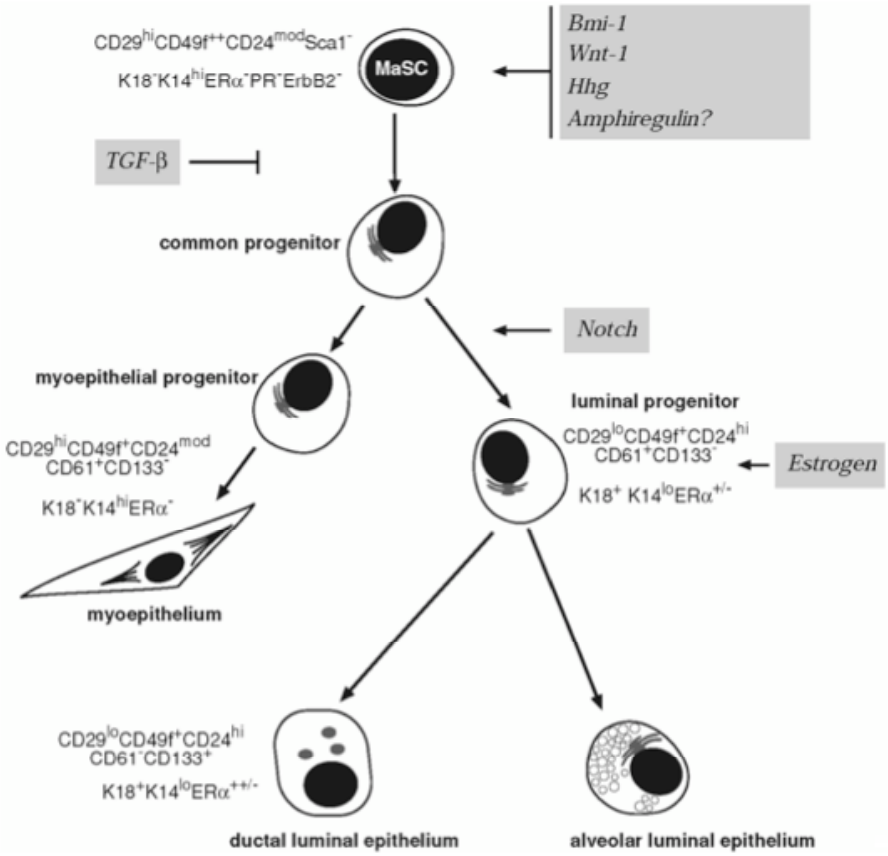


Figure 1.4 Schematic model of the mouse mammary epithelial hierarchy. The cell surface immunophenotypes and keratin expression profiles of the stem and intermediate cell types are shown. Genes and pathways that have been implicated in regulating cells in the hierarchy are indicated. (MaSC) Mammary stem cell. Adapted from (Asselin-Labat et al. 2008)

1.1.2 Current Mammary Stem/Progenitor Cell Markers

Stem cell markers are either cell-surface proteins that are preferentially expressed in stem/progenitor cells or functional markers that take advantage of 'intrinsic' stem cell properties. Although cell -surface proteins and functional stem cell markers lack the lineage resolution provided by retroviral-tagging cells, they provide a means of enrichment by either FACS (Floresence Activated Cells Sorting) or magnetic bead purification methods. The current hierarchical organization of stem and progenitor cells in the murine mammary gland based on cell surface proteins and fat pad transplantation experiments is summarized in figure 4 (Asselin-Labat et al. 2008). Cell surface proteins, CD-49f (α6 integrin), CD-29 (β-1 integrin) and CD-24 (heat- stable antigen), in combination with lineage depletion protocols, have been used to purify and isolate murine cells with reconstitution potential (Shackleton et al. 2006) (Stingl et al. 2006). Interestingly, in human breast tissue, luminal cell-restricted and bipotent mammary progenitors appear to express CD-49f (Stingl et al. 2001). Sca-1 (Stem cell antigen -1), which was previously used by (Welm et al. 2002) to enrich for murine MESC was not expressed in the CD49f^{hi}, CD24+ subset of cells with reconstitution potential. It was shown that expression of Sca-1 is induced during primary culture, and thus alone is regarded as a controversial marker (Shackleton et al. 2006) (Stingl et al. 2006). In addition, Sca-1 lacks a human homologue and thus somewhat of an unattractive murine MESC marker (Holmes and Stanford 2007). CD-24 has been localized to the apical plasma membrane of human luminal cells (Jones et al. 2004), although similar studies of mouse mammary tissue suggested a more ubiquitous pattern of expression (Shackelton et al. 2006). CD49f was shown to have a basal pattern of expression in the mammary

gland. CD29 expression was apparent in the cap cell region of terminal end buds, relative to mature ducts in which high expression was predominantly basolateral (Stingl et al. 2006). While these markers serve to enrich cells with reconstitution potential from bulk cultures, *in situ* they are unable to resolve a stem or progenitor compartment during mammary gland development.

Functional stem cell markers are more biologically relevant; as they reflect 'intrinsic' properties of stem cells, and therefore are not subjected to the caveat of induced expression when using cell-surface proteins. A side population phenotype (SP) has been used to enrich for MESC and other stem/progenitor cells including hematopoietic stem cells (HSC), and neural precursors (Goodell et al. 1996), (Bhattacharya et al. 2003), (Alvi et al. 2003). Exclusion of dyes, such as Hoechst -3221, by stem/progenitor cells is attributed to ABC transporter proteins (ATP-binding Cassette) (Zhou et al. 2001). Differentiated cells presumably lack these proteins and thus retain dyes appearing as bright cells on a side scatter plot. In contrast, stem/progenitor cells efflux the dye via ABC transporter proteins and thus appear as dim populations. The caveat to using SP is that the labeling dye can incorporate into DNA, and might compromise the reconstitution abilities of the labeled cells. Furthermore, the mammary repopulating cells found in recent studies seem to be negative for the SP phenotype (Shackeltosn et al. 2006). Thus the search for functional stem/progenitor cell markers remains an active area of research.

Recently, another marker has been added to the functional tool-kit, aldehyde dehydrogenase (ALDH) (Ginestier et al. 2007, Ibarra et al. 2007). This is an enzyme that converts reactive aldehydes to the less toxic carboxylic acid. In HSC and neural stem

cells, ALDH expression was shown to be elevated and responsible for the resistance of HSC to cyclophosphamide (Hess et al. 2006) (Corti et al. 2006). To detect ALDH activity, a substrate of ALDH has been conjugated to a BIODPSY fluorophore (commercially available by Stem Cell Technology), which can be converted into a negatively charged product by the enzyme. This conversion ensures the fluorophore is retained in the cell and thus detectable by FACS. In contrast, cells lacking ALDH activity leads to an efflux of the substrate and thus reduced fluorescence. Interestingly, ALDH positive cells derived from AML patients have increased NOD/SCID engraftment potential relative to ALDH negative cells, suggesting that these cells represent primitive leukemic stem cells (Cheung et al. 2007).

Linking tumor-initiating cells to stem/progenitor cells necessitates the identification of a common stem cell marker that distinguishes their- self-renewing potential from their differentiated counterparts. Thus the first-steps of my thesis work, described in Chapter 2, sought to define markers and assays for the elucidation of self-renewing populations within the mammary gland. In this search we realized that microRNAs could potentially serve the dual purpose of both markers and biological significance as regulators of self-renewal.

1.2 MicroRNAs: Biology and Application

The discovery of small RNAs has revolutionized our understanding of how genes are post-transcriptionally regulated, adding to an already complicated regulatory cellular network. MicroRNAs (miRNAs) are a subclass of small RNA molecules typically 21-22 nucleotides in length, and are found in all metazoa studied to date (Reviewed by (Bartel 2004), with phylogenetic studies demonstrating high evolutionary conservation from worms to humans (Pasquinelli et al. 2003). Approximately, 700 miRNAs have been reported in mammals, and with the advent of deep sequencing technology, we are beginning to find less abundant but more cell-type specific miRNAs.

MicroRNA genes can be found in genomic clusters or as scattered singletons. Knowledge of their genomic location provides insights into upstream transcriptional regulators or epigenetic marks that regulate their expression. Approximately, 40% of miRNA loci are found within intronic regions of non-coding transcripts, with 10% in exonic regions. Of those located in introns of protein-coding genes, many have been hypothesized to use the promoter of the surrounding gene (Kim 2005). In some instances, miRNAs are clustered and transcribed as a single polycistronic unit (He et al. 2005), although, they are not necessarily expressed at the same levels.

1.2.1 miRNA Biogenesis

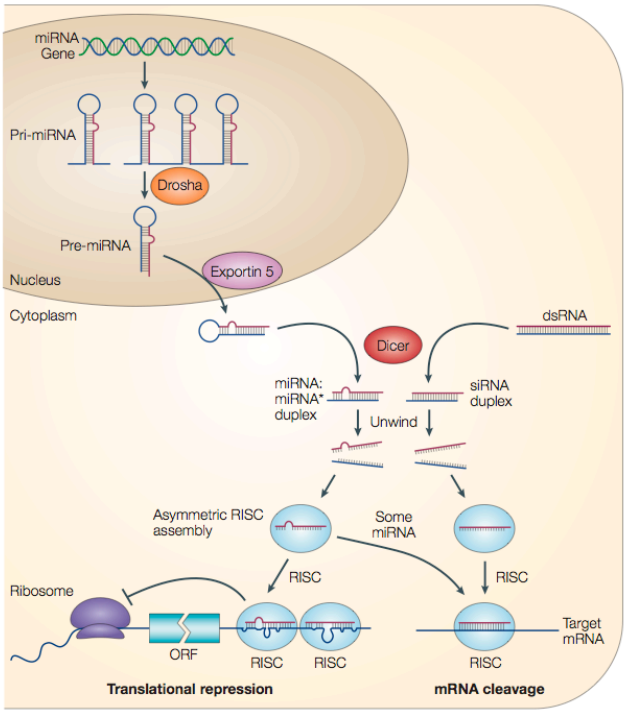


Figure 1. 5 The current model for the biogenesis and post-transcriptional suppression of microRNAs and small interfering RNAs. The nascent pri-microRNA (pri-miRNA) transcripts are first processed into approx70-nucleotide pre-miRNAs by Drosha inside the nucleus. Pre-miRNAs are transported to the cytoplasm by Exportin 5 and are processed into miRNA:miRNA* duplexes by Dicer. Dicer also processes long dsRNA molecules into small interfering RNA (siRNA) duplexes. Only one strand of the miRNA:miRNA* duplex or the siRNA duplex is preferentially assembled into the RNA-induced silencing complex (RISC), which subsequently acts on its target by translational repression or mRNA cleavage, depending, at least in part, on the level of complementarity between the small RNA and its target. ORF, open reading frame. Adapted from (He and Hannon 2004)

The biogenesis of miRNAs is a multi-step process and mutations of key RNAi components have been shown to affect stem cell maintenance and differentiation. Work by (Lee et al. 2004) and (Cai et al. 2004) has shown that most primary miRNA transcripts are driven by RNA polymerase II (Pol II), though (Borchert et al. 2006) have found a subset of miRNA associated with Alu repeats that are transcribed by RNA Pol III. Primary transcripts vary in length but are capped and polyadenylated similar to mRNA transcripts. Rnasen (also known as Drosha), an RNase III-type protein enzyme, and

DiGeorge syndrome critical region 8 protein (DCGR8) form a microprocessor complex that recognizes the single-stranded RNA (ssRNA) segments at the base of the hairpin (Lee et al. 2003). This leads to the cleaving of the primary transcript at exactly 11 base pairs from the ssRNA hairpin base to generate a 70-nt precursor molecule (pre-miRNA) (Lee et al. 2003), (Denli et al. 2004), (Gregory et al. 2004) and (Han et al. 2004)). Wang et al. demonstrated that DCGR8 deficient embryonic stem cells (ESC) do not fully down-regulate pluripotency factors in the presence of a strong inducer of differentiation, retinoic acid, suggesting that miRNAs are essential for driving differentiation programs (Wang et al. 2007). The precursor miRNA is transported out of the nucleus and into the cytoplasm by exportin-5 (Lund et al. 2004), (Yi et al. 2005), (Bohnsack et al. 2004). Dicer, another RNase III enzyme, is the next key component in the maturation of miRNAs (Bernstein et al. 2001). It recognizes the terminal loop of the hairpin structure and cleaves 22 -nucleotides from the pre-existing terminus generated by Rnasen, resulting in ~21 nucleotide RNA duplex. Dicer-mutant mice die early in development with a loss of Oct4-positive multipotent stem cells (Bernstein et al. 2003). Interestingly, In *Drosophila* ovaries, *dcr-1* mutant germ-line stem cells are depleted within 3 weeks of *Dicer* loss (Jin et al. 2007).

Strand selection occurs during the final step of miRNA biogenesis to generate the mature single-strand miRNA, and is dictated in part by their thermodynamic stability at the ends of the duplex (Schwarz et al. 2003) and (Khvorova et al. 2003). In mammalian cells, miRNAs are incorporated into miRNA-induced silencing complexes (miRISCs), with the multi-domain Argonaute protein as the key effector protein. The mode of miRNA-mediated regulation is dependent on the degree of complementarity between

miRNA:mRNA duplexes. Translational repression (Olsen and Ambros 1999) and (Doench et al. 2003) of target mRNA occurs if there are mismatches of the miRNA (Grimson et al. 2007). Destabilization of mRNA transcripts (Hutvagner and Zamore 2002), Martinez et al. 2002 and (Zeng et al. 2002)) is thought to result from perfect complementation between miRNA and the target mRNA substrate. In most instances, the regulatory effectors (consisting of Argonaute proteins and miRNAs) bind to the 3'UTR of protein-coding genes, though recent data suggests that 5'UTR regions of genes can also be a target site (Orom et al. 2008).

1.2.2 Application of MiR-Biology

MiRNA-sensor technology takes advantage of the binding requirement mentioned above. In plants miR-sensors were first developed to study the dynamic spatiotemporal expression patterns of miRNAs during plant development. In plants, miR-sensors take advantage of the perfect complementarity required between miRNA:mRNA duplex for recognition and cleavage. Thus perfect miRNA binding sites are placed in the 3'UTR of a reporter gene allowing for detection of miRNA activity. In zebrafish, miRNA sensors were utilized to study the substrates requirements for let-7 function in the developing zebrafish embryo (Kloosterman et al. 2004). In Drosophila, miR-sensors where used to study the expression patterns of bantam (Brennecke et al. 2004). The use of miR- sensors for tracking cell populations *in vitro* and *in vivo* are described in detail in chapter four.

1.3 Cancer Stem Cell Theory

My thesis work has largely been motivated by the potential relationship between stem cells and cells within a tumor mass with stem-like properties. The concept of stem-cell origin of cancer is a 30 years old hypothesis that has gained experimental proof only within the last 13 years. The concept that haematological malignancies have a stem cell component that is responsible for their aberrant self-renewal properties arose originally from studies on acute myeloid leukemia (AML) in which it was found that a small subpopulation of the leukaemic cells was capable of initiating the leukaemia when transplanted into immunocompromised mice (Bonnet and Dick 1997). Subsequently, evidence emerged for a similar stem cell-like component in other haematological malignancies, including chronic myeloid leukemia (CML) (Jamieson et al. 2004), acute lymphocytic leukaemia T-ALL (Cox et al. 2004) and myeloma (Matsui et al. 2004). It was earlier experiments, however, by Philip Fialkow and colleagues, using patterns of inactivation of X-linked genes, that demonstrated leukaemia, CML and AML, were clonal in origin (Fialkow et al. 1967), (Fialkow et al. 1981). Further analysis of X-inactivation patterns in various haematopoietic lineages revealed that early stem and progenitor cells were suspect in the development of CML and AML (Fialkow et al. 1979), thus planting the seed for stem cells as originators and propagators of disease. Since these studies, attempts have been made to isolate the tumorigenic population in solid tumors from tissues such as breast, brain, colon, pancreas, (Al-Hajj et al. 2003), (Singh et al. 2003), (Barker et al. 2009).

In breast cancer, the ‘cancer stem cell’ was defined by the expression of CD-44 and CD-24. Using fluorescent-sorted populations of dissociated primary breast tumors

(based on CD-44 and CD-24 expression), it was demonstrated the tumorigenic potential resided within the CD24 negative/low subset. In addition, these tumors recapitulated the same heterogeneous phenotype in the primary tumor, demonstration of their self-renewal potential. The tumor-initiating cells (and those with metastatic potential) have been further refined from primary breast tumors and established cell lines based on ALDH (Aldehyde Dehydrogenase) activity (Charafe-Jauffret et al. 2009), (Ginestier et al. 2007).

Pathologists have long appreciated the heterogeneity of breast tumors (du Toit et al. 1989), (Elston et al. 1982). However, the advent of DNA microarray technology in early 2000 verbalized these pathologically distinct specimens and defined the vocabulary for the classification of breast cancer today (Sorlie et al. 2001 and Perou et al. 2001). Both pathological and molecular profiles aim to categorize breast cancers and correlate their phenotypes to clinical prognosis. Approximately 80% of patients fall into a luminal category. In terms of gene expression, the luminal signature is characterized by over represented genes involved in estrogen receptor signaling and fatty acid metabolism. These patients tend to respond well to current chemotherapy (Sorlie et al. 2001 and Perou et al. 2001). In contrast, basal tumors, affecting the other 20%, are characterized by expression of cytokeratin-5, 6, 17, and are associated with poor clinical outcome.

The emerging picture is of one in which tumors have a hierarchal organization with respect to self-renewal and differentiation potential similar to their normal tissue counterparts. With respect to breast cancer the cancer stem cell theory aims to explain the different types of breast cancers and the heterogenous composition of tumor cells. Luminal breast cancer arises from mutations in luminal progenitors and basal cancers arise from mutations in basal progenitors. Because there is an enticing congruence

between normal and malignant stem-cell function, it is probable that if we understand one, we will have major insights into the other.

1.4 Summary

The mammary epithelium undergoes massive tissue remodeling during normal development. Evidence for the existence of stem cells within the epithelium I think is clear. However, where these cells reside and how they respond to environmental cues during development remains unknown. Markers that resolve stem and progenitors cells within the mammary gland will shed light on the general mechanism of mammary development. Furthermore, we know that miRNAs in plants target a majority of transcription factors that aide in the specification of cell-fates. In mammals, miRs also function in a tissue-specific manner, however, the expression patterns of miRNAs in normal mammary development is an understudied field. The cancer stem cell hypothesis is an interesting theory and highly controversial. Whether tumor-initiating cells are stem cells remains a therapeutically significant question. We know that solid tumors are composed of cells with differential tumorigenic potential and that differences in this potential may be key to targeted therapeutics.

CHAPTER 2

A role for microRNAs in the maintenance of mouse mammary epithelial progenitor cells

Ingrid Ibarra[1,2], Yaniv Erlich[1], Senthil K. Muthuswamy[1], and Gregory J. Hannon[1*]

[1]Watson School of Biological Sciences
Howard Hughes Medical Institute
Cold Spring Harbor Laboratory
1 Bungtown Road
Cold Spring Harbor, NY 11724

[2]Program in Genetics
Stony Brook University
Stony Brook, New York 11794, USA
* To whom correspondence should be addressed: hannon@cshl.edu

microRNA expression profiles are often characteristic of specific cell-types. The mouse mammary epithelial cell line, Comma-Dβ, contains a population of self-renewing progenitor cells, which can reconstitute the mammary gland. We have purified this population and determined its microRNA signature. Several microRNAs, including miR-205 and miR-22 are highly expressed in mammary progenitor cells, while others, including let-7 and miR-93, are depleted. Let-7 sensors can be used to prospectively enrich self-renewing populations, and enforced let-7 expression induces loss of self-renewing cells from mixed cultures.

Introduction

MicroRNAs (miRNAs) hone patterns of gene expression by programming the RNAi machinery to regulate complementary targets. While miRNAs have been integrated into numerous biological circuits, seminal studies in *C. elegans* implicated these small RNAs in regulation of differentiation and development (Reinhart et al. 2000). In this regard, some miRNAs are characteristic of particular differentiated cell types (Lagos-Quintana et al. 2002), while others are specifically expressed during early development and in stem and progenitor cells (Houbaviy et al. 2003; Giraldez et al. 2006; Nielson et al. 2007).

Evidence for a role of miRNAs in stem cell maintenance and differentiation is accumulating from analysis of mutations in key RNAi components. For example, Dicer-mutant mice die early in development with a loss of Oct4-positive multipotent stem cells (Bernstein et al. 2003). Even in the presence of a strong differentiation inducer, *DGCR8/pasha* knock-out ES cells fail to inactivate self-renewal programs (Wang et al. 2007). In *Drosophila* ovaries, *dcr-1* mutant germ-line stem cells are depleted within 3 weeks of *Dicer* loss (Jin et al. 2007), and homozygous mutation of *loqs*, an obligate Dicer partner, causes defects in egg chamber development (Forstemann et al. 2005; Jiang et al. 2005).

In many tissues, stem and progenitor cell populations are becoming increasingly well defined. In the mammary gland, this was elegantly demonstrated by the

reconstitution of a functional gland from a single stem cell, which was isolated using cell surface markers, CD49f, CD29 and CD24 (Shackelton et al. 2006, Stingl et al. 2006). Recently, ALDH (aldehyde dehydrogenase) has been used to isolate and characterize normal human mammary stem cells (Ginestier et al. 2007). Hematopoietic stem cells and neuronal progenitor cells have also been isolated on the basis of ALDH activity (Hess et al. 2006, (Corti et al. 2006) Interestingly, ALDH positive cells derived from AML patients have increased NOD/SCID engraftment potential relative to ALDH negative cells, suggesting that these cells represent primitive leukemic stem cells (Cheung et al. 2007). ALDH-positive cells from human breast tumors were also shown to be the tumorigenic population, with as few as 500 cells being capable of initiating disease (Ginestier et al. 2007).

Comma-Dβ cells harbor a permanent population of undifferentiated basal cells that are able to reconstitute the mammary tree (Deugnier et al. 2006), thus providing an excellent system in which to study the role of miRNAs in stem cell maintenance, self-renewal and differentiation. By combining ALDH and Sca-1 (stem cell antigen) expression criteria, we performed an unbiased characterization of miRNAs in mammary progenitor populations using deep sequencing. These studies identified miRNAs that are highly expressed in the progenitor fraction as well as miRNAs that are relatively depleted in this population. By manipulating expression of at least one of these miRNAs, we link miRNAs to progenitor self-renewal.

Results and Discussion

ALDH is a marker of mammary progenitor cells

Sca-1^{high} Comma-Dβ cells have retained the ability to reconstitute a functional mammary gland upon transplantation of as few as 1000 cells into the fat pad of a syngeneic virgin female (Deugnier et al. 2006). 2-D and 3D cultures, including mammosphere assays, have provided evidence of the self-renewal and differentiation capacity of these cells as they can generate both myoepithelial and luminal cells *in vitro* Since Sca-1 expression was not enriched in the recently defined murine mammary stem/progenitor cells *(*Shackleton et al. 2006*)*, we asked whether ALDH expression could be used to isolate progenitor populations from Comma-Dβ. We also tested whether a combination of ALDH and Sca-1 markers provided increased specificity for progenitor cells, at least in cultured populations.

ALDH activity can be measured in living cells by using a fluorogenic substrate, ALDEFLUOR (Corti et al. 2006, Hess et al. 2006, Ginestier et al. 2007) ALDH induces retention of this substrate, resulting in increased florescence. Truly positive cells can be identified by comparison to cells cultured in ALDEFLUOR in the presence of DEAB, an ALDH inhibitor. The Comma-Dβ cell line contains ALDH^{bright}Sca-1^{high} cells that comprise ~2% of the total population (Fig. 1A). This number is consistent with the number of side population (SP) cells we observed in this cell line (Supplementary Fig. 1).

Colony formation on irradiated feeders or Matrigel is commonly used to assess the proliferative capacity of purified epithelial stem and progenitor cells. In several studies, this capacity has been shown to correlate with *in vivo* morphogenic potential (Shackelton et al. 2006, Deugnier et al. 2006). We therefore examined the colony forming capacity of four sorted populations ($ALDH^{bright}Sca\text{-}1^{high}$, $ALDH^{bright}Sca\text{-}1^{neg}$, $ALDH^{neg}Sca\text{-}1^{high}$, and $ALDH^{neg}Sca\text{-}1^{neg}$). Only the two $ALDH^{bright}$ populations yielded significant numbers of colonies, with the $ALDH^{bright}Sca\text{-}1^{high}$ subset exhibiting a 3-fold greater colony-forming frequency and substantially larger colonies (Fig.1B). $ALDH^{bright}$ cells gave rise to both luminal and myoepithelial colonies, based on morphology (Fig. 1C). A third colony morphology was also observed that fit neither the dispersed tear-drop shape characteristic of myoepithelial cells nor the tightly arranged cells with distinct cell borders that indicate luminal cells (Stingl et al. 1998).

$ALDH^{bright}Sca\text{-}1^{high}$ cells plated at clonogenic density in Matrigel expanded and formed spheroids (avg. 46/well n=4) ($p<0.001$), whereas the $ALDH^{bright}Sca\text{-}1^{neg}$ cells grew poorly under these conditions (Fig. 1D). $ALDH^{neg}$ cells were unable to form colonies. These results are consistent with previous studies showing the inability of $Sca\text{-}1^{neg}$ cells to grow in Matrigel (Deugnier et al. 2006).

Resistance to a group of anticancer drugs called oxazaphosphorines has been linked to ALDH activity (Bunting et al. 1996). We therefore reasoned that mafosfamide (MAF) treatment might enrich the population of $ALDH^{bright}$ progenitor cells. We treated cells for four days and analyzed the surviving population by FACS. This resulted in a 15-

fold enrichment in ALDH^{bright}Sca-1^{high} cells. Thus, the progenitor population resident within Comma-Dβ can be selected by this method, and these progenitors are intrinsically resistant to at least some anti-cancer drugs (Fig.1E,F). Following selection, we also noted a 2-fold expansion in the ALDH^{neg}Sca-1^{hi} compartment. We could not distinguish between resistance of this population to MAF and its arising by differentiation of selected ALDH^{bright}Sca-1^{high} cells.

MAF is a cyclophosphamide derivative that is active in cultured cells. Cyclophosphamide is commonly used as part of a first-line therapy for breast cancer (Smith et al. 2003). Thus, the fact that treatment with MAF can enrich ALDH-positive cells has profound implications, since recent data shows that tumor-initiating cells can be isolated from human breast tumors on the basis of ALDH activity (Ginestier et al. 2007).

A microRNA fingerprint of mouse mammary epithelial progenitor cells

To probe potential roles for miRNAs in the maintenance and differentiation of mammary epithelial progenitor cells, we constructed small RNA libraries from Sca-1^{high}, Sca-1^{neg}, ALDH^{bright}Sca-1^{high}, and MAF-treated Comma-Dβ cells. These were deeply sequenced on the Illumina 1G platform and mapped to the mouse genome using a customized bioinformatics pipeline. Reads were annotated by BLAT (Kent, 2002) to a unified database comprised of mouse entries from miRbase (Griffiths-Jones et al. 2006), NONCODE (Liu et al. 2005), tRNAs in "The RNA Modification Database" (Limbach et al. 1994), and rRNA entries in the Entrez Nucleotide Database.

Approximately 50% of all sequences that mapped to the genome corresponded to known miRNAs (Supplementary Table 1) for Sca-1^{high}, ALDH^{bright}Sca-1^{high} and Sca-1^{neg} libraries. In the MAF library, 80% of reads mapped to miRNAs. Breakdown products of noncoding RNAs such as rRNAs, tRNAs, snRNAs, snoRNAs, and others represented less than .5% of total sequences for all four libraries. An estimated 25% of sequences mapped neither to known miRNAs nor other annotated small RNAs in the sorted libraries whereas only 5% remained unidentified for the MAF library. The top 50 miRNAs sorted based on abundance in the ALDH^{bright}Sca-1^{high} library are shown in Supplementary Table 2.

Expression signatures are often presented as heat maps, illustrating the relative signal for an individual species in two samples. Although there are undoubtedly biases in the cloning of specific RNAs, the available sequence data permitted us to examine both differential expression and approximate abundance. We reasoned that focusing on highly expressed miRNAs would maximize the possibility of identifying those that are biologically relevant. A "bubble plot" can be used to depict both the abundance of a particular miRNA (given as the sum of the reads in the two libraries) and its relative expression (plotted as a $\log_2$ of the ratio of reads in each library).

The ALDH^{bright}Sca-1^{high} (Fig. 2A.) and the MAF libraries (Fig. 2B) were compared to the Sca-1^{neg} library to identify differentially expressed miRNAs. Two abundant miRNAs, miR-205 and miR-22, were consistently enriched in the progenitor population. Both were also abundant in Sca-1^{hi} library, suggesting that they may be important for the basic physiology or identity of basal cells. MiRNA expression profiling

of various tissues showed that miR-205 was preferentially expressed in breast and thymus (Baskerville et al. 2005). In human embryonic stem cells, Nanog and Sox2 binding sites are located near the miR-205 and miR-22 promoters (Boyer et al. 2005). However, in comparing our dataset to ES cell-specific miRNAs no consistent overlap in patterns was found.

Other miRNAs showed substantially lower expression in the progenitor compartment. Let-7b, let-7c and miR-93 were the most abundant miRNAs that showed preferential expression in Sca-1^{neg} cells. Collectively, let-7b and let-7c represented only 8.8% of the total miRNA sequences in the ALDH^{bright}Sca-1^{high} library compared to 24% of miRNA sequences in Sca-1^{neg} cells. Interestingly, miR-20a is part of a polycistronic cluster containing 17-5p, miR-18a, miR-19b. These are also underrepresented in the progenitor compartment. miR-21 was the most abundant miRNA found in relatively equal amounts in all four libraries constituting a consistent average of 30% of mapped miRNAs sequences.

Overall, the trends in miRNA representation seen upon comparison of ALDH^{bright}Sca-1^{high} to Sca-1^{neg} cells were reproduced upon examination of the MAF-treated library. However, miR-200a and miR-429, both of which are part of miR-8 family, were found at substantial levels in the MAF library only.

We performed an independent verification of differential miRNA expression using quantitative stem-loop PCR (qRT-PCR) as previously described (Chen et al. 2005) (Fig. 2C). We examined expression of let-7b, let-7c, miR-93, miR-23b, miR-23a, miR-205, miR-31 in the Sca-1^{hi} fraction vs Sca-1^{neg} libraries. In all 7 cases, differential

expression was confirmed, though the absolute magnitudes of expression ratios did not precisely agree with those determined from sequencing data.

let-7 depletes the self-renewing ALDH compartment in Comma-Dβ

We sought to investigate a role for reduced let-7 expression in mammary progenitor cells. However, it was first necessary to investigate whether the ALDH^{br}Sca-1^{hi} compartment was receptive to signals known to expand stem/progenitor populations and whether miRNA expression patterns responded similarly. Enforced expression of β-catenin in Comma-Dβ cells was shown to expand the Sca-1^{hi} compartment and increase mammosphere-forming capacity (Chen et al. 2007). Similarly, we observed a 3.5-fold increase in the ALDH^{bright}Sca-1^{high} population upon the ectopic expression of Wnt-1 (Fig. 3A). Wnt-1-expressing cells survived higher doses of MAF than empty vector control cells (Fig. 3B), consistent with the ALDH^{bright}Sca-1^{high} compartment having intrinsically higher drug resistance.

In concert with changes in the progenitor compartment, we observed that Wnt-1-expressing cells expressed 6-fold higher levels of miR-205 when compared to empty vector control cells with no observed reduction in let-7b, let-7c, or miR-93 expression, as might be expected since the differentiated compartments were still prominent in this mixed population (data not shown).

To probe the functional relevance of differential miRNA expression patterns, we examined the consequences of enforced expression of let-7c. Comma-Dβ let-7c cells showed a substantial, 6-fold reduction in the ALDH^{bright} compartment (n=4). In concert, we observed the emergence of distinctly Sca-1^{neg} and Sca-1^{lo} populations (Fig. 3C).

These results suggest that differences in miRNA expression between differentiated and self-renewing populations within Comma-Dβ cells have substantial impacts on cell identity and physiology.

A let-7 sensor marks the progenitor compartment

Convenient markers for rare cell populations have proven difficult to identify. miRNA sensors have been used in plants and animals to visualize the expression patterns of individual small RNA species. We wished to investigate the general possibility that miRNA sensors, as directed by our observed expression patterns, could be used to mark rare cell populations and permit their isolation. We constructed a let-7c sensor by introducing its perfect complement into the 3' untranslated region of DsRed, thus specifying silencing by RNAi in the presence of the miRNA (Fig. 4A). (Charafe-Jauffret et al. 2008)

Since let-7c expression is low in $ALDH^{bright}Sca\text{-}1^{high}$ cells, we predicted that the sensor would not be silenced, thus marking the progenitor compartment by DsRed expression. Where let-7c expression is high in the more differentiated cell types, we predicted that the sensor would be silenced (Fig. 4B). Indeed we found that overall, DsRed-positive cells (DsR+) constituted ~.8% of the population (Fig. 4C). DsRed-positive cells are enriched for $Sca\text{-}1^{high}$ and ALDH expressing cells, as expected (data not shown).

We tested DsRed+ cells for their ability to self-renew and differentiate *in vitro*. DsR+ cells formed spheroids with 10-fold greater efficacy than DsR- cells (Fig.4D), with only DsR+ cells forming spheroids greater than 50μm in size (Fig. 4E). Confocal images of spheroids co-stained with Keratin 5 (K5) and Keratin 18 (K18) revealed that a single DsR+ cell was able to give rise to a K5-positive, basal, outer layer and an inner layer of luminal, K18-positive cells (Fig. 4F), though, consistent with previous observations, not all spheres had such an apparent luminal structure (Deugnier et al. 2006). To probe the ability of DsR+ cells to differentiate into myoepithelial cells we co-stained spheroids with K5 and smooth muscle actin (SMA) and indeed observed spheroids with an outer layer of K5- and SMA-positive cells (Fig. 4G).

These studies demonstrate that a lack of let-7c expression, can be used to prospectively isolate mammary progenitor cells. Perhaps in combination with additional sensors, this allows the experimentally determined miRNA expression signature to be converted into a functional tool that can augment existing markers of murine progenitors and likely also tumor initiating cells.

Overall, our results support the notion that miRNA expression patterns form both a characteristic signature of a given cell type and help to reinforce cell fate specification. Even within a single cell line, distinct compartments containing progenitor cells and more differentiated cells have unique miRNA patterns, suggesting that such signatures can be used not only to define and track rare cell populations *in vitro* and *in vivo* but that

manipulation of these signatures might be used to expand or deplete stem cell and tumor initiating cell populations for therapeutic benefit.

Acknowledgements

We are grateful to Pamela Moody, Frances Kittrell, and Daniel Medina for help with Flow Cytometry and culturing Comma-Dβ cells. We thank Avi Rosenberg (CSHL), Max Wicha, Christopher Ginestier (Univ. Michigan), Susan Ludeman, and Mike Colvin for advice and for sharing protocols and reagents. We would also like to thank Monica Dus with help in confocal imaging. This work was supported in part by grants from the NIH (II, SM, GJH) and a kind gift from Kathryn W. Davis (GJH). YE is a Goldberg-Lindsay Fellow of the Watson School of Biological Sciences.

2.3 Methods

Cell culture

Comma-Dβ cells were grown in DMEM:F12 (HyClone) supplemented with 2% FCS, 5ng/ml murine EGF(Sigma), 10μg/ml human insulin (Sigma), and 50 μg/ml gentamicin (Gibco). Cells were only used within passages 17-35. Phoenix cells were maintained in DMEM supplemented with 10% FBS (Hyclone), and penicillin-streptomycin (Gibco).

Constructs and Infections

For construction of Let7c stable expression vectors, the following primers were used: forward-5' GGC CAG ATC TGT GTG GTC AAG GAG ATG TTA G-3' and reverse 5'GAT CCT CGA GTA ACA GCC CGT GAG AAA TAG-3' containing Bgl-II/Xho1 restriction sites. A 500bp fragment was PCR amplified from mouse genomic DNA and cloned into an MSCV vector carrying a hygromycin cassette (Clontech). Phoenix cell transfections were performed using LT-1 transfection reagent (Mirus) according to manufacture's instructions. To construct Wnt-1MSCV, the 1.9kb fragment of wingless cDNA (nucleotides 284-2181) in pMV7 (kind gift of Anthony Brown) was subcloned into an MSCV-hygro vector. For construction of the Let7c sensor, miRNA-complementary oligonucleotides were annealed and cloned into a Marx vector that directs dsRED expression (a kind gift Scott Hammond).

ALDEFLUOR and SP cell staining and Flow Cytometry

Cells were stained at 10^6 cells/ml in assay buffer containing 1μmol BAAA for 1hr at 37C. The ALDEFLUOR kit was purchased from StemCell Technologies, Durham,NC, USA. An aliquot of stained cells were treated with 50mmol/L DEAB as a negative control. After ALDEFLUOR staining, cells were co-stained with anti-Sca-1-PE (BD Pharmigen) for 20 minutes on ice. For small RNA cloning cells were FACS sorted directly into Trizol LS reagent (Invitrogen). ALDEFLUOR was excited at 488 nm, and fluorescence emission was detected using a standard fluorescein isothiocyanate (FITC) 530/30-nm band-pass filter. For SP analysis, cells were stained with Hoescht $_{33342}$ Dye as previously described (Goodell et al. 1996).

In vitro assays

Colony formation assays on feeders were essentially performed as described by Shackleton et al 2006. Three dimensional (3D) cultures were performed as described in (Debnath et al 2003).

Antibodies and Immunofluorescence

The following primary antibodies were used: anti-Sca-1 PE (BD pharmigen); mouse anti-cytokertain peptide 18 (Sigma), mouse anti-α –Sma (Sigma), rabbit polyclonal anti-cytokeratin 5 (Covance). Flurochrome-conjugated secondary antibodies included anti-rabbit IgG-Alexa$_{488}$ and anti-mouse IgG-Alexa$_{647}$ (Molecular Probes).

Small RNA cloning

1-4 ug of total RNA from sorted cells was used for small RNA cloning performed as described in Pfeffer et al. 2005, and a detailed protocol is available upon request. Illumina 1G sequencing and analysis was performed as described Stark et al. 2007.

miRNA expression analyses

mature miRNAs were quantified using the TaqMan MicroRNA Assays previously described by Chen et al. 2005 (Applied Biosystems). Data was normalized to Actin using SuperScript III SYBR Green One-Step qRT-PCR system (Invitrogen). The experiments were repeated twice and all reactions were run in triplicate.

Cell Viability Assay

To assess the cytotoxic effects of MAF cells were seeded at 5,000 cells per well in a 96-well format. 24h or 48h later cells were treated with various doses of mafosfamide L-lysine salt (D-17930) (NIOMECH in der IIT GmbH) freshly dissolved in water. Cell viability was measured using CellTiter-Glo® Luminescent Cell Viability Assay (PROMEGA).

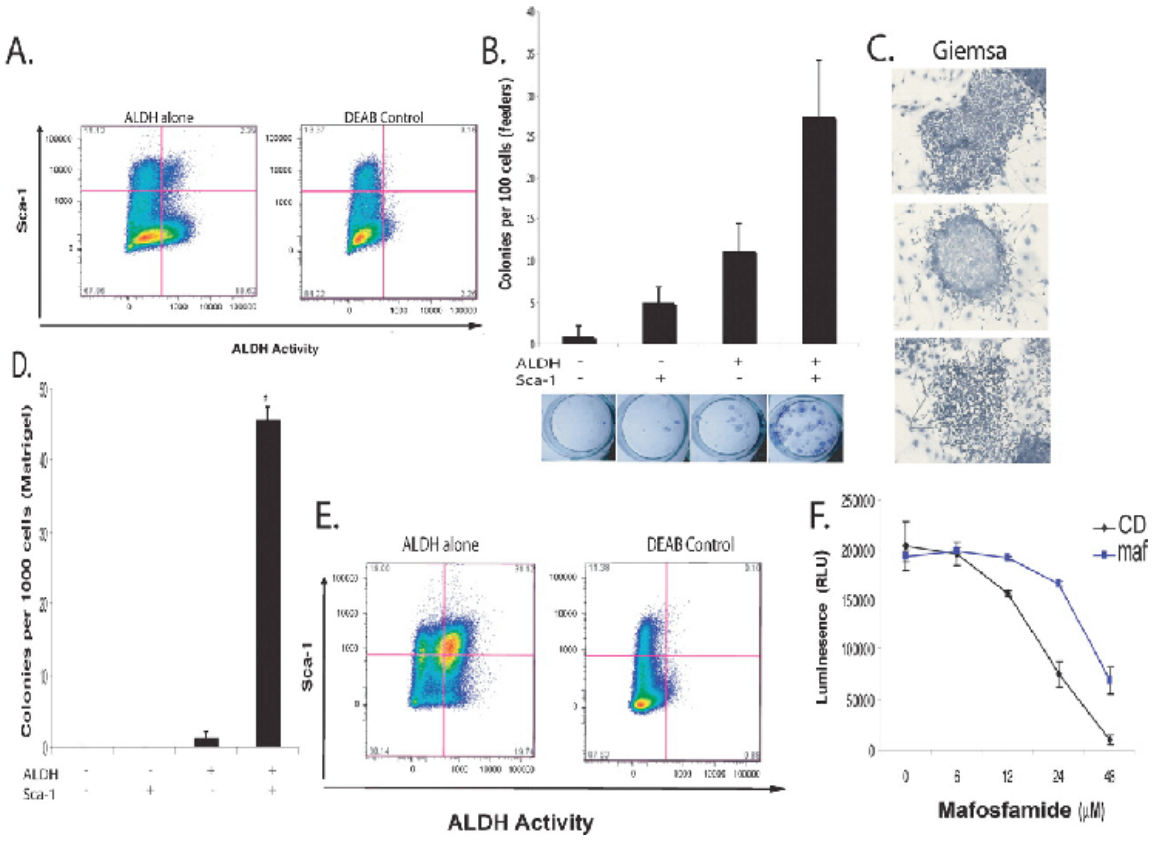


Figure 2.1 Characterization of ALDH as MESC marker in Comma-Dβgeo
(A) A FACS pseudo color dot plot showing ALDH activity and Sca-1 expression in Comma-Dβgeo cells. Cells incubated with ALDEFLOUR substrate (BAAA) and stained with Sca-1(left), cells stained with BAAA and co-stained with Sca-1 and incubated with ALDH inhibitor, DEAB, to establish background fluorescence (right). Shown are 100,000 events. (B) Histogram showing the colony-forming capacity of 4 sorted populations based on ALDH activity and Sca-1 expression seeded at clonal density on irradiated NIH3T3 feeders. Data represents the meanof four independent experiments. (C) Giemsa staining of $ALDH^{br}Sca^{hi}$ colonies grown on irradiated feeders for 6 days. Based on morphology myoepithelial (top), luminal (middle), and mixture (bottom) colonies were observed. (D) Histogram showing the colony-forming ability of 4 sorted populations embedded at clonal density in Matrigel (n=4). (E) Mafosfamide selects for $ALDH^{br}Sca\text{-}1^{hi}$ cells. Comma-Dβgeo cells were treated with a 6uM dose of MAF for 5-6 days and resultant population FACS analyzed. Cells resistant to Mafosfamide incubated with ALDEFLOUR and stained with Sca-1 (left), DEAB control for gating (right). (F) Cell viability assay after a 24hr treatment with various doses of mafosfamide of Comma-Dβgeo cells (black) and mafosfamide resistant cells (blue). Data represents the mean ± SD (error bar) of 2 independent experiments done in triplicate.

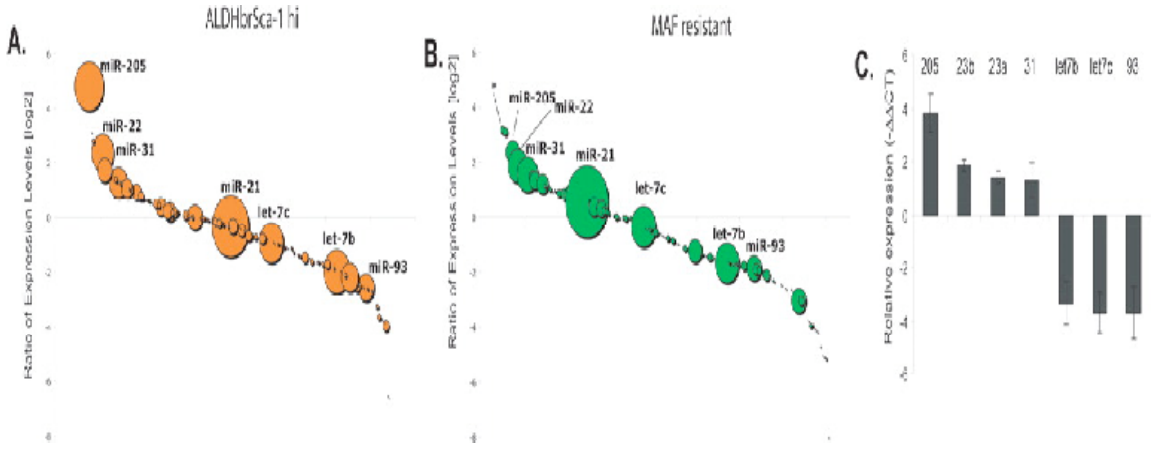


Figure2.2 MicroRNA profile of stem/progenitor compartment in Comma-Dβgeo
MicroRNAs are differentially expressed in self-renewing compartments. (A-B) Bubble-plots depicting the relative abundance and log2 ratio of ALDH^{br}Sca^{hi} mapped microRNAs (A) and mafosfamide resistant cells (B) relative to Sca-1^{negative} library. (C) Stem-loop quantitative (q) RT-PCR for the mature forms of selected differentially expressed microRNAs sequenced by Solexa. Shown are relative expression levels ΔΔCT of each microRNA from sorted Sca-1^{hi} and Sca-1 ^{negative} Comma-Dβgeo cells.

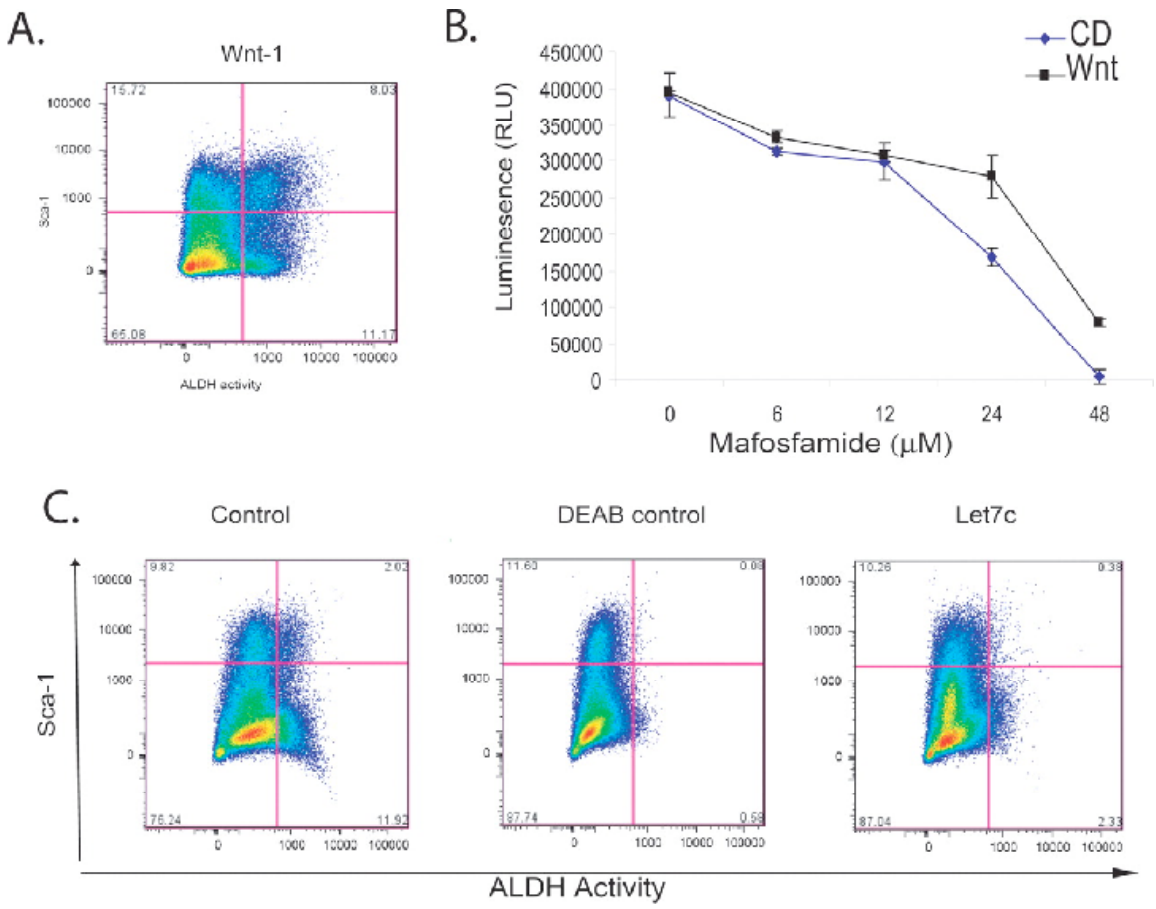


Figure 2.3 Let7c depletes the self-renewing compartment in Comma-Dβgeo

(A) Ectopic expression of Wnt expands the ALDH^{br} Sca-1^{hi} compartment. FACS plot of Comma-Dβgeo overexpressing Wnt-1 co-stained with ALDEFLOUR and Sca-1. (B) Cell viability assay using cell titer glo (promega) after a 48hr treatment with various doses of mafosfamide of Comma-Dβgeo cells (blue) and Wnt-1 expressing cells (black). Data represents the mean ± SD (error bar) of 2 independent experiments done in triplicate. (C) FACS profile of empty vector control Comma-Dβgeo cells co-stained with ALDEFLOUR and Sca-1(right), DEAB control for empty vector cells(middle), Comma-Dβgeo cells ectopically expressing Let7c(right) and also stained with ALDEFLOUR and Sca-1. A depletion of the ALDH compartment is observed upon introduction of Let7c.

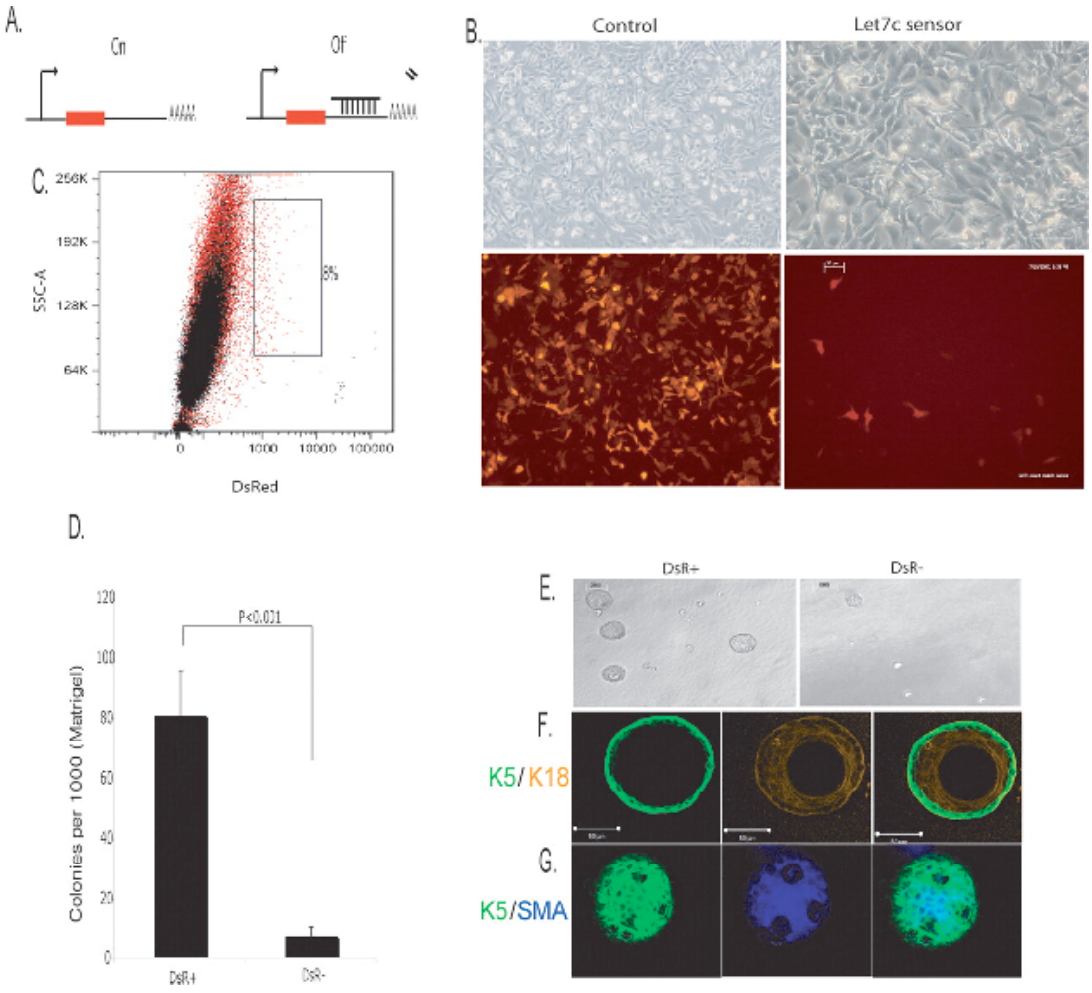
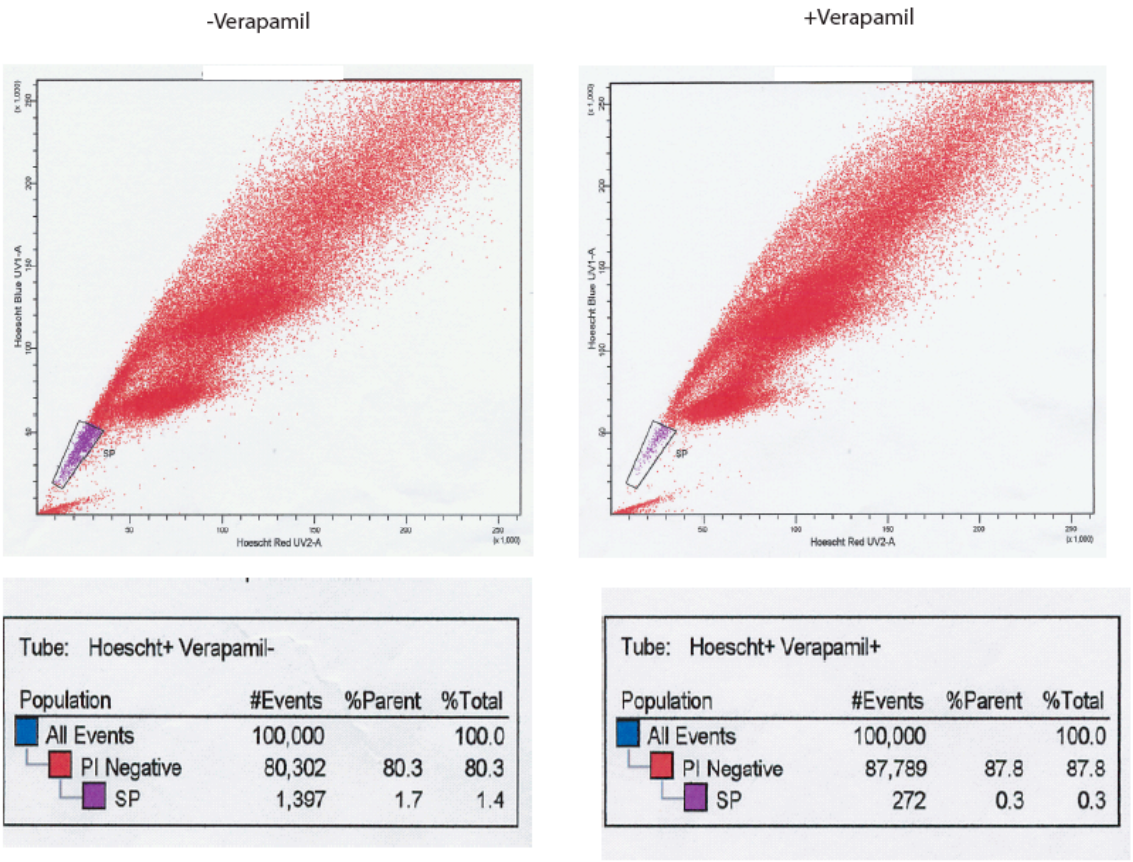


Figure 2.4 Self-renewal and differentiation of a single Let7c negative cell *in vitro.* (A) Cartoon depicting Let7c sensor construct. (B) Phase contrast images of Comma-Dβgeo cells expressing construct with no let7c binding sites (control) and Comma-Dβgeo cells expressing sensor construct containing Let7c complementary sites. (C) Overlay FACS dot plot of Let7c sensor cells (red) and uninfected Comma-Dβgeo cells (black) as an unstained control. DsR+ cells constitute .8% of the total population. (D) Histogram showing the colony-forming ability of DsR+ and DsR-cells embedded at clonal density in Matrigel (n=4). (E) Phase contrast images of resultant DsR+ and DsR- spheroids grown on Matrigel. DsR+ cells gave rise to substantially larger colonies greater than >50μm whereas DsR- cells never exceeded this size. (F-G) Confocal images of spheroids derived from DsR+ cells. (F) Representative cross section through the middle of a sphere co-stained with basal K5 and luminal K18 antibodies. (G) Representative image through the top of spheroid co-stained with basal K5 and α-Sma antibodies. DsR+ cells in 3-D culture give rise to both luminal K8+ and myoepithelial K5+ α-Sma cells.

Supplementary Figure



Supplementary Figure 1. Comma-Dβ cells contain a side-population

FACS profile of Comma-Dβ cells stained with Hoescht 33342 Dye and with fluorescence displayed at two wavelength emissions, blue (FL7) and red (FL8). (A) Cells incubated in the absence of ATP transporter inhibitor, verapamil and (B) cells stained and cultured in the presence of verapamil. As indicated by the FACS profile SP represents approximately 2% of total number of events collected.

Name	Condition		
	Sca-1 ^{Negative}	Sca ^{High}/ALDH ^{bright}	MAF
Total number of successful Solexa reads:	4,099,736	2,270,791	1,860,259
Biological Products:			
miRNA	2,2059,799 (54%)	1,067,613 (47%)	1,472,429 (79%)
mRNAlike	474,384 (12%)	255,991 (11%)	221,731 (12%)
tRNA	36,719 (0.9%)	11,849 (0.5%)	4,949 (0.3%)
piRNA	1,8825 (0.46%)	5,653 (0.25%)	4,886 (0.26%)
rRNA	1,250 (<0.1%)	1,109 (<0.1%)	282 (<0.1%)
snoRNA	504 (<0.1%)	192 (<0.1%)	140 (<0.1%)
snRNA	47 (<0.1)	58 (<0.1%)	0 (0%)
Other RNAs	2,646 (<0.1%)	420 (<0.1%)	454 (<0.1%)
Technical Artifacts:			
Adaptor self-ligation	134,510 (3.28%)	159,210 (7.01%)	17,444 (0.93%)
Spiked-in radio-labeled RNA marker	60,873 (1.48%)	295,279 (1.30%)	47,906 (2.57%)
Undefined:			
Undefined	1,163,999 (28.39%)	473,417 (20.84%)	90,038 (4.8%)

Supplementary Table 1: Distribution of sequencing results for each compartment.

All the successful Solexa reads were compared using BLAT [1] to a database that was comprised of mouse mature miRNA from miRBase[2], mouse non-coding RNA from NONCODE[3], mouse tRNA from [4].
Undefined represents the class of sequences that could not be annotated using this database.

[1] Kent, W.J. BLAT - The BLAST-Like Alignment Tool. Genome Res. 12(4), 656-664 (2002).
[2] miRBase: microRNA sequences, targets and gene nomenclature.
Griffiths-Jones S, Grocock RJ, van Dongen S, Bateman A, Enright AJ.
NAR, 2006, 34, Database Issue, D140-D144
[3] NONCODE: an integrated knowledge database of non-coding RNAs
Nucleic Acids Research, 2005, Vol. 33, Database issue D112-D115
[4] Limbach P.A., Crain P.F., McCloskey. J.A. 1994. Summary: the modified nucleosides of RNA. Nucleic Acids Res. **22**: 2183-2196.

Table 2.1 The 50 most abundant differentially expressed microRNAs

MiRNAs cloned from the four distinct libraries sorted by abundance in ALDH^{br}Sca^{hi} library. Data represents raw counts for each miRNA.

Supplementary Table 2. The 50 most abundant differentially expressed microRNAs cloned from the four distinct libraries sorted by abundance in ALDHbrScahi library. Data represents raw counts for each miRNA.

Name	Sca-	Sca+	Sca+/ALDH	MAF
mmu-miR-205	19863	245719	282099	68275
mmu-miR-21	481446	983326	194865	472852
mmu-miR-22	53768	177050	140987	131022
mmu-miR-31	70341	350889	86879	138207
mmu-let-7c	252067	151885	67186	134301
mmu-miR-29a	35707	79858	59601	36501
mmu-let-7b	350721	136707	44986	73731
mmu-miR-24	76273	194739	39141	76414
mmu-miR-29b	29736	69223	31442	45792
mmu-let-7a	42194	68022	25015	41828
mmu-let-7f	23513	58940	22726	40008
mmu-miR-130a	31343	32538	20878	36732
mmu-miR-143	169575	107243	18784	13747
mmu-let-7i	30523	27783	18424	25134
mmu-miR-20a	112710	98273	15711	31599
mmu-miR-103	36344	66014	14593	31678
mmu-miR-93	146002	90496	12521	26717
mmu-miR-16	6130	46814	10002	21188
mmu-let-7g	17060	27643	9857	15469
mmu-let-7d	23400	42094	8432	13661
mmu-miR-30a-5p	9121	13441	8238	17062
mmu-miR-26a	11669	17514	7253	14106
mmu-miR-10a	8238	5136	7205	8316
mmu-let-7e	9472	13042	6887	7454
mmu-miR-125b	21591	58415	6789	4243
mmu-miR-221	14678	37425	6499	3543
mmu-miR-320	11634	7985	6293	5242
mmu-miR-140*	19716	31507	5848	3780
mmu-miR-92	1219	7993	4265	2043
mmu-miR-99b	5934	7434	4180	4099
mmu-miR-30d	2963	5209	3966	4423
mmu-miR-210	8556	4564	3932	2592
mmu-miR-27b	21564	52185	3929	4537
mmu-miR-181a	5993	4373	3489	3961
mmu-miR-99a	1605	2413	3454	2588
mmu-miR-100	2547	3668	3055	3448
mmu-miR-27a	29974	53643	3051	4661
mmu-miR-652	19402	10406	2997	3359
mmu-miR-191	5599	6619	2978	6834
mmu-miR-23a	33020	171674	2936	6008
mmu-miR-200a	1987	1477	2691	11914
mmu-miR-674	7689	6755	2372	3005
mmu-miR-183	2800	7719	2296	3091
mmu-miR-218	2183	4141	2187	1877
mmu-miR-101b	3773	7742	1966	1841
mmu-miR-429	1788	1758	1893	10349
mmu-miR-23b	11674	104212	1739	3286
mmu-miR-125a	3172	10349	1698	827
mmu-miR-26b	5200	6697	1635	4009
mmu-miR-107	3264	6585	1635	2427
Total Number of Known miRNA*	2,469,404	3,980,114	1,274,811	1,676,774
Total Number of Successful Reads:	4,099,736	6,648,439	2,270,791	1,860,259
Total Number of Reads with known erro	4,783,145	6,844,356	2,433,920	2,336,839

*Using BLAT by Kent WJ. Parametes:
-minIdentity=90 -minScore=17 -tileSize=6 -minMatch=1
Database: All Mus Musculus entries in mature.fa from mirBase v9.2

Literature Cited

Baskerville,S., Bartel, D.P. Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes. 2005. *RNA* **11**:241-247.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchinson, E.P., Alcorn, H., Li, M.Z., Mills, A., Elledge, S.J., Anderson, K.V., Hannon, G.J. 2003. Dicer is essential for mouse development. *Nat. Genet.* **35**: 215–217.

Boyer, L.A., Lee, T.I., Cole, M.F., Johnstone, S.E., Levine, S.S., Zucker, J.P., Guenther, M.G., Kumar, R.M., Murray, H.L., Jenner, R.G., Gifford, D.K., Melton, D.A., Jaenisch, R., Young, R.A. 2005. Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122**:947-56.

Bunting, K.D., Townsend, A.J. 1996. De novo expression of transfected human class 1 aldehyde dehydrogenase (ALDH) causes resistance to oxazaphosphorine anti-cancer alkylating agents in hamster V79 cell lines. Elevated class 1 ALDH activity is closely correlated with reduction in DNA interstrand cross-linking and lethality. *J Biol Chem*. **27**:11884-11890.

Chen C, Ridzon, D.A., Broomer, A.J., Zhou, Z, Lee, D.H., Nguyen, J.T., Barbisin, M, Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., Guegler, K.J. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res*. **33**: e179.

Chen, M.S., Woodward, W.A., Behbod, F., Peddibhotla, S., Alfaro, M.P., Buchholz, T.A., Rosen, J.M. 2007. Wnt/beta-catenin mediates radiation resistance of Sca1+ progenitors in an immortalized mammary gland cell line. *J Cell Sci*. **120**:468-77.

Cheung, A.M., Wan, T.S., Leung, J.C., Chan, L.Y., Huang, H., Kwong, Y.L., Liang, R., Leung, A.Y. 2007. Aldehyde dehydrogenase activity in leukemic blasts defines a subgroup of acute myeloid leukemia with adverse prognosis and superior NOD/SCID engrafting potential. *Leukemia* **21**:1423-30.

Corti, S., Locatelli, F., Papadimitriou, D., Donadoni, C., Salani, S., Del Bo ,R., Strazzer, S., Bresolin, N., Comi, G.P. Identification of a primitive brain-derived neural stem cell population based on aldehyde dehydrogenase activity. 2006. *Stem* Cells **24**:975-85.

Debnath, J., Muthuswamy, S.K. & Brugge, J. S. 2003. Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures. *Methods* **30**: 256–268.

Deugnier, M.A., Faraldo, M.M., Teuliere, J., Thiery, J.P., Medina, D., Glukhova, M.A. 2006. Isolation of mouse mammary epithelial progenitor cells with basal characteristics from the Comma-Dbeta cell line. *Dev Biol*. **293**:414-25.

Förstemann, K., Tomari, Y., Du, T., Vagin, V.V., Denli, A.M., Bratu, D.P., Klattenhoff, C., Theurkauf, W.E., Zamore, P.D. Normal microRNA maturation and germ-line stem cell maintenance requires Loquacious, a double-stranded RNA-binding domain protein. 2005. *PLoS Biol.* **3**:236.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C., Liu, S., Schott, A., Hayes, D., Birnbaum ,D., Wicha, M.S., Dontu G. 2007. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* in press.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van, Dongen, S., Inoue, K., Enright ,A.J., Schier , A.F. 2006. Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. *Science* **312**:75-9.

Goodell, M.A., Brose, K., Paradis, G., Conner, A.S., Mulligan, R.C. 1996. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med.***183**:1797-806.

Griffiths-Jones, S., Grocock R.J., van Dongen S., Bateman A., Enright A.J. 2006. miRBase: microRNA sequences, targets and gene nomenclature. NAR, 34, Database Issue, D140-D144

Hess, D.A., Wirthlin, L., Craft, T.P., Herrbrich, P.E., Hohm, S.A., Lahey, R., Eades, W.C., Creer, M.H., Nolta, J.A. Selection based on CD133 and high aldehyde dehydrogenase activity isolates long-term reconstituting human hematopoietic stem cells. 2006. *Blood* **107**:2162-2169.

Houbaviy, H.B., Murray, M.F. and Sharp, P.A. 2003. Embryonic stem cell-specific microRNAs. *Dev. Cell* **5**: 351–358.

Jiang, F., Ye, X., Liu, X., Fincher, L., McKearin, D., Liu, Q. Dicer-1 and R3D1-L catalyze microRNA maturation in Drosophila. 2005. *Genes Dev.* **19**:1674-1679

Jin ,Z., Xie, T. Dcr-1 maintains Drosophila ovarian stem cells. 2007. *Curr Biol.* **17**:539-44.

Kent, W.J. BLAT -- The BLAST-Like Alignment Tool. 2002. Genome Research **4**: 656-664

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr. Biol.* **12**: 735–739. Limbach P.A., Crain P.F., McCloskey. J.A. 1994. Summary: the modified nucleosides of RNA. Nucleic Acids Res. **22**: 2183-2196.

Liu C., Bai B., Skogerbø G., Cai L., Deng W., Zhang Y., Bu D., Zhao Y., Chen R. 2005. NONCODE: an integrated knowledge database of non-coding RNAs. Nucleic Acids Res. **33** (Database issue):D112-5.

Neilson, J.R., Zheng, G.X., Burge, C.B., Sharp, P.A. Dynamic regulation of miRNA expression in ordered stages of cellular development. 2007. *Genes Dev.* **21**:578-89.

Pfeffer, S., Sewer, A., Lagos-Quintana, M., Sheridan, R., Sander, C., Grasser, F.A., van Dyk, L.F., Ho, C.K., Shuman, S., Chien, M. 2005. Identification of microRNAs of the herpesvirus family. *Nat. Methods* **2**: 269-276.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., Ruvkun, G. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. 2000. *Nature* **403**:901-6.

Shackleton, M., Vaillant, F., Simpson, K.J., Stingl, J., Smyth, G.K., Asselin-Labat, M.L., Wu, L., Lindeman, G.J., Visvader, J.E. 2006. Generation of a functional mammary gland from a single stem cell. Nature **439**:84-8.

Smith, R.E., Bryant, J., DeCillis, A., Anderson, S; National Surgical Adjuvant Breast and Bowel Project Experience. 2003. Acute Myeloid Leukemia and Myelodysplastic Syndrome After Doxorubicin-Cyclophosphamide Adjuvant Therapy for Operable Breast Cancer: The National Surgical Adjuvant Breast and Bowel Project Experience *Journal of Clinical Oncology* **21**:1195-1204

Stark, A., Kheradpour,P., Parts, L., Brennecke, J., Hodges, E., Hannon, G.J., Kellis, M. 2007. Systematic Discovery and characterization of fly microRNA using 12 Drosophilia genomes. *Genome Research* in Press

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

Stingl, J., Eaves, C.J., Kuusk, U., Emerman, J.T. 1998. Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. *Differentiation* **63**:201-213.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., Blelloch, R. 2007. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet*. **39**:380-385.

Chapter 3

The Application of Sensor Technology to Human Breast Cancer

The *in vitro* experiments and small RNA profiling analysis in this chapter were performed by Ingrid Ibarra. The *in vivo* experiments performed in NOD/SCID mice and described in figure 5 were conducted by Suling Liu and Christophe Ginestier in the Wicha Laboratory.

3.1 Introduction

The invasive and metastatic properties of cancer cells, either intrinsic or acquired, account for the deaths of 90% of cancer patients (Gupta and Massague 2006). Metastasis is generally viewed as a multi-step process involving migration, invasion, and colonization of tumor cells (Nguyen et al. 2009). These tumor cells seem to employ several strategies to achieve these steps. For example, morphological changes or transitions from an epithelial-state to a mesenchymal state (known as epithelial to mesenchymal transition, or EMT) allow cells to adopt a fibroblastic spindle morphology, which increases motility and thus mediates the first steps of metastasis (migration and invasion) (reviewed in (Thiery 2002), (Yang et al. 2004). In order to accomplish the final steps of metastasis (colonization) tumor cells are thought to revert from a mesenchymal to an epithelial state (mesenchymal to epithelial transition, or MET). The roles of miRNAs, such as the miR-200 family and miR-205, have recently been implicated as repressors of metastatic potential in tumor cells (Bracken et al. 2008) (Korpal et al. 2008) (Park et al. 2008) (Gregory et al. 2008).

In order to define a common signature between normal stem/progenitor cells with self-renewal capacity and tumor cells with tumorigenic and metastatic potential, we extended our miRNA analysis from chapter 2 to human breast cancer-derived cell lines, using ALDH as common cellular marker. By characterizing the miRNA signatures of human tumor-initiating cells, we would be able to extend the application of miR-sensor technology (outlined in chapter 4) to the study of tumor metastasis. Metastatic ALDH+ cells could be tracked *in vivo* using miR-sensors and their ability to migrate to distant

sites could be readily visualized. Furthermore, miR-sensors would aid in the resolution of tumorigenic compartments currently lacking any cellular markers.

Normal human mammary stem/progenitor cells have been shown to express elevated levels of ALDH (aldehyde dehydrogenase) (Ginestier et al. 2008). In addition, differential ALDH activity has been used for the identification and isolation of tumor-initiating cells from breast tumors. More recent work by Charafe-Jauffret et al. demonstrated that breast cancer cells sorted on the basis of ALDH levels differed in their metastatic potential. Cells with elevated levels of ALDH were shown to metastasize to distant sites whereas their ALDH negative counterparts lacked the migratory capacity (Charafe-Jauffret et al. 2009). These studies are appealing because they show that human breast cancer cell lines retain the tumor-initiating populations and that these cells are responsible for metastasis.

A substantial number of luminal breast cancer cell lines are thought to possess a tumor-initiating compartment, but resolution of these cells has proven to be more difficult than cancers of the basal sub-type. For instance, while basally classified breast cancer cell lines have a defined ALDH population, luminally classified tumors often times lack an ALDH compartment (Max Wicha personal communication). Luminal tumors comprise 80% of breast cancers found in patients, which emphasize the need for better characterization with cellular markers (Perou et al. 2000; Sorlie et al. 2001). miR-sensors have been shown to be able to mark specific cellular compartments, and may therefore serve to resolve the tumor-initiating populations in luminal breast cancer cell lines.

This chapter covers the miRNA profiles of a metastatic human breast cancer cell-line, Sum-159. It provides data on potential markers for luminal breast cancer cell lines and lastly, this chapter extends functional studies on let-7 in human breast lines *in vivo*.

3.2 Results and Discussion:

3.2.1 Human breast tumor-initiating cells and self-renewing murine MESC share a similar miRNA expression pattern

We sought to assess the miRNA composition of a human breast cancer cell line in which the tumorigenic and metastatic potential has been well established. We focused on the Sum-159 cells derived from an aggressive, anaplastic breast tumor (Flanagan et al. 1999). Approximately, 5% of Sum-159 cells express elevated levels of ALDH and as little as 500 ALDH ^{positive} cells can generate a tumor mass when injected into a 'humanized' fat pad (Charafe-Jauffret et al. 2008). The gene expression pattern of Sum-159 cells is indicative of a mesenchymal phenotype, and the receptors for estrogen and ERRB-2 are absent.

To assess the miRNA content in the tumor- initiating metastatic Sum-159 cells, we isolated (using FACS) and generated small RNA libraries from 800,000 ALDH ^{positive} cells and 1.5 million ALDH ^{negative} cells. Figure 1a depicts the miRNAs profiles of ALDH ^{positive} and ALDH ^{negative} compartments. MiR-31, miR-141 miR-200c, and miR-205 were the most abundant and differentially expressed miRs in ALDH ^{positive} cells, representing 3%, 4.2%, 0.79% and 11.9% of total miRs sequenced respectively. In contrast, these miRs constituted less than 0.1% of ALDH ^{negative} library. MicroRNA-10a, though not as abundant, was significantly differentially expressed between the two

libraries, constituting 0.47% of ALDH ^{positive} cells and only 0.01% in ALDH ^{negative} cells (table 1). Interestingly, these results overlap with miRNA profiles of COMMA-Dß cells sorted on the basis of ALDH activity.

Recently, miR-200 and miR-205 have been implicated in EMT. Enforced miR-200 and miR-205 expression in MDCK cells alters their fibroblastic phenotype to an epithelial state. Interestingly, inhibition of miR-200 and miR-205 increases MDCK motility in transwell migration experiments. ZEB1 and ZEB2, repressors of E-cadherin are validated miR-200 targets and support the functional migration experiments (Bracken et al. 2008) (Korpal et al. 2008) (Park et al. 2008) (Gregory et al. 2008). These studies demonstrate that miR-205 and miR-200 can enforce an epithelial phenotype and thus suggest that these cells are less likely to metastasize. However, this conclusion is based on the assumption that cells must undergo a transition from epithelial to mesenchymal state in order to migrate, which has not been definitely shown *in vivo*.

miR-31 and miR-10a have also been implicated in invasion and metastasis (Valastyan et al. 2009) (Ma et al. 2007). However, these studies were performed with mixed populations and thus it is difficult to gauge if their results are indicative of the miRNAs, or the differential cell types in their populations. What we can postulate with our current data is that ALDH ^{positive} Sum-159 cells share a miRNA expression patterns comparable to murine mammary progenitor cells, which possess proliferative capacity, are receptive to Wnt signaling, and can reconstitute a functional mammary gland.

Among the differentially down-regulated miRs in ALDH ^{positive} Sum-159 cells were Let-7c, miR-221, let-7i, miR-99a and miR-100, with a contribution less than 0.1% of total miRNA sequences in the library. In their ALDH ^{negative} counterparts, these miRs

contribute between 1-8% of total miRNA reads, suggestive of higher levels of expressions. Let-7c and miR-99a are clustered on the same chromosome, and it has been proposed that they might be regulated in a co-ordinate fashion by the same factors. In addition, miR-100 and miR-99 share the same seed sequence, suggesting that they might share a set of targets, and act combinatorially in regulating gene expression in ALDH ^{negative} cells. Table 1 shows the most abundant up-regulated and down-regulated differentially expressed miRNAs in the Sum-159 ALDH compartments.

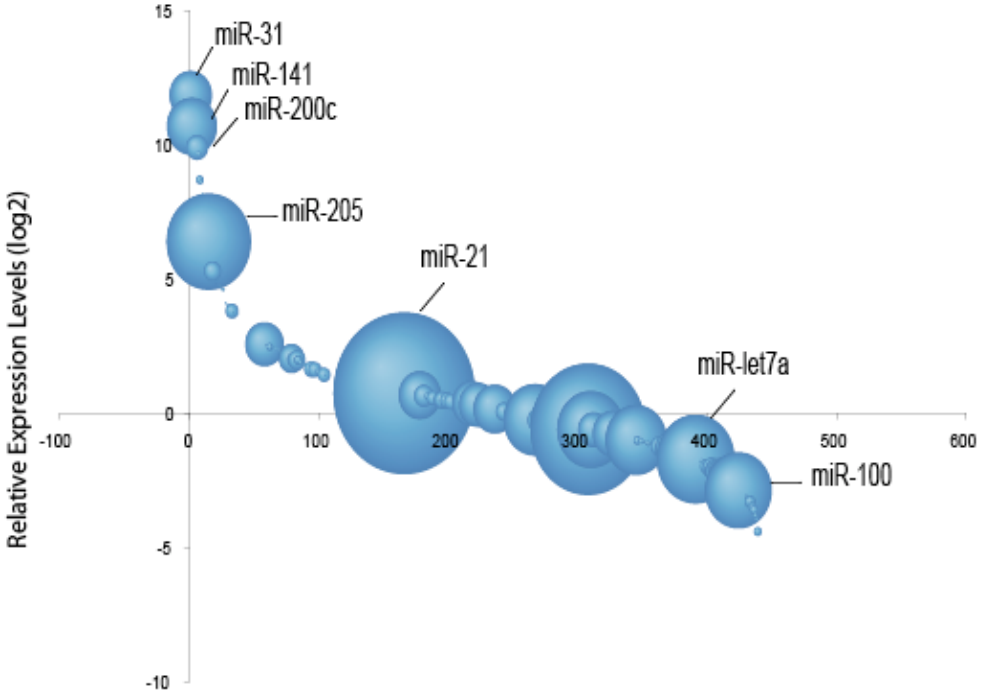


Figure 3.1 MiRNA profiles of Sum-159 cells sorted on the basis of ALDH activity
Small RNA libraries were generated from Sum-159 cells sorted on the basis of ALDH activity. Bubble-plots depicts the relative abundance of miRNA (log2) in ALDH ^{positive} library and ALDH^{negative} libraries. Data was sorted on the basis of differential expression.

3.2.2 miR-sensors are Applicable to Human Breast Cancer Cell Lines

While basal tumors have the worst clinical prognosis (Liu et al. 2008), invasive ductal carcinoma (IDC) constitutes approximately 80% of all breast cancers (Sorlie et al. 2001, (Perou et al. 2000). miR-sensors haven proven to be useful functional markers in resolving a subset of cells within heterogeneous populations that show differential proliferative potential (Ibarra et al. 2007). More recently, we have shown that a subset of HCC-1954 and MDA-MB-453 breast cancer lines can also be marked by let-7 and miR-93 sensor constructs (fig 2a,b). We are currently working with Christophe Ginestier at the University of France to characterize the potential of sensor marked human breast cancer cells to form tumors in NOD/SCID mice.

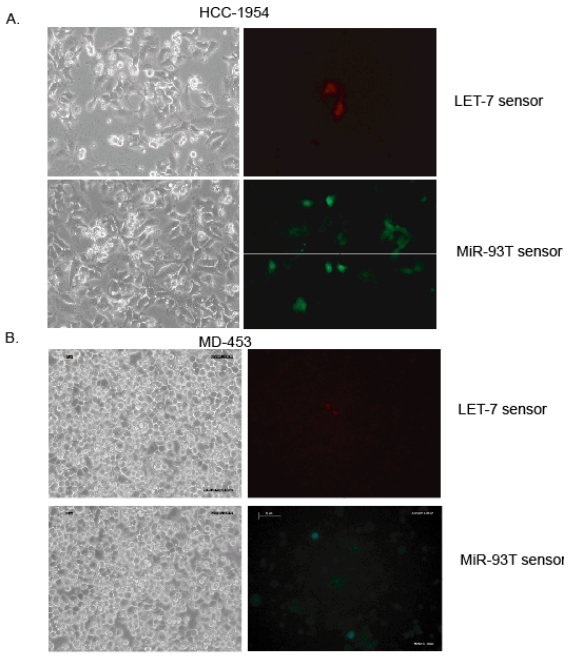


Figure 3.2 Micro-RNA sensors mark a subset of human breast cancer cells.
(a) Bright field images of HCC-1954 cells expressing let-7 sensor (top) and miR-93 tandem GFP sensor (bottom). (b) Bright field images of MDA-MB-453 cell lines expressing let-7 sensor (top) and miR-93 tandem sensor bottom.

3.2.3 MiR-205 and miR-93 mark Tumor-Initiating Cells with Luminal Phenotype

We sought to compare the miRNA composition of ALDH^{positive} cells derived from luminal, basal, and mesenchymal breast cancer cell lines. We reasoned that miRs that are exclusive to basal compartments could be used as positive selectable markers for luminal cells, and, in combination with miRs down-regulated in ALDH^{positive} cells, could resolve populations with tumorigenic potential. MDA-MB-453 breast cancer cells are classified as luminal in origin and contain an ALDH ^{positive} compartment of ~7% (Charafe-Jauffret et al. 2008). Sum-149 and Sum-159 represent basal and mesenchymal gene expression patterns respectively, and also contain ALDH ^{positive} compartments (Charafe-Jauffret et al. 2008). HCC-1954 cells are classified in the basal-like subtype of HER2-amplified breast cancer and also contain cells with elevated ALDH activity. These cell lines should provide a broad and comprehensive analysis of the major breast cancer types.

Small RNA libraries were generated from FACS sorted ALDH^{positive} cells from the cell lines mentioned above and deep sequenced using the Illumina 1G platform. By contrasting the MDA-MB-453 ALDH^{positive} miRNA expression patterns with the ALDH^{positive} fraction of basal and mesenchymal cell lines, we could identify luminal-specific miRNAs. We observed that miR-205, miR-31, miR-181, miR-100, miR-452, and miR-423 were excluded from ALDH^{positive} MDA-MB-453 cells, contributing to less that 0.01% of the total miRNA reads (fig 3a,b,c). In contrast, HCC-1954 ALDH^{positive}, basal Sum-149 ALDH^{positive}, and Sum-159 ALDH^{positive} cells expressed high levels of miR-205, contributing 7.7%, .89% and 12% of total miRNA sequences respectively (fig. 3a, 3b). miR-31 contributed to 2.0% of miRNAs in HCC-1954 ALDH^{positive} cells and 3.1%

miRNAs in Sum-159 ALDH^{positive} cells. However, less than 200 reads (out of millions) of miR-31 was sequenced in basal Sum-149 and luminal MDA-MB-453 ALDH^{positive} cells, indicating that the miRNA is expressed at a very low level.

Among the differentially upregulated miRs in MDA-MB-453 cells were miR-214, miR-199b, miR-130a, miR-320 and the miR-221 family. These miRNAs contributed between 1-4% of the total miRNAs sequences. Let-7c and clustered miR-99a comprised 5.8% and 2.0% of MDA-MB-453 ALDH^{positive} libraries respectively, while miR-199b-3p contributed 3% of total miRNA sequences. When compared to basal and mesenchymal libraries these percentages are higher due to the reduced contribution of the let-7 cluster (less than 0.2%). When we assess the overlap between ALDH^{positive} cells and GFP+ let-7 sensor cells we only observe a 13% overlap, suggesting that not all ALDH^{positive} cells down-regulate let-7 (data not shown).

I concluded from this analysis that miR-205 and miR-200 family is expressed exclusively by basal cells. Luminal MDA-MB-453 ALDH+ cells lack miR-205 and miR-200 expression and thus sensor constructs designed to detect miR-205 and miR-200 activity could potentially be used to mark cells executing luminal cell-fate programs. In combination with other down-regulated miRs (Tables 2-4), sensors can be used to capture a broader range of cells with tumor-initiating potential that do not express ALDH at an elevated levels.

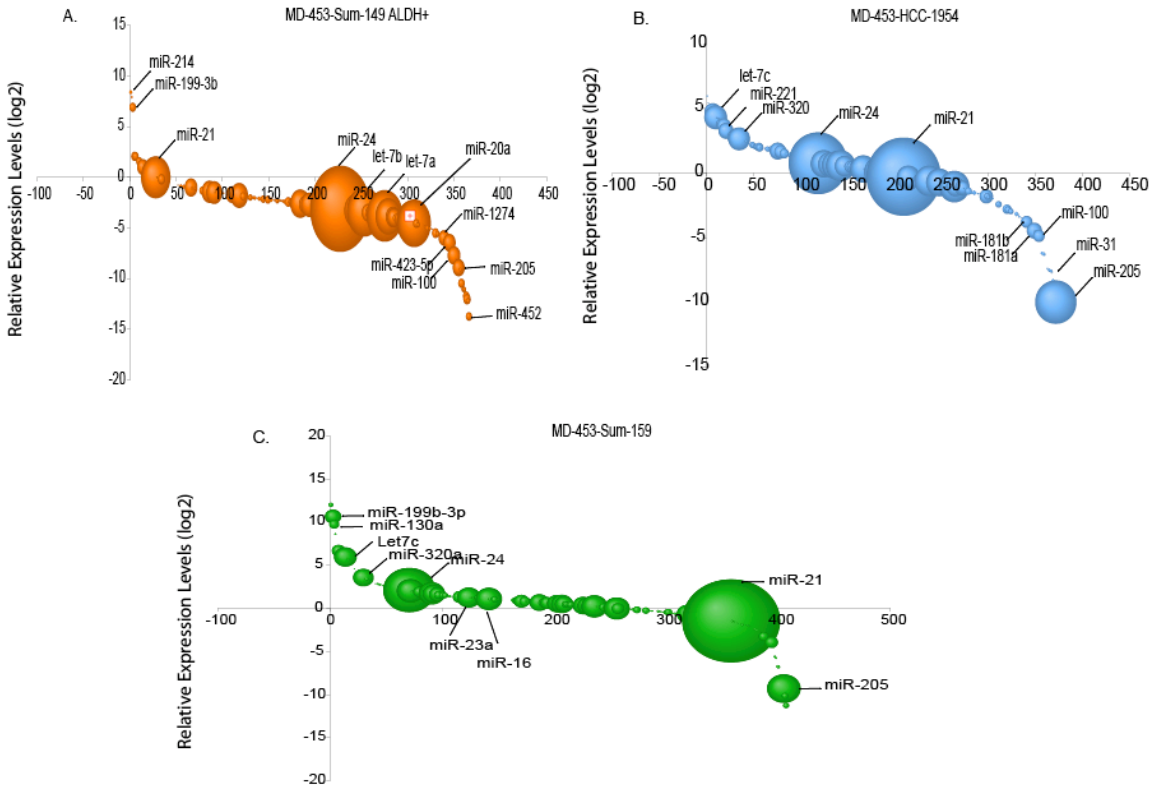


Figure 3.3 MicroRNA profiles of ALDH^{positive} tumor-initiating human breast cancer cells. miRNAs expression patterns that differentiate luminal ALDH^{positive} cells were compared to ALDH^{positive} compartments of (a) basal Sum-149 (b) basal-subtype HCC-1954 (c) mesenchymal Sum-159 cells. Bubble-plots depict the relative abundance and log2 ratio.

3.2.4 Enforced Expression of Let-7 and miR-93 impedes tumor growth of breast cancer xenografts

We have previously shown that enforced expression of let-7c results in a significant decrease in the ALDH ^{positive} compartment of Comma-Dß cells (Ibarra et al. 2007). In addition, Yu et al. showed that over-expression of let-7 in human breast cancer cell line SKBR-3 compromises mammosphere formation and the ability of these cells to form tumors in NOD/SCID mice (Yu et al. 2007a). Let-7 has also been shown to inhibit the growth of multiple human lung cancer cell lines in culture, as well as the growth of cancer cell xenografts in immunodeficient mice (Takamizawa et al. 2004) (Esquela-

Kerscher et al. 2008). We thus sought to test the effects of enforced expression of miRs downregulated in ALDH^{positive} cells on the maintenance of tumor mass *in vivo*. Our initial studies focused on Let-7c and miR-93, because they contributed less than 0.001% of miRNA reads in Sum-159 ALDH^{positive} libraries, and were abundant in ALDH^{negative} cells. We decided to use a tet-inducible lentiviral system in order to control the expression of miR in a dose and time dependent manner. Figure 4a shows a schematic of the lentiviral vector used, pTRIPZ, to achieve tet-regulatable expression of miR-93 and let-7c. The two main components on the pTRIPZ vector are the tetracycline response element (TRE) and the trans-activator. The TRE consist of operators fused to the CMV minimal promoter. A turbo red florescent protein (RFP) reporter is driven by the TRE. The pTRIPZ, transactivator, known as rtTA, binds to and activates expression from TRE promoters in the presence of Dox. Figure 4b shows elevated levels (~200 fold) of let-7c in RNA samples from Sum-159 cells treated with Dox compared to cells cultured in the absence of Dox. Furthermore, expression was specific to let-7 since the levels of miR-100 were constant between the treated and untreated cells. FACS analysis of untreated and treated cells illustrated the induction of miR only upon dox treatment indicated by DsRed expression (fig. 4c, e). Similar results were obtained with miR-93 expression (fig. 4d).

To test if the enforced expression of miR-93 or let-7c would perturb the number of ALDH^{positive} cells, the fat pads of NOD-SCID mice were 'humanized' and injected with 100,000 Sum-159 cells infected with TRIPZ-let-7c or TRIPZ-miR-93. The mammary fat pad of mice are 'humanized' by injecting human fibroblast cells into the fat pad of NOD/SCID mice (Al-Hajj et al. 2003).

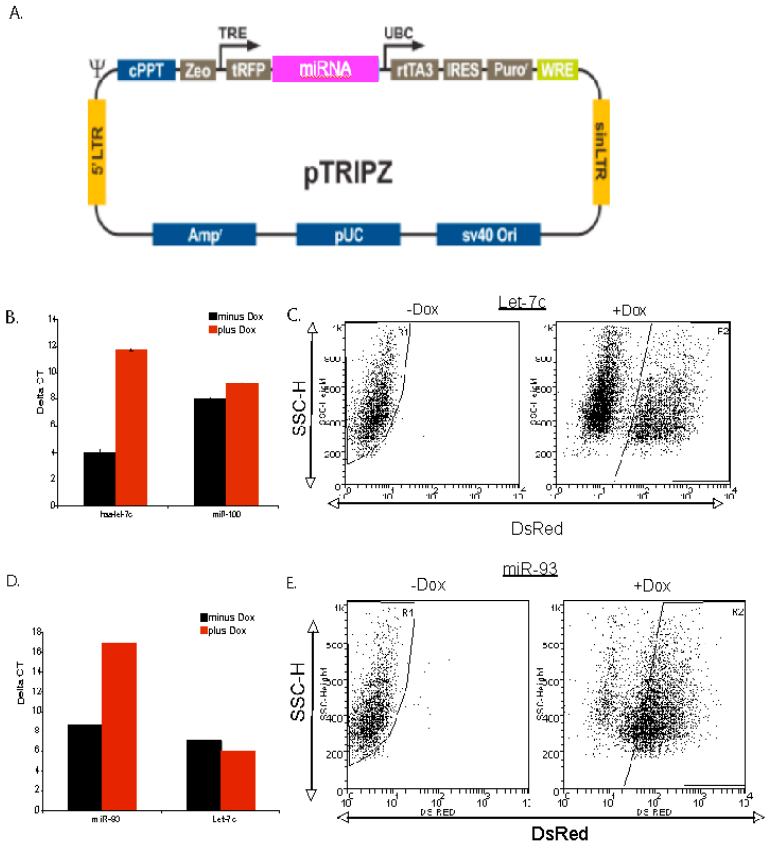


Figure 3.4 TRIPZ lentiviral inducible system for tight regulatable miRNA expression (a) Schematic of pTRIPZ vector (b,d) qRT-PCR for the mature forms of human let-7, miR-93 and miR-100 normalized to U6 RNA from Sum-159 cells used for fat-pad transplantation experiments. (c,e) FACS profiles of Sum-159 let-7c cells (c) or Sum-159 miR-93 cells (e) untreated or treated with Dox. miR expression appears to be tightly regulated as seen by the absence of DsRed cells in the absence of Dox treatment.

Enforced expression of let-7c or miR-93 in

Sum-159 cells and MDA-MB-453 cells impedes tumor growth (fig.5 a,c). Interestingly, immunohistochemical staining of Sum-159 tumors showed a significant reduction in ALDH^{positive} cells when compared to untreated tumors (Figure 5d). In Comma-Dβ cells we also observe a decrease in the number of ALDH+ cells.

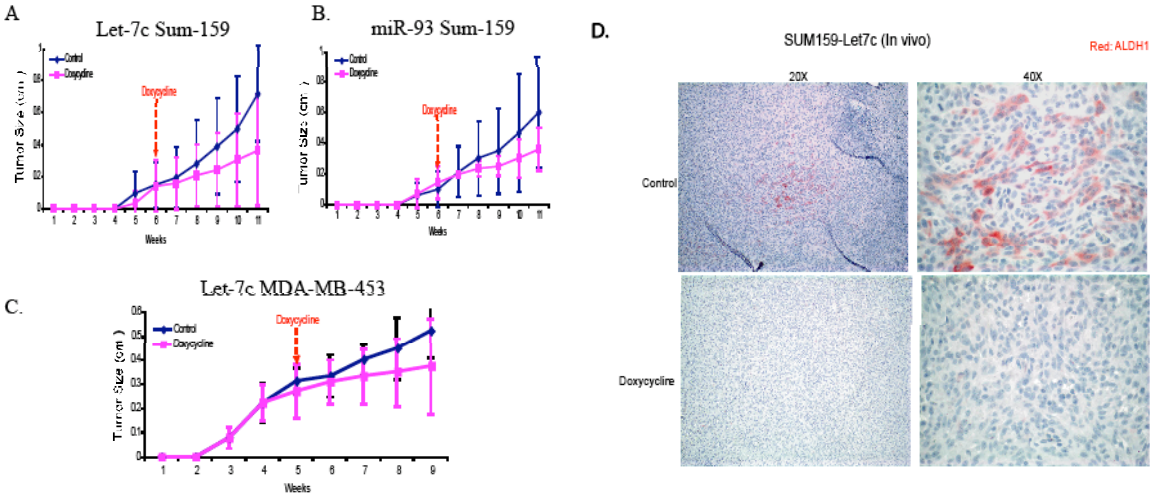


Figure 3.5 Let-7c and miR-93 impedes tumor growth of breast cancer xenografts (a,b) Sum-159 cells (c) MDA-MB-453 cells infected with TRIPZ let-7c or TRIPZ-miR-93 ex vivo and injected into the fat pad of 10 NOD/SCID mice using 1 x 105 cells per fat pad. At .2cm tumor size 5 mice of each group were treated with dox in drinking water every 3 days. Measurements of tumor size were taken once a week for 9 to 11 weeks and averaged. (d) Immunostaining of Sum-159-derived tumors for ALDH expression (red).

To confirm that the depletion of ALDH positive cells leads to the elimination of the tumorigenic compartment, we would need to perform secondary transplantation experiments using cells from dox treated tumors, and assess their capacity to generate a tumor mass. Furthermore, experiments in which cells infected with let-7c or miR-93 are first sorted on the basis of ALDH and than transplanted would provide insights into the differential capacity of these cells to generate tumors in the presence of miRNAs with

potential tumor suppressive roles. Careful *in vitro* and *in vivo* experiments that measure the proliferative rates of ALDH+ cells and ALDH- cells are needed to establish the affect of miRNAs on cell growth and viability in general.

In conclusion, ALDH+ tumor-initiating cells, share a similar miRNA profile to murine ALDH+ progenitor cells. These cells like the ALDH+ progenitors of murine mammary glands down-regulate the expression of highly conserved miRNAs, let-7 and miR-93. The prediction is that manipulation of these cells either by forced differentiation, reduced proliferation or cell death would have enormous impacts on tumor-propagation and maintenance of tumor mass with time. Secondary transplants of dox treated tumors are critical in illustrating that cells with a given capacity can be manipulated by miRNAs. Currently preliminary experiments suggest that tumor-growth inversely correlates with the presence of ALDH+ cells and perhaps elimination of these specific cell-types may be of therapeutic significance.

Table 1

Differentially Up-regulated miRs sorted on the basis of log2 ALDH+ library

Name	Sum-159 ALDH+	Sum-159ALDH-	% A+ library	% A- library
hsa-miR-31	68988	25	3.08133871	0.00083683
hsa-miR-141	95321	77	4.25749822	0.00257743
hsa-miR-429	1529	2	0.06829256	6.6946E-05
hsa-miR-200a	6114	8	0.2730809	0.00026778
hsa-miR-217	1499	2	0.06695261	6.6946E-05
hsa-miR-200c	17911	25	0.79999214	0.00083683
hsa-miR-200b	1225	2	0.05471444	6.6946E-05
hsa-miR-375	2185	7	0.0975927	0.00023431
hsa-miR-141*	589	2	0.0263076	6.6946E-05
hsa-miR-122	421	2	0.0188039	6.6946E-05
hsa-miR-216b	138	1	0.00616375	3.3473E-05
hsa-miR-135b	115	1	0.00513646	3.3473E-05
hsa-miR-451	432	6	0.01929522	0.00020084
hsa-miR-338-5p	66	1	0.00294788	3.3473E-05
hsa-miR-205	268162	4260	11.9774157	0.14259538
hsa-miR-153	39	1	0.00174193	3.3473E-05
hsa-miR-491-5p	38	1	0.00169726	3.3473E-05
hsa-miR-10a	10726	365	0.47907519	0.01221768
hsa-miR-135b*	28	1	0.00125062	3.3473E-05
hsa-miR-134	22	1	0.00098263	3.3473E-05
hsa-miR-338-3p	64	3	0.00285855	0.00010042
hsa-miR-133a	80	4	0.00357319	0.00013389
hsa-miR-675	218	11	0.00973694	0.0003682
hsa-miR-10b	57	3	0.0025459	0.00010042
hsa-miR-302a	19	1	0.00084863	3.3473E-05
hsa-miR-1268	496	27	0.02215377	0.00090377
hsa-miR-142-5p	14	1	0.00062531	3.3473E-05

Differentially Down-regulated miRs sorted on the basis of log2 ALDH+ library

Name	Sum-159 ALDH+	Sum-159 ALDH-	% A+ library	% A- library
hsa-let-7c	3901	23759	0.17423758	1.06119219
hsa-miR-28-3p	579	3533	0.02586095	0.15780092
hsa-miR-221	3943	24082	0.17611351	1.07561893
hsa-miR-574-3p	66	417	0.00294788	0.01862524
hsa-miR-190	9	57	0.00040198	0.0025459
hsa-miR-891a	3	20	0.00013399	0.0008933
hsa-miR-130a	1088	7329	0.04859536	0.32734869
hsa-let-7i	8103	57699	0.36191928	2.57711721
hsa-miR-99a	2285	18091	0.10205918	0.80803181
hsa-miR-519d	1	8	4.4665E-05	0.00035732
hsa-miR-519c-3p	4	34	0.00017866	0.0015186
hsa-miR-1226	5	43	0.00022332	0.00192059
hsa-miR-520f	3	26	0.00013399	0.00116129
hsa-miR-143	778	6775	0.03474925	0.30260436
hsa-miR-130b*	3	28	0.00013399	0.00125062
hsa-miR-199b-3p	1516	14254	0.06771191	0.63665278
hsa-miR-100	20384	198916	0.91044831	8.88455342
hsa-miR-1305	1	10	4.4665E-05	0.00044665
hsa-miR-520d-3p	1	10	4.4665E-05	0.00044665
hsa-miR-625*	1	10	4.4665E-05	0.00044665
hsa-miR-105	4	42	0.00017866	0.00187592
Total miRNA	2238897	2987474		

Table 3.1 Top differentially regulated miRNAs from Sum-159 libraries sorted on the basis of ALDH

Table 2

Differentially Upregulated miRs sorted on the basis of log2 ALDH+ library

Name	MD-453A+	Sum-149A+	%miRNA	%miRNA
hsa-miR-214	17799	2	0.483114341	0.000216102
hsa-miR-199a-5p	6366	1	0.172790937	0.000108051
hsa-miR-199b-3p	66855	21	1.814630556	0.002269071
hsa-miR-222*	957	1	0.02597564	0.000108051
hsa-miR-222	52483	475	1.424534522	0.051324219
hsa-miR-99a	46205	421	1.254132149	0.045489466
hsa-miR-449a	190	2	0.005157128	0.000216102
hsa-miR-125b-2*	1012	12	0.027468493	0.001296612
hsa-miR-375	1003	12	0.027224208	0.001296612
hsa-miR-221	45025	574	1.222103669	0.062021267
hsa-miR-502-5p	78	1	0.002117137	0.000108051
hsa-miR-10a	2581	35	0.070055515	0.003781785
hsa-miR-99a*	1009	15	0.027387065	0.001620765
hsa-let-7c	129167	2429	3.505951462	0.262455848
hsa-miR-188-5p	1821	37	0.049427002	0.003997887
hsa-miR-484	4715	101	0.127978208	0.01091315
hsa-let-7g*	42	1	0.001139997	0.000108051
hsa-miR-374a*	40	1	0.001085711	0.000108051
hsa-let-7i*	115	3	0.00312142	0.000324153
hsa-miR-31	148	4	0.004017131	0.000432204

Differentially Down-regulated miRs sorted on the basis of log2 ALDH+ library

Name	MD-453A+	Sum-149A+	%miRNA	%miRNA
hsa-miR-1274b	2732	6478	0.074154075	0.699954294
hsa-miR-1201	2	5	5.42856E-05	0.000540255
hsa-miR-664*	25	68	0.000678569	0.007347467
hsa-miR-193a-5p	259	726	0.00702998	0.078445017
hsa-miR-1304	14	41	0.000379999	0.00443009
hsa-miR-376c	1	3	2.71428E-05	0.000324153
hsa-miR-423-5p	2665	8816	0.072335509	0.952577502
hsa-miR-1	7	32	0.000189999	0.003457632
hsa-miR-1826	25	152	0.000678569	0.01642375
hsa-miR-196b	87	651	0.002361422	0.070341193
hsa-miR-184	2	15	5.42856E-05	0.001620765
hsa-miR-100	1178	9463	0.031974195	1.022486491
hsa-miR-200a*	4	44	0.000108571	0.004754243
hsa-miR-584	36	399	0.00097714	0.043112344
hsa-miR-34a	8	96	0.000217142	0.010372895
hsa-miR-1287	5	66	0.000135714	0.007131365
hsa-miR-205	454	8240	0.012322822	0.890340134
hsa-miR-923	11	218	0.000298571	0.023555115
hsa-miR-200b	26	543	0.000705712	0.058671686
hsa-miR-155	46	2432	0.001248568	0.262780001
hsa-miR-125b-1*	1	79	2.71428E-05	0.008536028
hsa-miR-138	13	1043	0.000352856	0.11269718
hsa-miR-224	814	10	0.000271428	0.087953503
hsa-miR-200b*	91	1	2.71428E-05	0.00983264
hsa-miR-218	2294	18	0.00048857	0.247868964
hsa-miR-429	3154	20	0.000542856	0.340792813
hsa-miR-452*	1067	6	0.000162857	0.115290403
hsa-miR-452	2072	4	0.000108571	0.223881645
Total miRNA	3684221	925489		

Table 3.2 Top differentially regulated miRNA from libraries generated from ALDH+ sorted MDA-MB-453 cells and ALDH+ Sum-149 cells.

Table 3

Differentially Upregulated miRs sorted on the basis of log2 ALDH+ activity

Name	MD-453	HCC-1954	%miRNA	%miRNA
hsa-miR-199a-5p	6366	1	0.172980528	4.19908E-05
hsa-miR-214	17799	7	0.483644427	0.000293935
hsa-miR-199b-3p	66855	27	1.816621619	0.001133751
hsa-miR-130a	25345	19	0.688688579	0.000797825
hsa-miR-675	1215	2	0.033014663	8.39815E-05
hsa-miR-30a*	3039	16	0.082577415	0.000671852
hsa-miR-125b-2*	1012	6	0.027498633	0.000251945
hsa-miR-99a	46205	290	1.255508218	0.012177322
hsa-miR-296-3p	304	2	0.008260459	8.39815E-05
hsa-miR-30c-2*	131	1	0.003559606	4.19908E-05
hsa-miR-30a	13874	135	0.376992122	0.005668753
hsa-miR-125b	15331	151	0.416582545	0.006340606
hsa-miR-375	1003	10	0.027254079	0.000419908
hsa-let-7c	129167	1341	3.50979829	0.056309616
hsa-miR-450a	175	2	0.004755198	8.39815E-05
hsa-miR-424	552	7	0.014999254	0.000293935
hsa-miR-99a*	1009	13	0.027417115	0.00054588
hsa-miR-483-3p	150	2	0.004075884	8.39815E-05
hsa-miR-503	292	5	0.007934388	0.000209954
hsa-miR-217	109	2	0.002961809	8.39815E-05
hsa-miR-302c	53	1	0.001440146	4.19908E-05
hsa-miR-935	414	11	0.011249441	0.000461898
hsa-miR-195*	29	1	0.000788004	4.19908E-05
hsa-miR-155	46	2	0.001249938	8.39815E-05
hsa-miR-342-3p	4854	218	0.131895615	0.009153987
hsa-miR-137	911	43	0.024754204	0.001805603
hsa-miR-628-3p	105	5	0.002853119	0.000209954
hsa-miR-190	58	3	0.001576009	0.000125972
hsa-miR-497	1478	78	0.040161046	0.00327528
hsa-miR-320a	99336	5413	2.699213599	0.227296013

Differentially Down-regulated miRs sorted on the basis of log2 ALDH+ activity

Name	MD-453	HCC-1954	%miRNA	%miRNA
hsa-miR-181b	2384	14699	0.064779387	0.61722226
hsa-miR-923	11	74	0.000298898	0.003107317
hsa-miR-556-5p	1	7	2.71726E-05	0.000293935
hsa-miR-196b	87	648	0.002364013	0.027210016
hsa-miR-365	7	53	0.000190208	0.002225511
hsa-miR-1287	5	41	0.000135863	0.001721621
hsa-miR-376a	3	26	8.15177E-05	0.00109176
hsa-miR-181a	2615	25773	0.071056249	1.082227996
hsa-miR-642	7	81	0.000190208	0.003401252
hsa-miR-146a	17	250	0.000461934	0.010497691
hsa-miR-584	36	1128	0.000978212	0.047365583
hsa-miR-1826	25	876	0.000679314	0.03678391
hsa-miR-181a*	3	199	8.15177E-05	0.008356162
hsa-miR-147b	39	2729	0.00105973	0.114592799
hsa-miR-376c	1	80	2.71726E-05	0.003359261
hsa-miR-34a	8	1434	0.00021738	0.060214758
hsa-miR-31	148	48571	0.004021539	2.039533466
hsa-miR-452*	6	2304	0.000163035	0.096746723
hsa-miR-205	454	184737	0.012336343	7.757248026
hsa-miR-224	10	7221	0.000271726	0.303215317
hsa-miR-452	4	6254	0.00010869	0.262610247
Total miRNA	3680183	2381476		

Table 3.3 Top differentially regulated miRNAs from libraries generated from MDA-MB-453 ALDH+ and HCC-1954 ALDH+ cells.

Table 4

Differentially Upregulated miRs sorted on the basis of log2 ALDH+ activity

Name	MD-453A+	Sum159A+	%MD-453	%Sum159
hsa-miR-214	17785	402	0.80141926	#DIV/0!
hsa-miR-199b-3p	66826	1516	3.0112816	#DIV/0!
hsa-miR-199a-5p	6364	157	0.28677156	#DIV/0!
hsa-miR-939	35	1	0.00157715	#DIV/0!
hsa-let-7c	128811	3875	5.80442036	#DIV/0!
hsa-miR-222*	956	30	0.04307882	#DIV/0!
hsa-miR-302d	318	10	0.01432957	#DIV/0!
hsa-miR-137	911	29	0.04105105	#DIV/0!
hsa-miR-125b-2*	1012	34	0.04560227	#DIV/0!
hsa-miR-302a*	54	2	0.00243332	#DIV/0!
hsa-miR-130a	25279	1083	1.13911034	#DIV/0!
hsa-miR-302a	432	19	0.01946658	#DIV/0!
hsa-miR-574-5p	3786	171	0.17060294	#DIV/0!
hsa-miR-99a	46160	2193	2.08004009	#DIV/0!
hsa-miR-624*	35	2	0.00157715	#DIV/0!
hsa-miR-222	52592	3291	2.36987583	#DIV/0!
hsa-miR-196a*	74	5	0.00333455	#DIV/0!
hsa-miR-940	87	6	0.00392035	#DIV/0!
hsa-miR-342-5p	273	19	0.0123018	#DIV/0!
hsa-miR-9*	1076	76	0.0484862	#DIV/0!
hsa-miR-7	2535	185	0.11423097	#DIV/0!
hsa-miR-340*	41	3	0.00184752	#DIV/0!
hsa-miR-483-3p	150	11	0.00675923	#DIV/0!
hsa-miR-628-3p	105	9	0.00473146	#DIV/0!
hsa-miR-221	44833	3923	2.02024344	#DIV/0!
hsa-miR-501-5p	294	26	0.01324809	#DIV/0!
hsa-miR-181c*	350	31	0.01577153	#DIV/0!
hsa-miR-1226	56	5	0.00252345	#DIV/0!
hsa-miR-130b*	33	3	0.00148703	#DIV/0!
hsa-miR-320	106650	10193	4.80581186	#DIV/0!

Differentially Down-regulated miRs sorted on the basis of log2 ALDH+ activity

Name	MD-453A+	Sum159A+	%MD-453	%Sum159
hsa-miR-181b	2326	19143	0.06367552	0.86261281
hsa-miR-653	1	9	2.7376E-05	0.00040555
hsa-miR-432	1	9	2.7376E-05	0.00040555
hsa-miR-371-3p	4	38	0.0001095	0.00171234
hsa-miR-142-3p	2	19	5.4751E-05	0.00085617
hsa-miR-122	43	420	0.00117715	0.01892584
hsa-miR-146a	17	187	0.00046538	0.00842651
hsa-miR-584	36	458	0.00098552	0.02063818
hsa-miR-181a	2677	35092	0.07328434	1.58129911
hsa-miR-217	109	1499	0.00298393	0.06754723
hsa-miR-216b	9	138	0.00024638	0.00621849
hsa-miR-200a*	4	62	0.0001095	0.00279381
hsa-miR-196b	84	1333	0.00229955	0.06006702
hsa-miR-100	1177	20377	0.03222102	0.91821874
hsa-miR-184	1	23	2.7376E-05	0.00103642
hsa-miR-708	4	120	0.0001095	0.00540738
hsa-miR-200b	28	1226	0.00076652	0.05524543
hsa-miR-204	3	132	8.2127E-05	0.00594812
hsa-miR-155	46	2116	0.00125928	0.09535019
hsa-miR-451	9	431	0.00024638	0.01942152
hsa-miR-376c	1	55	2.7376E-05	0.00247838
hsa-miR-200b*	1	92	2.7376E-05	0.00414566
hsa-miR-34a	6	601	0.00016425	0.02708198
hsa-miR-125b-1*	1	106	2.7376E-05	0.00477652
hsa-miR-429	14	1530	0.00038326	0.06894414
hsa-miR-452*	6	681	0.00016425	0.0306869
hsa-miR-138	13	1486	0.00035588	0.06696143
hsa-miR-181a*	2	251	5.4751E-05	0.01131044
hsa-miR-224	10	1912	0.00027376	0.08615764
hsa-miR-452	4	1524	0.0001095	0.06867377
hsa-miR-31	148	68983	0.00405158	3.10847932
hsa-miR-205	446	267667	0.01220949	12.0614837
Total miRNA	3652895	2219188		

Table 3.4 Top differentially regulated miRNAs from libraries generated from ALDH+ MDA-MB-453 and Sum-159 ALDH+ cells

Methods:

Cell Culture

HCC-1954, MDA-MB-453, Sum-149, and Sum-159 were obtained by Max Wicha laboratory and cultured as previously described (Charafe-Jauffret et al. 2009). Lentivirus was made using the Translentiviral system from Openbiosystems. The 'humanization' of the mammary fat pad was performed by the Wicha laboratory. Briefly, the fat pads were cleared pre-puberty and humanized by injecting a mixture of irradiated and non-irradiated immortalized human fibroblasts (1:1 irradiated:non-irradiated, 50,000 cells/100µl Matrigel/fat pad). Irradiated fibroblasts (4Gy) support growth of normal and cancer epithelial cells by secreting a variety of growth factors, collagen and possibly directly interacting with the epithelial cells (Orimo et al. 2005); (Tlsty 2001). For a detailed protocol see Ginestier et al. 2008.

Small RNA cloning

1ug of total RNA from sorted cells was used for small RNA cloning performed as described in Pfeffer et al. 2005. A detailed protocol is available upon request. RNA was electroeluted. A current of 100v for 18minutes was applied to elute RNA from polyacrylamide gels. Ethanol precipitations were performed overnight at -20C and samples were centrifuge for 45 minutes at 4C to maximize recovery of RNA.

miRNA expression analyses

Mature miRNAs were quantified using the TaqMan MicroRNA Assays previously described by (Chen et al. 2005) (Applied Biosystems). Data was normalized to U6 RNA. Three biological replicates were performed, with the reactions run in triplicate.

Vectors

Approximately 100bp upstream and downstream of let-7c and miR-93 were amplified from genomic DNA using the following primers and subcloned in ClaI/Mlu sites of pTRIPZ: Human Let7c

forward 5' GGC CAT CGA TCC TAT GAG AAA CAG TTA GCA GC
reverse 5' CTAGACGCGTTGAACATGGAGTGACAACCCA
hsa-miR-93
forward 5' GCG AAT CGA TGC CTT TTC CCC ACT TCT TAA CC
reverse 5' GTATACGCGTGGAGACCAGACCCTTTTGAAC

miR-sensor for let-7 and miR-93 sequences were previously described in Ibarra et al. 2007.

CHAPTER 4

Let-7 Expression in Mice

4.1 Introduction

Despite its discovery over nine years ago, the expression patterns and functional role of Let-7 in mammalian systems remains largely unexplored. Let-7 expression is generally observed in differentiated cell types, and interestingly, less differentiated cell types employ mechanisms that tightly down-regulated let-7 activity. Recent work in embryonic stem cells revealed several mechanisms in which let-7 can be post-transcriptionally regulated during its biogenesis. Lin-28 is an RNA-binding protein that has been shown to bind specifically to the conserved nucleotides in the loop region of the Let-7 precursor and thus inhibit processing of the primary transcript by Drosha (Viswanathan et al. 2008 and Newman et al. 2008). RNAs that mimic the binding site can competitively suppress Lin28 activity, and restore Let-7 processing (Newman et al. 2008). Heo et al. showed another Lin28-mediated regulation of let-7 through induction of uridylation at its 3' end of the precursor let-7 (pre-let-7) in the cytoplasm. The uridylated pre-let-7 is inaccessible to Dicer processing, and undergoes degradation (Heo et al. 2008). Lin-28 mediated regulation is not restricted to embryonic stem cells, but has also been observed in neuronal progenitors (Rybak et al. 2008). Coupled with the presence of mature let-7 in differentiated cells, it can be proposed that Lin-28 is part of the network regulating pluripotency. This also implies that stem cells have evolved an intricate mechanism to regulate differentiation through the modulation of the small RNA pathway, and perturbation of this network might lead to expansion or reduction in the stem cell/progenitor compartments in various tissues, including the mammary gland.

Temporal expression of let-7 has been shown to be under the control of various hormones. For example, expression of the primary, precursor and mature let-7 in

Drosophila coincides with an ecdysone pulse at the late third-sate of third-larval instar and prepupae stage (Bashirullah et al. 2003), (Sempere et al. 2002), (Sempere et al. 2003). Furthermore, nuclear receptor DAF-12 and its steroidal ligand directly activate promoters of let-7 microRNA family members in *C. elegans* to down-regulate the microRNA target, hbl-1, which drives the progression of epidermal stem cells from the second to third larval stage (Bethke et al. 2009). However, the expression of let-7 in response to hormonal cues in mammals is unknown, and it is uncertain if the evolutionarily conserved hormonal response is present in mammalian systems.

In *C. elegans*, let-7 is expressed during late larval stages and required for the adoption of a fully differentiated state in seam cells (Reinhart et al. 2000). In *D. melanogaster*, let-7 is expressed during larvae to adult metamorphosis and guides the transitions from 3rd instar to the pupal stage (Bashirullah et al. 2003), (Sempere et al. 2002). The expression pattern and function of Let-7 in mammalian tissue development remains elusive. Recent studies revealed distinct patterns of Let-7 family miRNAs expression in chick and mouse limbs, hypothesized to regulate embryonic limb development (Kloosterman and Plasterk 2006), (Schulman et al. 2005). However, there have been no robust mouse models to analyze the spatial and temporal expression of let-7, which has restricted our ability to identify the functional domains of let-7 during tissue formation and differentiation.

To assess the dynamic spatiotemporal expression of let-7 *in situ* during mammalian tissue development (with the focus on the mammary gland and bone marrow), we adapted an assay previously described (Parizotto et al. 2004) Brennecke et al. 2004, (Hutvagner and Zamore 2002), based on the ability of miRNAs to inactivate

genes by classical RNAi. A construct was generated containing a GFP reporter under the control of the tetracycline-inducible CMV promoter, with two perfectly complementary let-7 target sites in the 3′UTR. A similar construct without the let-7 target sequences in the 3′ UTR was used as a control. The presence of let-7 miRNA should result in the degradation (by RNAi) of the GFP transcript containing the let-7 target sequence, leading to the reduction of the reporter expression. This will provide an *in vivo* sensor for let-7 activity, as cells that express the miRNA will appear dimmer under GFP excitation conditions.

The following chapter describes the generation of a let-7 sensor mouse. We initially focused on the bone marrow for several reasons: (1) it is an easily assessable tissue and bypasses any staining procedures; (2) stem and progenitor compartments have been well-defined; (3) it's comprised of cells at different developmental stages allowing assessment of let-7 levels during cell differentiation ;(4) bone marrow cells provide a means of quantification of let-7 expression based on GFP expression using flow-cytometry. To begin uncovering the function of let-7 in mammalian cells, this chapter also contains experiments describing the use of small Locked Nucleic Acid (LNA) oligonucleotides, a new generation of miRNA inhibitors, targeting let-7. Finally the chapter concludes with let-7 expression patterns in mammary gland development and its potential role during mammary gland differentiation.

4.2 Results:

4.2.1 Generation of Let-7 sensor ESC-derived mice

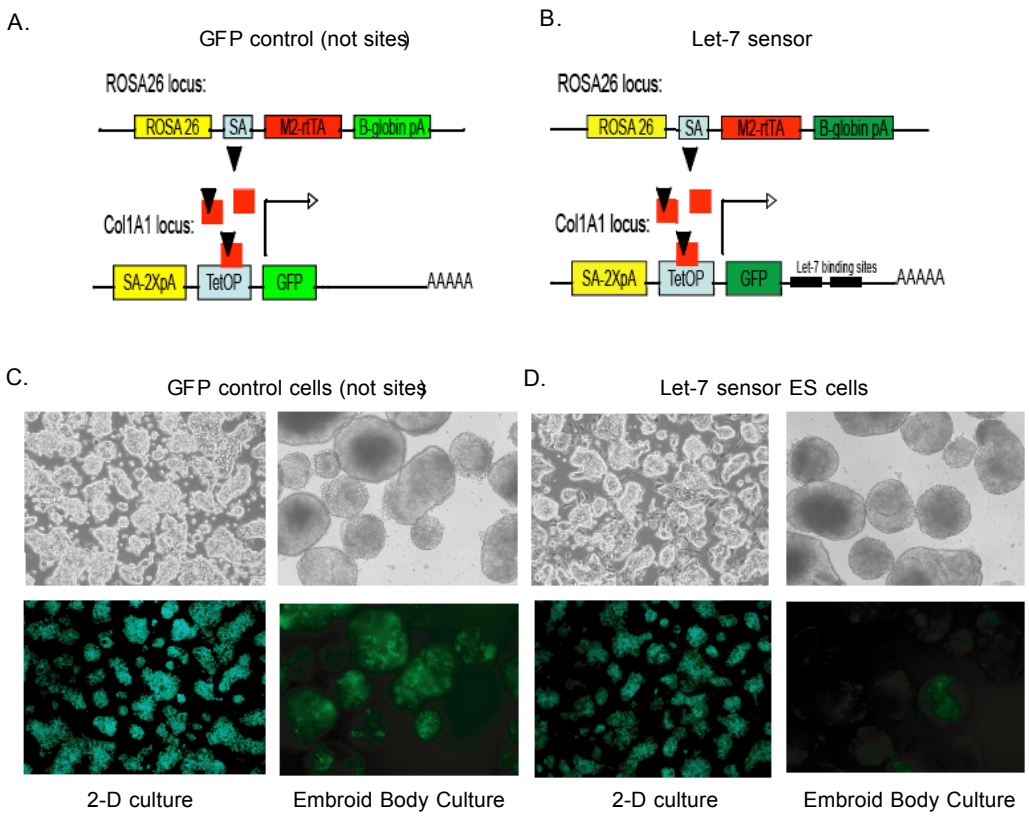


Figure 4.1 Characterization of Let-7 sensor ES cells in vitro.
(A-B) Schematic representation of transgenes found in the GFP ES cells and Let-7 sensor cells and mice. (A) M2-rtTA gene was targeted into the ROSA26 locus under control of the ROSA26 promoter. An eGFP cassette was placed under the control of the doxycycline (DOX)-responsive promoter, and inserted downstream of the collagen-1A1 locus (Beard et al. 2006). SA, splice acceptor; pA, polyadenylation signal; TetOP, tetracycline/doxycycline-responsive operator. Black arrows indicate transcriptional start sites. (B) A similar system to (A) was used, but two let-7 sites were cloned in tandem into the 3'UTR of eGFP. This would allow let-7 to repress GFP expression via RNAi (C-D) Sensing let-7 activity under differentiation conditions. (C) Micrographs of GFP control ES cells cultured in 2D cultures in the presence of LIF (left) or as embryoid bodies for seven days in the absence of LIF and presence of retinoic acid. (D) Micrographs of let-7 sensor ES cells cultured in 2D (left) or as embryoid bodies for seven days in the presence of retinoic acid.

Embryonic stem cells (ESC) are pluripotent cells derived from the inner cell mass of an early stage embryo known as a blastocyst, that have the ability to contribute to the

development of an entire embryo (Martin 1981), (Evans and Kaufman 1981) . Using a myriad of gene-targeting techniques, scientists have been able to genetically modify these ESCs to answer specific biological questions. These modified ESC can then be introduced into host embryos, and develop into all tissues types, including the future germline. This has enabled the creation of genetically modified mice containing (or missing) genes of interest.

A major obstacle to making genetically modified mice has been the difficulty in ensuring germline transmission of the modification to the next generation. To circumvent this, a technique known as ESC-tetraploid aggregation was developed (Wood et al. 1993). By aggregating ESC with tetraploid host embryos, it was possible to generate a newborn that is completely derived from the ESC (known as ES-cell-derived animals). Although initially inefficient, the use of improved ES cell lines has made the production of ES-cell-derived animals a viable and practical approach (Eggan et al. 2002), (Eggan et al. 2001). This has enabled us to utilize this technology in the generation of our miRNA-responsive sensor mouse model.

To generate a miRNA-responsive sensor, we took advantage of the KH2 ESC engineered in the laboratory of Rudolph Janeich (Beard et al. 2006). This ES cell line expresses the M2 reverse tetracycline transactivator (M2-rtTA) constitutively from the ROSA26 locus (ROSA26::rtTA (Urlinger et al. 2000), and contains a FRT docking site in the collagen-1A1 locus, which allows the Flp-mediated site-specific integration of a eGFP reporter with or without perfect let-7 binding sites (Cola1a::TetOP-let7-sensor) (McCreath et al. 2000) (Fig.1a,b). This dual system has previously been shown to

operate in many tissues types, with minimal variations in expression level between independently generated ES cell lines (Beard et al. 2006) (Hochedlinger et al. 2005)

Under defined culture conditions containing various growth factors and cytokines, ESCs can differentiate into a variety of cell types, including cardiomyocytes, dopaminergic neurons, pancreatic β cells, hematopoietic cells, endothelial cells, and adipocytes (Kawasaki et al. 2000). In most scenarios, the formation of an embroid body (EB) is an intermediate step in the differentiation to these specific cell-fates (Nagy, 2003). Essentially EB are formed by culturing ES cells under low -attachment conditions in the presence of retinoic acid and the absence of Leukemia Inhibitory Factor (LIF) for at least 7 days. In ES cells and neuronal precursors, mature Let-7 expression is detected by northern blot analysis only under induced differentiation conditions, such as EB formation (Viswanathan et al. 2008), (Newman et al. 2008). To validate whether our let-7c sensor was 'responsive' to let-7 activity, we induced the let-7c sensor ESC into embryoid bodies, and look for reduction in the expression of GFP (as a result of let-7-mediated repression). GFP expression in control ES cells cultured in 2D self-renewing conditions (in the presence of LIF), or differentiating conditions (embryoid bodies formation and in the presence of retinoic acid) retained similar levels of GFP expression (Figure 1c). In contrast, let-7c-sensor ES cells show a down-regulation of GFP expression when cultured under differentiating conditions, suggesting a higher level of let-7 expression when compared to undifferentiated (self-renewing) ES cells.

To generate let-7-sensor mice, we employed ESC-tetraploid aggregation (Eggan et al. 2001, (Eggan et al. 2002) a technique successfully optimized by Sang Young Kim at CSHL. In this procedure, ten to fifteen ES cells were injected into a four-cell stage

embryo. In this procedure, ten to fifteen modified ES cells were injected into a tetraploid blastocyst, and then transplanted to pseudo-pregnant females to allow normal embryonic development. We were able to obtain nineteen GFP control mice and fourteen let-7c sensor mice at birth (within 3 weeks of injection). Shown in figure 2, is genotype of founding males containing both the tTA and GFP transgene.

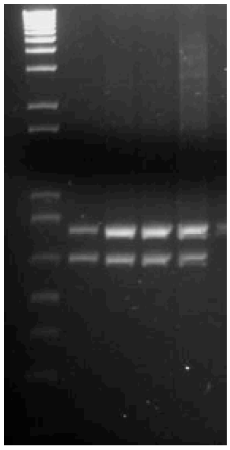


Figure 4.2 ES- Mice Genotype
750 base pair GFP (upper band) and approximately 650 basepair rtTA (lower) band.

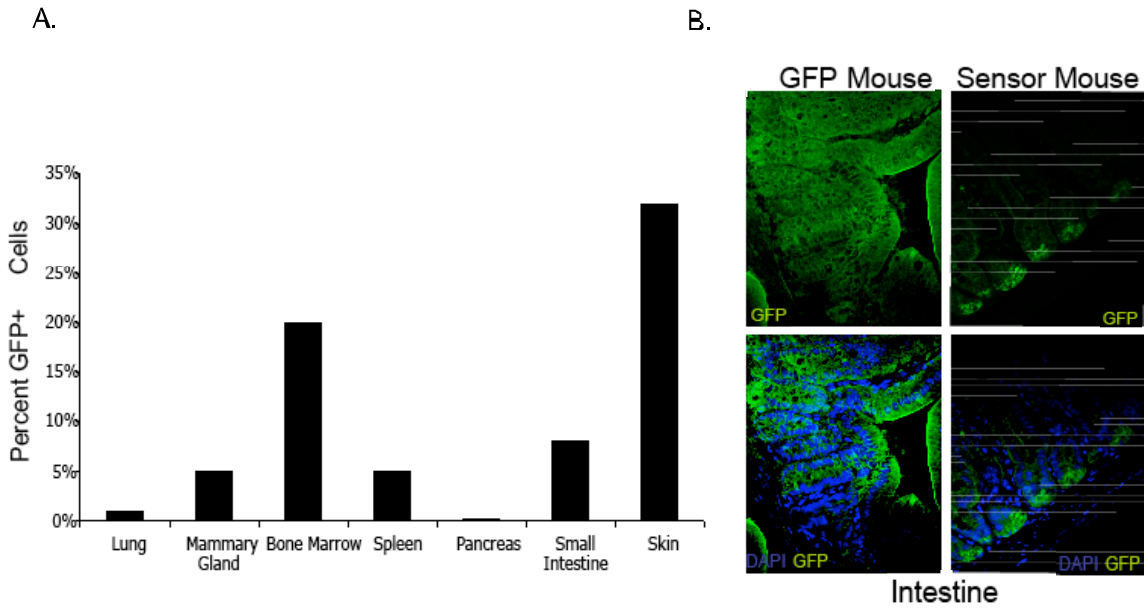


Figure 4.3 Expression of Let-7 GFP reporter in Tissues
Let-7 expression is abundant in mammalian tissues. (A) FACS analysis of GFP expression in dissociated tissues. Tissues were dissociated using collagense IV. (B) Confocal images of paraffin-embedded intestinal tissue derived from control and let-7-sensor mice treated with DOX for 4 days. The tissue is immunolabeled for GFP expression and the nuclei are stained with DAPI. In the sensor mice, GFP expression is restricted to a few cells located at the crypt base.

4.2.2 Let-7 expression in Tissues

During the initial characterization of let-7 in mammals, Reinhart et al. 2001 detected the microRNA expression using northern blot analysis. In figure 3a, we are able to show a more quantitative analysis of let-7 expression by examining the percent of GFP cells in various tissues dissociated from the let-7c sensor mice and analyzed by flow cytometry. Interestingly, GFP expression is highest in the bone marrow, suggesting an overall down-regulation of let-7 in bone marrow derived cells. Reinhart et al. have also reported this observation, lending credence to our sensor mouse model as a faithful representation of let-7 levels in the mouse. In contrast, lung and pancreas tissues have

less than 0.2% percent of GFP positive cells, indicating that let-7 is more abundant in these tissues.

Immunodetection of GFP in intestinal tissue sections of control mice and sensor mice revealed differential fluorescence between cells within the crypt base and villi (figure 3b). Interestingly, these correlate with previously reported self-renewing cells of the intestine that were shown to be found at the bottom of the intestinal crypt (Barker et al. 2009).

4.2.3. Let-7 expression is up-regulated in mature erythroblasts and down-regulated in self-renewing stem cells in the bone marrow

The bone marrow and spleen are heterogenous tissues comprised of cells at different developmental stages (Rossi et al. 2008). The various compartments within the bone marrow contain short and long-term repopulating stem cells, as well as committed myeloid and lymphoid progenitors (Kiel et al. 2005). Given the role of let-7 as a heterchronic gene regulator in worms and flies, we reasoned that let-7 would also be downregulated in more 'primitive' (less committed) cellular compartments of the bone marrow and that differential GFP expression (let-7 activity) might be used for the prospective enrichment of HSC and progenitors cells.

We first sought to assess let-7 expression in erythrocytes, a fully committed and differentiated lineage that comprise a substantial percentage of cells within the bone marrow. The cellular marker, Ter119 (a molecule associated with glycophorin A), identifies the late stages of murine erythroid lineage (Kina et al. 2000), with approximately 30% of bone marrow cells positive for Ter-119 (figure 4a). Figure 4b

shows elevated levels (~3 fold) of let-7c in RNA samples from sorted Ter-119+ cells compared to Ter-119- cells. To test if relative GFP intensities of the let-7 sensor could be correlated with let-7 levels *in vivo*, I compared the GFP mean fluorescence intensities of Ter-119- cells derived from GFP control and let-7 sensor mice. Figure 4c shows a histogram overlay of control (green) and sensor (black) Ter-119^{negative} cells. A minor shift in GFP intensity (blue arrow) is observed in the sensor cells, which is not significant enough to suggest reporter repression. In contrast, Ter-119^{positive} cells derived from the sensor mice resulted in a 2-log shift in GFP intensity when compared with the GFP control (fig 4c), indicating a significant repression of the reporter transgene as a result of let-7 expression (figure 4c), consistent with our Q-PCR analysis. It is worth noting that a secondary peak (of higher fluorescence) can be detected. This can be explained by the presence of polychromatic erythroblasts in the Ter+ subset, thought to be a progenitor population and possibly expressing less let-7. From these experiments, it can concluded that let-7 is upregulated in mature erythrocytes and that its expression can be qualitatively measured in mixed populations of cells at different developmental stages based on GFP intensity.

To test the hypothesis that HSC and progenitor cells are enriched within the GFP positive fraction of bone marrow cells, we assayed the colony-forming potential of cells isolated from the sensor mice with differential GFP expression. Various populations were sorted onto methylcellulose-based media, supplemented with Interleukin-3 (IL-3), Interleukin-6 (IL-6), Erythropoietin (EPO) and stem cell factor, to promote the growth of bone marrow progenitor cells, and is commonly used to assess HSC potential *in vitro* (Manz et al. 2002).. GFP^{bright} SSC^{lo} cells (population 5) formed colonies on

methylcellulose, whereas all other GFP compartments failed to grow under these conditions (Figure 4d). Furthermore, population 5 captured all previously characterized progenitor compartments, including colony-forming unit-erythroid (CFU-E), burst-forming unit-erythroid (BFU-E), CFU-granulocyte, marcrophage (CFU-GM) and CFU-granulocyte, erythroid, macrophage, megakaryocyte (CFU-GEMM)(figure 4d). Q-PCR analysis of GFP ^{bright} and GFP ^{negative} cells confirmed the down-regulation of mature let-7c (~3-fold) in GFP ^{bright} cells (figure 4e). In addition to these functional assays, immuno-phenotyping experiments revealed that 90% of lineage-depleted, ckit+ and sca-1+ cells resided in the GFP bright compartment (log4-log5)(data not shown). Short and long-term repopulating cells have previously been found within these c-kit and sca-1 fraction of lineage depleted cells in the bone marrow (Uchida and Weissman 1992), (Weissman et al. 2001), suggesting that self-renewing cells also is correlated with downregulated let-7 expression.

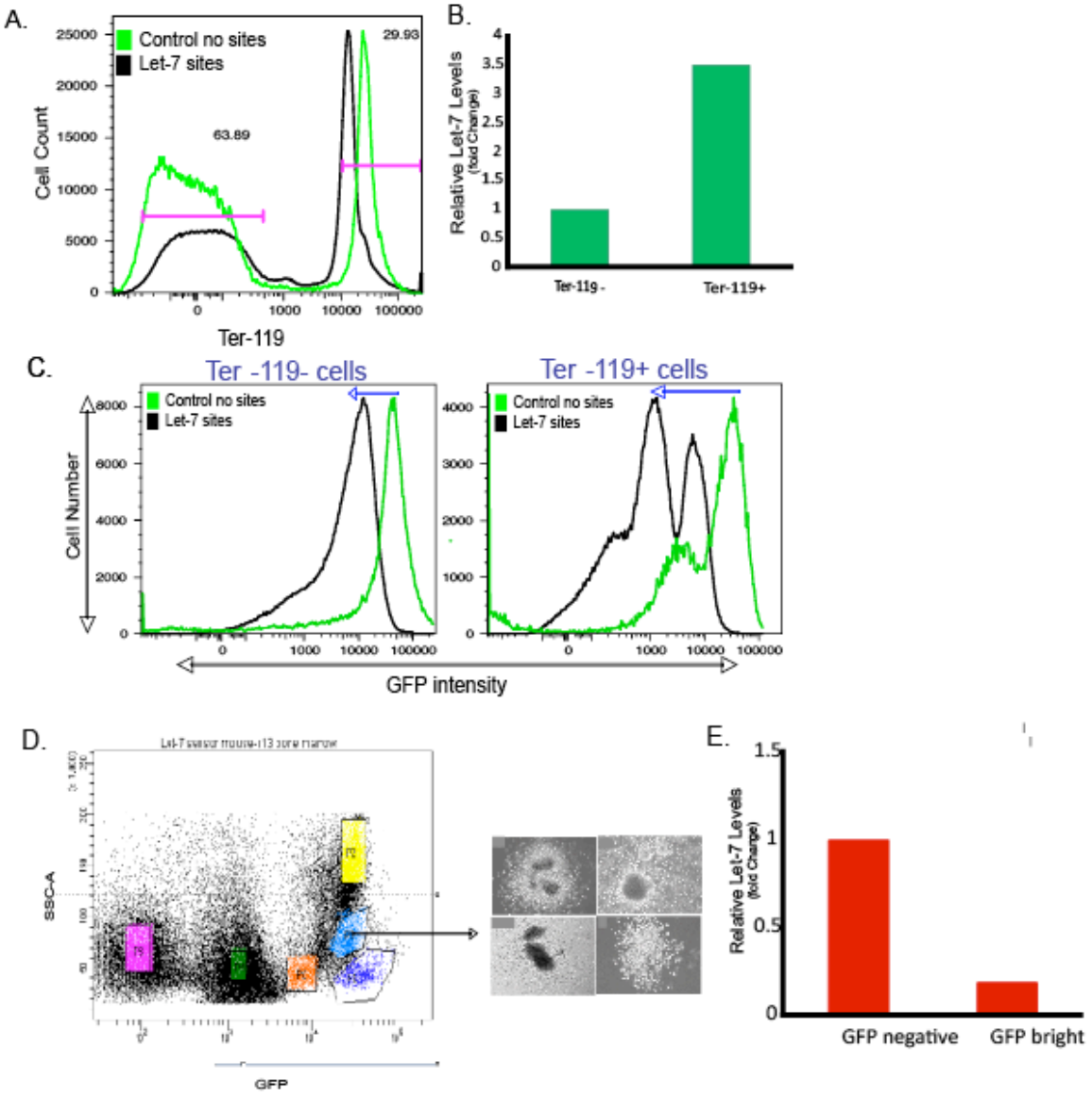


Figure 4.4 Let-7 is expressed in committed Ter-119+ cells and is down-regulated in Colony-forming cells

(a) Histogram overlay of bone-marrow cells derived from control GFP mice and let-7 sensor mice stained with Ter-119. Gates shown were used for FACS analysis and sorting for subsequent Q-PCR. (b) Q-PCR analysis for the mature form of let-7c (normalized to U6 RNA), showing a 3-fold up-regulation in erthyroid lineage Ter119+ cell (c) Histrogram overlay of Ter-119- (left) and Ter-119+(right) subsets from control and sensor mice showing a 2-log shift in GFP mean fluoresence intensity of Ter-119+ subset. (d) Gated populations that were isolated and cultured onto methylcellulose. Morphology of day-10 colonies derived from one thousand sorted cells (inverted light microscope) shown. Colonies: CFU-granulocyte/ erythroid/ macrophage/ megakaryocyte (GEMM), CFU-granulocyte/macrophage (GM), burst-forming units/erythroid (BFU), and CFU-megakaryocyte (M). (e) Q-PCR analysis for mature form of let-7c in cells sorted on the basis of GFP expression. GFP bright (log-4 –log5) and GFP negative (log-2). Data was normalized to U6 RNA.

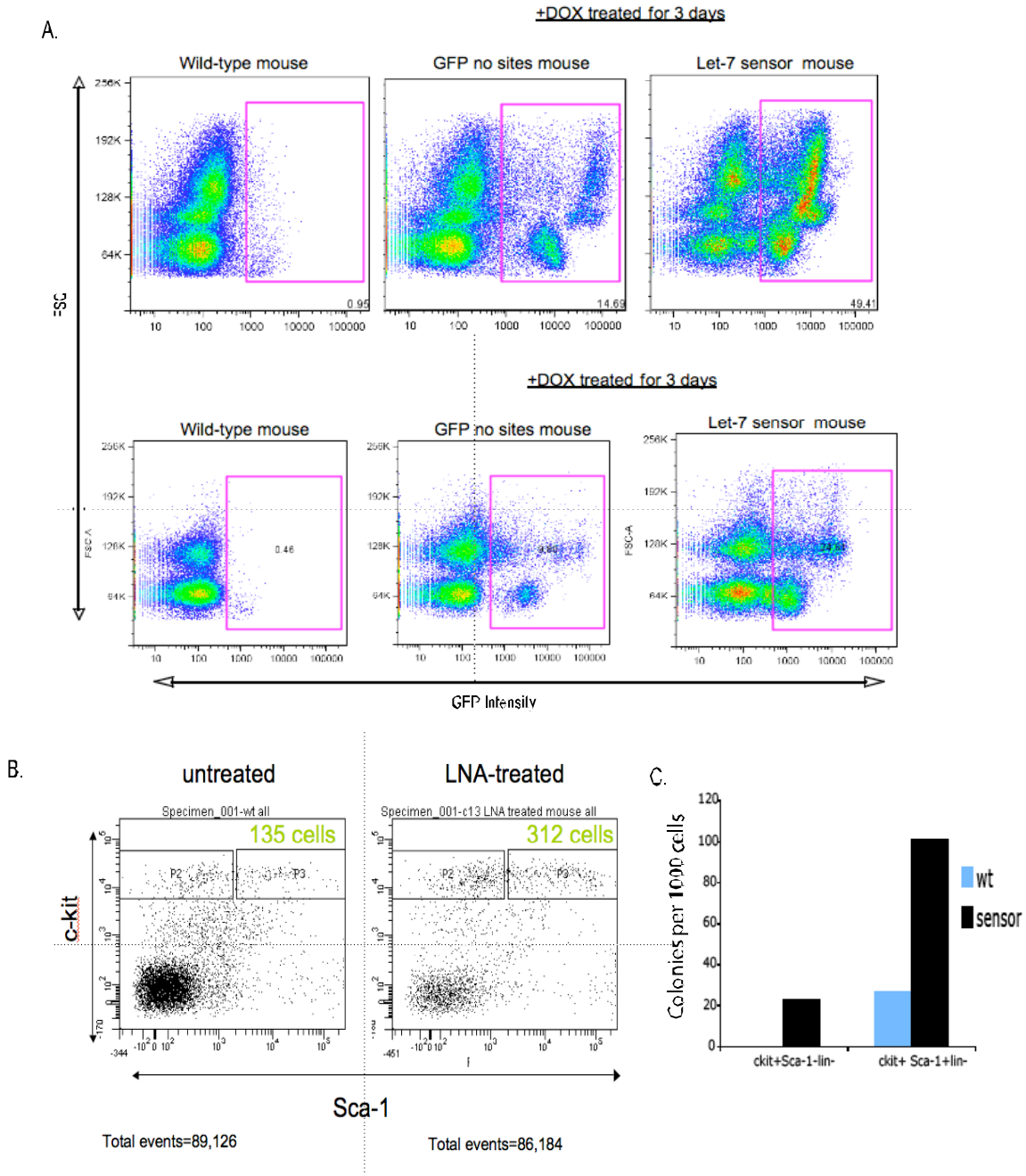


Figure 4.5: **Inhibition of let-7 expression results in an increase of ckit+,sca-1+ cells** An eight-nucleotide LNAs molecule blocks let-7 activity in sensor mice. Control mice and sensor mice injected (intravenous) with 10mg/kg/body weight for three days and analyzed 72hrs after the last injection. (a) FACS plots of bone marrow (top) and spleen (bottom) cells of untreated wild-type mice (left), control GFP mice (middle) and sensor mice (right). A block in let-7 expression increases ckit+sca-1+ cells (b) FACS plots of ckit+sca-1+ fraction of untreated (left) and 24 day-old mice injected with 10mg/kg/body weight let-7 LNA and analyzed 3 weeks after last injection (right). (c) To test colony-forming efficiency and differentiation capacity of ckit+sca-1+ cells, untreated and LNA treated mice these cells were FACS sorted unto methylcellulose and counted after 10d culture. Shown are number of colonies obtained/1000 sorted cells.

4.2.4 Inhibition of let-7 expression in vivo results in the expansion of ckit+sca-1+ cells

Recently, our lab is collaborating with Santaris Pharma to test the efficacy and specificity of a new generation of locked nucleic acids (LNA) miRNA inhibitor. These locked nucleic acids (LNA) inhibitors are eight-nucleotide RNA that induces a steric block through complementary binding of the LNA molecule to the seed sequence of miR (nucleotides 2-8). The cellular outcome of such binding is still unclear, with reports suggesting either simple sequestration by stoichiometric complex formation between the mature microRNA and the LNA inhibitor (Chan et al. 2005) or a yet unknown mechanism leading to degradation of the target microRNA (Krutzfeldt et al. 2007), (Krutzfeldt et al. 2005); (Esau et al. 2006).

In order to evaluate the potential of the LNA inhibitor in microRNA inhibition in cells *in vivo*, we intravenously injected LNA inhibitors against let-7 into control GFP mice and let-7 sensor mice. FACS analysis of bone marrow cells and dissociated spleens from untreated and treated mice are shown in Figure 5a. De-repression of the GFP reporter was observed in let-7 sensor mice (relative to control) as seen by an increase in GFP + cells from 15% to 49%s (5a top). In the spleen we also observed de-repression of GFP reporter from 9% (GFP mouse) to 24% in the let-7 sensor mice (figure 5a bottom).

We then assess the biological consequence of let-7 inhibition in the HSC and progenitor compartments of the bone marrow. The let-7 sensor mice treated with the same regimen as described above were analyzed 3 weeks after the last injection. It is important to note that that these initial experiments were performed in 24-day old mice and that older mice tended to give variable results with respect to the fold changes in the

HSC compartment (data not shown). Figure 5b shows the FACS analysis and colony-forming assay on 1000 sorted cells from c-kit+, sca-1+ compartment of LNA-treated and wild type mice. LNA treated mice resulted in an approximately 3-fold increase in c-kit+ Sca-1+ fraction. Furthermore, we observe a 5-fold increase in colony-forming efficiency of LNA-treated cells, suggesting an increase in proliferative capacity of ckit+sca-1+ cells due to let-7 inhibition (figure 5c).

4.2.5 Let-7 expression During Mammary Gland Development

The application of miRNA sensors for the study of mammary gland biology should reveal upstream regulators of miRNA expression. By understanding the location of miRNA activity *in vivo*, we can attempt to correlate their activity with cellular or developmental processes. Though the expression patterns of miRNAs have been assessed in a majority of tissues using northern blot analysis, Q-PCR and miRNA- array platforms, the function of most miRNAs, including let-7 remain elusive in mammalian systems.

The mammary gland is an ectoderm-derived tissue, and is different to most other tissues in that its development occurs postnatally in response to hormonal responses. It undergoes various morphological transitions in response to ovarian-secreted hormones, to achieve a structure that maximizes milk production. The understanding of mammary gland development inevitably led to the elucidation of genetic and epigenetic mutations that cause breast cancer. Key cellular processes such as proliferation, apoptosis and differentiation are often disrupted in cancers, and are reflected in specific developmental stages during mammary gland development. We have previously shown that let-7 expression is abundant in the differentiated cell-types of the mammary epithelium and downregulated in less-differentiated mammary progenitor cells (Ibarra et al. 2007).

However, we have only observed this in an immortalized cell line derived from a mid-pregnant female, thus capturing a glimpse of let-7 expression during a particular stage of development. The generation of a mouse model that would allow the *in vivo* assessment of let-7 (a strong inducer of differentiation programs), will shed light into the general biology of mammary epithelial stem cells.

While FACS analysis provided preliminary evidence of the dynamic regulation of let-7 expression in mammary gland development, it cannot provide the spatial resolution in the tissue of interest. We thus sought to assess the expression patterns of let-7 (via GFP expression) in the developing mammary gland by indirect immunofluoresence to uncover tissue localization of let-7. To ensure both specificity and quantitative comparisons, we tested a range of antibody dilutions and chose the maximum dilution that can still discriminate between positive signals from noise. Images were acquired using same exposure time and laser intensity to allow semi-quantitative comparisons.

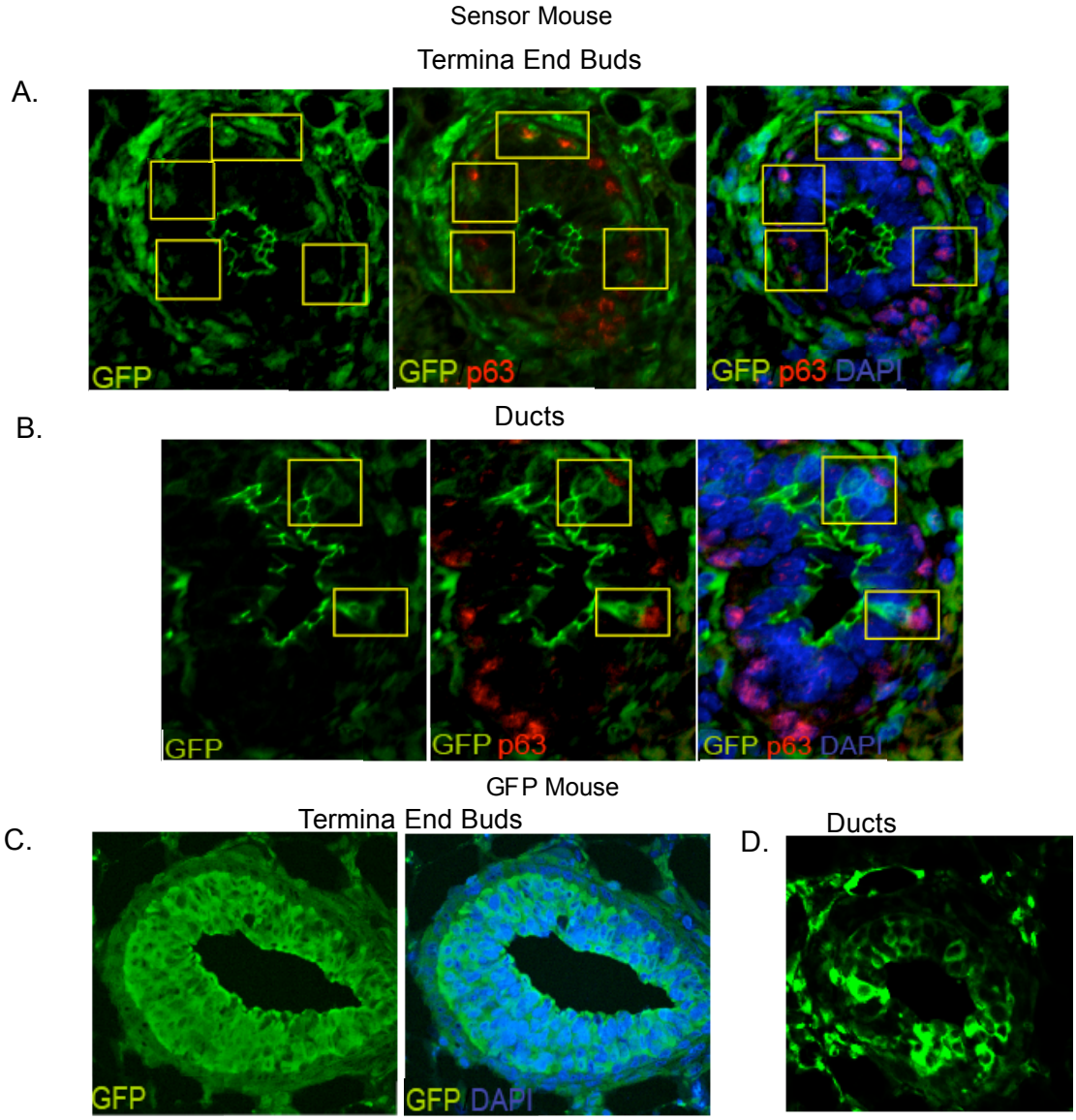


Figure 4.6. Spatial expression pattern of Let-7 in virgin mammary glands.
(a-b) Immunodetection of GFP (left), p63 (basal marker) (middle), DAPI-merge (right) in 5 week old virgin mammary glands of let-7 sensor mice treated with DOX for 4 day. (a) Terminal End buds (b) Ductal structures. Yellow boxes indicate cells that are in close proximity and differentially express let-7 and p63. (c) Indirect immunofluoresence of TEB of control virgin GFP mice stained with GFP (left), and DAPI (middle). (d) Ducts of control mice. Note the mosaic pattern of GFP expression in control mice.

Based on the known function of let-7 in worms and flies, we hypothesized that GFP expression (low let-7 activity) would be restricted to the cap layer of the terminal-end-bud (TEB) in virgin mammary glands. Cap cells of pre-pubescent glands are thought to contain stem cells that divide asymmetrically, giving rise to myoepithelial cells that surround the mammary ducts (Williams and Daniel 1983). Using p63 as a marker for cap and myoepithelial cells, we detected cells within the protruding end of the cap layer that were GFP ^{positive} and p63 ^{positive} (Figure 6 top). We also observed dual GFP ^{positive} p63 ^{positive} cells towards the bottom of the TEB. In certain instances we observe cells, in which a p63 ^{positive} GFP ^{negative} cell is directly adjacent to a GFP ^{positive} p63 ^{negative} cell. I have not detected GFP expression in body cells in the mammary glands analyzed thus far. Interestingly, the stromal cells adjacent to cap layer tended to stain positive for GFP (let-7 low). These cells, which outline the TEB structures and ducts of the mammary gland, play a critical role during early stages of mammary development, and may contain undifferentiated progenitors that repress let-7.

In the ducts of five-week old mice the majority of cells were GFP^{negative}, indicative of let-7 activity in luminal compartment (Figure 6 middle). However, we do observe cells within the ducts that are GFP ^{positive} and p63 ^{negative} (Figure middle), suggestive of let-7 downregulation in these intermingled luminal epithelial cells. Mammary glands from control GFP mice were also explanted (figure 6). A mosaic GFP pattern was observed in the ducts of control mice while in the TEB all cells were stained uniformly expressed GFP.

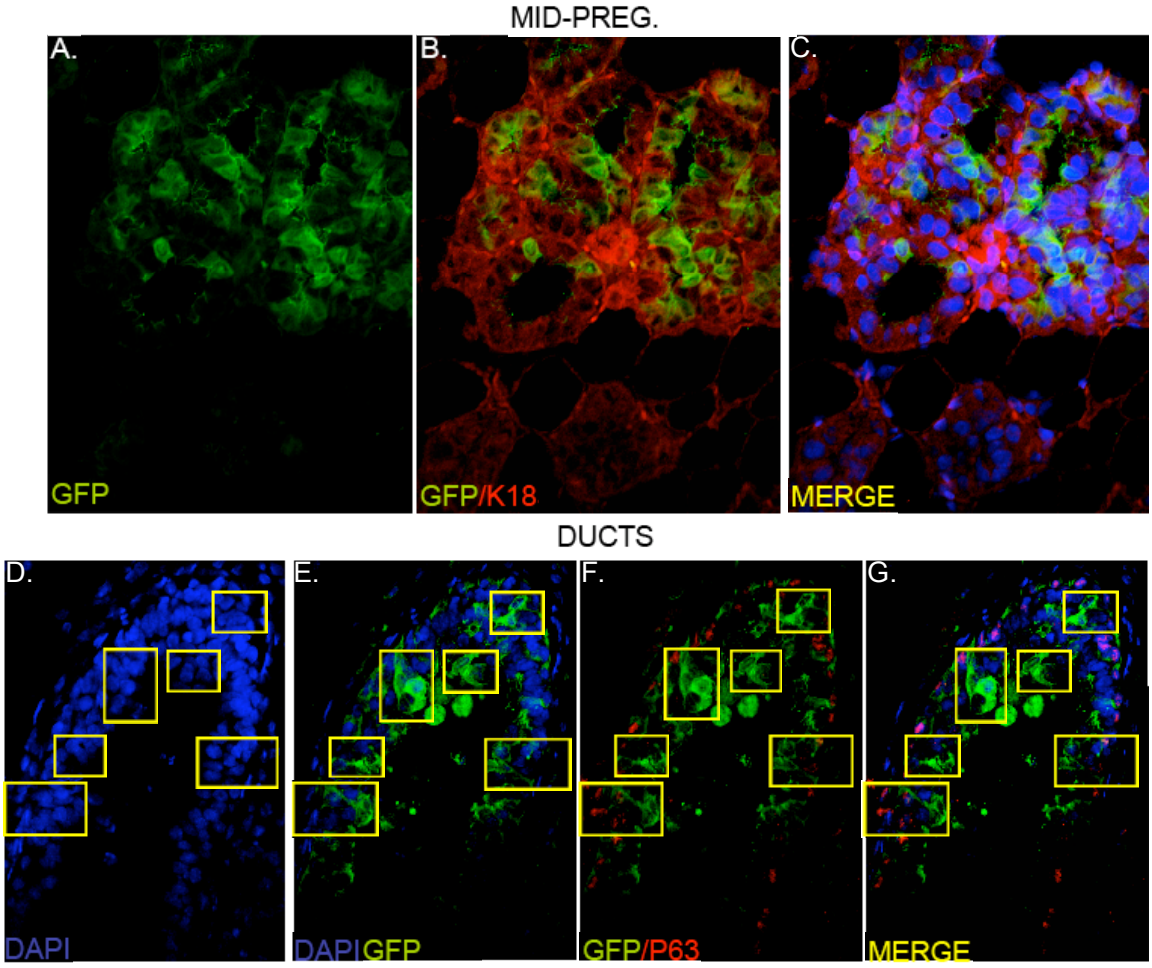


Figure 4.7 Let-7 expression pattern in mid-pregnant mammary glands.

(a-c)Indirect immunofluorescence for K18 (red), GFP (green) and DAPI(blue) in pregnant glands of the let-7 sensor mice treated with Dox for 4 days. GFP expression is observed in alveoli cells of mid-pregnant glands (sensor mice treated with Dox for 4days). (a) Confocal image show the GFP channel (left), K18 and GFP merged (b.), and the merge with nuclei stained blue (c). Note that compared to virgin glands, there is an increased number of GFP+ cells observed. (d-g) Confocal Images of ducts of pregnant sensor mice showing a predominately luminal cell staining pattern (p63 negative). Yellow boxes indicate the brightest GFP cell observed.

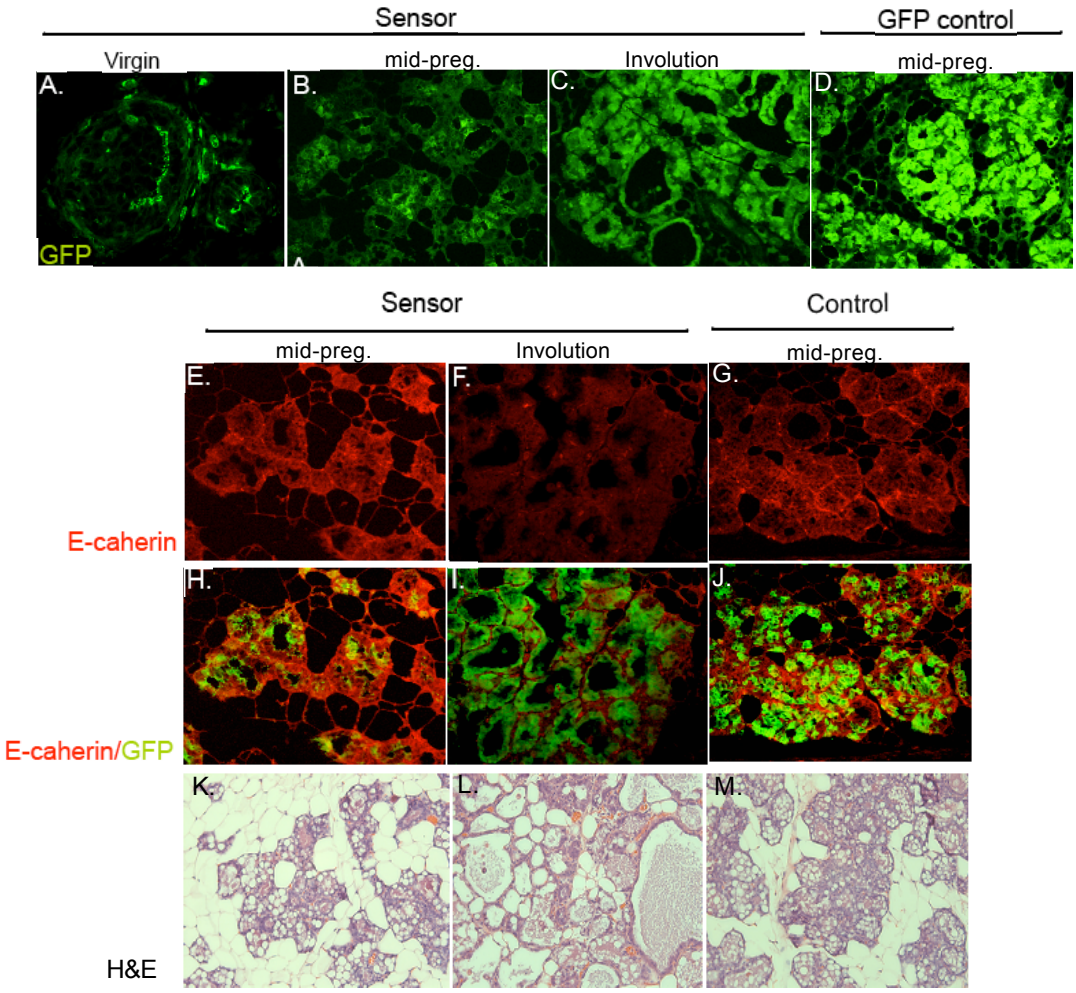


Figure 4.8 Spatial Expression Patterns of Let-7 and E-cadherin during Mammary Gland Differentiation. Immuno-labeling of paraffin-embedded sections of GFP in let-7 sensor mice (a-c) and control mice (d); 5 week old virgin (a), mid-pregnant (b,d), 24hr involuting gland (c). Note a dramatic increase in the number of GFP+ (low let-7) cells during involution. (e-g) Immuno-labeling of E-cadherin in mid-pregnant (e) and 24hr involution (f), of let-7 sensor mice and mid-pregnant control mice (g). Note E-cadherin expression is reduced during involution coinciding with Let-7 down-regulation. (h-j) merged GFP(green) and E-caderin channels (red). (k-m) H&E staining of mammary glands from let-7 sensor and control mice showing normal morphology.

Figure 7 shows the expression of GFP+ cells in alveolar cells of a mid-pregnant female. Interestingly, the percentage of GFP staining cell is increased compared to virgin glands. Figure 8a shows a summary of let-7 expression patterns in virgin, pregnant and involution stages of mammary gland development. At involution, a stage characterized by massive tissue remodeling, we observed the majority of cells in the mammary gland to be GFP + (figure 8c).

This observation prompted an investigation of E-caderin expression during pregnancy and involution, as E-cadherin plays a role in alveolar differentiation and adhesion-dependent survival (figure 8e,f). E-cadherin has also been shown to be important for epithelial differentiation and to decrease during involution (Forster et al. 2002). E-cadherin expression is nearly gone at 24hr involution, coinciding with the loss of let-7 expression (figure 8f). This is an interesting result, because changes in the normal expression pattern of the E-cadherin/catenin complex have been found in various human cancers. In breast cancer, partial or total loss of E-cadherin expression is associated with the loss of the differentiated state, acquisition of invasiveness, increased tumor grade, metastatic behavior and poor prognoses (Heimann et al. 2000) , (Siitonen et al. 1996) (Oka et al. 1993). Therefore, I am currently investigating let-7 expression during later stages of involution, when the mammary gland reverts to a virgin morphological state. It will also be interesting to see whether enforced let-7 expression can inhibit the involution of the mammary gland.

4.3 Discussion:

We now appreciate that miRs can act in a dose-dependent and cell-type specific manner (Lagos-Quintana et al. 2002), (Sokol and Ambros 2005). MicroRNA-sensors, designed to detect the specific miRNA activity, have proven to be useful tools for characterizing expression patterns and function in biologically relevant contexts (Brennecke et al. 2003), (Rybak et al. 2008). To elucidate the expression pattern of let-7 in various tissues, we generated a let-7 sensor mouse model that represses eGFP expression upon let-7-mediated silencing. Given the implicated role of let-7 in cellular differentiation (Yu et al. 2007a), (Wulczyn et al. 2007), (Ibarra et al. 2007) it will likely serve as an important tool.

Methods that take advantage of the slow-cycling nature of HSC have been devised to isolate self-renewing cells from their differentiated progeny (Lerner and Harrison 1990) (Van Zant 1984). Only one published mouse model generated in Elaine Fuchs laboratory, using the histone 2B (H2B)–green fluorescent protein (GFP) fusion protein, has been developed for these purposes (Tumbar et al. 2004). The epithelial stem cell niche in the skin (Tumbar et al. 2004) and more recently, hematopoiteic stem cells (Foudi et al. 2009) were shown to differentially retain GFP and thus served as a niche marker (skin) and enrichment marker (HSC) for the stem/progenitor cells.

The use of the 8-nt LNA inhibitors will prove extremely powerful in terms of revealing the functions of miRs. Remarkably, the targets for most miRNA molecules are unknown, as it has been difficult to reliably suppress miRNA function. Loss of function experiments using cholesterol-conjugated 31-nucleotide LNAs have proven useful in some studies (Krutzfeldt et al. 2005) (Cheng et al. 2005a), but are problematic in cell-

types that are sensitive to lipid- based transfection reagents, and require large amounts for efficacy. The specificity of 8-nt LNA is still under study by our lab and Santa Pharma, but the let-7 sensor mouse has proven quite useful in terms of titrating the minimum dose needed to de-repress GFP reporter and induce a loss-of-function phenotype. Furthermore, the let-7 sensor will also prove useful in testing the tissue availability and penetration of the LNA inhibitors, as let-7 expression can be detected in a wide variety of tissues.

The reprogramming of differentiated cell-types into pluripotent cells has become an active area of stem cell research due to the enormous therapeutic benefits that could be gained. Lin-28 has proven to be an essential component of cellular reprogramming (Yu et al. 2007b), and is a negative regulator of let-7 transcription (Newman et al. 2008, Viswanathan et al. 2008). This strongly suggests that the down-regulation of let-7 might be required to achieve a pluripotent state. Thus the generation of mice that reports the activity of let-7 may facilitate the search for genes/ cells that are amendable to reprogramming.

Though *in vitro* methylcellulose cultures demonstrated the differentiation potential of GFP+ cells, it remains to be seen if transplantation of these cells into irradiated mice can engraft and provide short-term and long-term repopulation of the bone marrow. In collaboration with Camila Dos Santos, a postdoctoral fellow in our lab, we have performed bone marrow transplants of 500 GFP+ sorted cells. However, only one animal out of 25 transplanted mice survived. In the single survivor, donor cells from sensor mice contributed to 30% of bone marrow cells (marked by CD45.2 expression).

These experiments require further optimization, as the animals are dying after 2 weeks of irradiation due to bacterial infections.

The mammary gland is the optimal tissue to study stem cell maintenance, mobilization and differentiation in response to hormonal controls. When coupled with the characterization of let-7 expression at various stages in its development, it will undoubtedly lead to a better understanding of miRNA regulation by hormonal response, as well as its role in regulating stem/progenitor cell renewal and differentiation. Given the correlation between undifferentiated cells and low let-7 expression levels, it might be possible to use differential let-7 expression to identify and enrich for MESC during the different mammary developmental transitions.

4.4 Methods:

Molecular Cloning and Gene Targeting in ES cells

We used a PCR method to introduce two let-7 complementary sequences with PmeI and SnabI as spacers into the 3'UTR of EGFP using the following primers:
Forward: 5' TGA CGgaattcC CAC AAC CAT GGT GAG CAA G.
Reverse: 3' TAGCTATGGCAGCTG TTGGTATGTTGGATGATGGAGT caaatttg atgcatTTGGTATGTTGGATGATGGAGT cttaag AT GA 5'
The GFP control mouse was made using the same forward primer and the following reverse primer: 3' TAGCTATGGCAGCTG cttaag AT GC 5'
PCR products were subcloned into the EcoRI site of pBS31. Electroporation and screening of ES cells were performed previously described by Beard et al. 2006. Tetraploid blastocyst injections were performed as previously described in Eggan et al. 2001. The following genotyping primers were used rtTA-Fwd: atggtgtcggtatcgaaggc reverse rtTA-Rev: cagcaggcagcatatcaagg. For GFP allele we used same primers as above.

Cell Culture

ES cells and Embryoid bodies were cultured following protocols obtained in Nagy, 2003. Briefly, ESC were cultured in DMEM Knockout media (Gibco), supplemented with 15% Fetal Bovine serum(Gibco), LIF (Chemicon), glutamine, ß-mercapthaethanol and pennicillin/ streptomycin. For embryoid cultures, we grew ESC on low-attachment plates (Corning) in the absence of LIF, and induced differentiation with retinoic acid (Sigma). Methylcellulose –based media was obtained from Stem Cell Technologies and followed manufacture protocol.

FACS analysis

Tissues were dissociated using Collagenase I or IV (Sigma). Bone Marrow cells were flushed using a 27-gauge syringe and passed through a 40um filter. Cells were analyzed on an AriaII. Immunophenotyping antibodies were obtained from ebioscience: Ckit-APC, CD-150-PE, CD48-PeCy5.5, Sca-1-PeCy.7.

Immunofluoresence

Tissues were fixed in formalin and embedded in paraffin. 5um sections were generated using a microtome. Slides were cleared using Xylgene, and antigen retrieval was performed using Trilogy (Protocol) and a pressured cooker. The following antibodies were used: anti-chicken GFP (1:300) Invitrogen, anti-mouse Cytokeratin 18 (1:100) (Sigma), anti-rabbit p63(1:100)(Santa Cruz), anti-mouse E-cadherin (1:100) (Santa Cruz).

miRNA expression analyses

Mature miRNAs were quantified using the TaqMan MicroRNA Assays previously described by (Chen et al. 2005) (Applied Biosystems). Data was normalized to U6 RNA. Each experiment was done in triplicate.

Chapter 5

Probing tumor phenotypes using stable and regulated synthetic microRNA precursors

Ross A Dickins[1,2], Michael T Hemann[2], Jack T Zilfou[2], David R Simpson[3], Ingrid Ibarra[2], Gregory J Hannon[1–3] & Scott W Lowe[1–3]

The authors from Cold Spring Harbor Laboratory conducted construction and characterization of p53.1224. The western blot experiments in this chapter were performed by Ingrid Ibarra.

[1]Howard Hughes Medical Institute, [2]Cold Spring Harbor Laboratory, [3]Watson School of Biological Sciences, 1 Bungtown Road, Cold Spring Harbor, New York 11724, USA. Correspondence should be addressed to S.W.L. (lowe@cshl.edu).

NATURE GENETICS VOLUME 37 [NUMBER 11 [NOVEMBER 2005 1 2 8 9

RNA interference is a powerful method for suppressing gene expression in mammalian cells. Stable knock-down can be achieved by continuous expression of synthetic short hairpin RNAs, typically from RNA polymerase III promoters. But primary microRNA transcripts, which are endogenous triggers of RNA interference, are normally synthesized by RNA polymerase II. Here we show that RNA polymerase II promoters expressing rationally designed primary microRNA–based short hairpin RNAs produce potent, stable and regulatable gene knock-down in cultured cells and in animals, even when present at a single copy in the genome. Most notably, by tightly regulating Trp53 knock-down using tetracycline-based systems, we show that cultured mouse fibroblasts can be switched between proliferative and senescent states and that tumors induced by Trp53 suppression and cooperating oncogenes regress upon re-expression of Trp53. In practice, this primary microRNA–based short hairpin RNA vector system is markedly similar to cDNA overexpression systems and is a powerful tool for studying gene function in cells and animals. The sequence specificity of RNA interference (RNAi) allows experimental gene inhibition by introduction of synthetic double-stranded RNA into cells. Stable suppression can be achieved by genomic integration, often through retroviral transduction, of expression cassettes producing short hairpin RNAs (shRNAs) that are processed into small RNA duplexes[1,2]. Early-design shRNAs have stems of 19–29 nucleotides (nt) with minimal flanking RNA sequence (Fig. 1a) and are traditionally expressed from strong RNA polymerase III (Pol III) promoters, such as those that drive expression of the endogenous U6 small nuclear RNA (snRNA) or the H1 RNA. This stem-loop style of shRNA can also be expressed from high-copy polymerase II (Pol II)

promoters, but only in the context of a minimal RNA[3–6]. In contrast to early-design shRNAs, naturally occurring primary microRNA precursors fold extensively beyond the core stem-loop structure (Fig. 1a). Using the entire miR-30 precursor RNA as a template, Cullen and colleagues substituted miR-30 stem sequences and showed effective and regulated target gene inhibition in transient transfection assays[7,8]. Furthermore, miR-30–based shRNAs inhibit gene expression more potently than traditional stem-loop shRNAs, again in transient assays[9,10]. These observations prompted the construction of genome-wide libraries of rationally designed shRNAs based on the primary miR-30 RNA, which currently target more than 28,000 human genes and 25,000 mouse genes[10] (RNAi Codex; seeURL in Methods). In this study, we found that miR-30–based shRNAs (called shRNA-mirs) could stably suppress gene expression when driven by Pol II promoters present at a single copy in the genome. Furthermore, we used tetracycline-responsive promoters to conditionally express shRNA-mirs in vitro and *in vivo*, illustrating the potential of this knock-down system for studying gene function.

RESULTS

Different styles of shRNA have distinct vector preferences

To determine whether shRNA-mirs could stably suppress gene expression in mammalian cells, we recovered a shRNA-mir called p53.1224 (the predicted small interfering RNA (siRNA) begins at nt 1,224 of the mouse Trp53 cDNA) from the shRNA-mir library[10]. We subcloned a U6 promoter-p53.1224 cassette into a murine stem cell virus (MSCV) and a self-inactivating (SIN) retroviral vector, generating MSCV/LTRU6miR30-PIG (LUMP)-p53.1224 or SIN-U6miR30-PIG (UMP)-p53.1224 constructs (Fig. 1b). We also

constructed similar vectors expressing a standard stem-loop shRNA targeting Trp53 (p53C)[11], producing MSCV/LTR-U6shRNA-PIG (LUSP)-p53C or SINU6shRNA-PIG (USP)-p53C constructs (Fig. 1b). We introduced all four constructs into early-passage mouse embryonic fibroblasts (MEFs). We assessed the resulting cell populations for knock-down of Trp53 expression after treatment with adriamycin (a DNAdamaging agent that stabilizes Trp53) and for the ability to form colonies when plated at low density (a functional read-out of Trp53 loss). The MSCV-based p53.1224 shRNA-mir (LUMP-p53.1224) was more effective at suppressing Trp53 than its SIN-based counterpart (UMP-p53.1224), producing nearly undetectable Trp53 levels as assessed by immunoblotting (Fig. 1c). Conversely, the SIN-based p53C shRNA (USP-p53C) was more effective at suppressing Trp53 than its MSCV-derived counterpart (LUSP-p53C), verifying that the U6 promoter is sufficient for stem-loop shRNA expression (Fig. 1c). The ability of each vector to suppress Trp53 correlated precisely with its ability to stimulate colony formation at low density, with cells expressing the MSCV-based p53.1224 vector producing as many colonies as Trp53-null cells (Fig. 1d). This vector preference was also observed for several other shRNA-mirs and stem-loop shRNAs targeting diverse gene products (data not shown). As noted for stem-loop shRNAs expressed from Pol III promoters, however, not every shRNA-mir was effective at knocking-down target gene expression in the MSCV-based context (data not shown). Nevertheless, miR-30–based shRNAs can be potent when stably expressed from retroviral vectors, particularly those with a functional 5' long terminal repeat (LTR).

Effective shRNA-mir expression from a Pol II promoter

The more potent knock-down produced by shRNA-mirs expressed from the MSCV vector compared with the SIN vector implies that the 5' LTR contributes to optimal shRNA-mir expression. To determine whether the LTR promoter, a strong Pol II promoter, is sufficient for effective gene knock-down using shRNA-mirs, we introduced the p53.1224 shRNA into an MSCV vector lacking a U6 promoter (MSCV/LTRmiR30-PIG (LMP); Fig. 1b). To facilitate comparison, we introduced this vector and its LUMP- and UMP-based counterparts into NIH3T3 cells at a low multiplicity of infection (<5% efficiency) such that most transduced cells contained single- or low-copy proviral integrations, as confirmed by Southern blotting (data not shown). Viral titers were similar for empty vector and shRNA-mir constructs (data not shown). Notably, both vectors carrying the MSCV LTR (LUMP-p53.1224 and LMP-p53.1224) suppressed Trp53 expression extremely efficiently and were far superior to UMPp53.1224, which expresses p53.1224 from the U6 promoter alone (Fig. 1e). Therefore, transcription of shRNA-mir cassettes from Pol II promoters is sufficient for highly effective knock-down of a target gene, even when expressed at a single copy. This feature is essential for knock-down screens using complex libraries where infected cells are unlikely to contain multiple copies of a given shRNA vector. Because Pol II transcribes endogenous primary microRNAs[12,13], the improved performance of these vectors seems to derive from natural mechanisms of RNA-dependent gene inhibition. Stable expression of stem-loop shRNAs can produce loss-of function phenotypes in mice[11,14]. To determine whether shRNAmirs expressed from Pol II promoters can efficiently modulate gene

expression *in vivo*, we targeted genes for which the null phenotype was known. For example, inactivation of the BH3-only protein Bcl2l11 (also called Bim, a proapoptotic member of the Bcl2 family) accelerates lymphomagenesis in Eμ-Myc transgenic mice[15]. Therefore, we tested whether shRNA-mir–mediated suppression of Bim would also cooperate with Myc during lymphomagenesis. Mice reconstituted with Eμ-Myc hematopoietic stem cells (HSCs) transduced with two independent shRNA-mirs targeting Bim (collectively called shBim and expressed from the LTR of MSCV/LTRmiR30-SV40-GFP (LMS)) formed tumors much more rapidly than mice reconstituted with HSCs expressing a control vector ($P < 0.05$; Fig. 2a). Lymphomas arising in mice transduced with shBim were green fluorescent protein (GFP)-positive, expressed low levels of Bim (Fig. 2b) and had a mature (IgM+) B-cell phenotype uniquely characteristic of Bim deficient lymphomas15 (data not shown). Therefore, *in vivo* loss-of function phenotypes can be recapitulated using miR-30–based shRNAs expressed from Pol II promoters.

siRNAs have been used to identify modulators of drug action but are not suitable for long-term assays or *in vivo* systems. In principle, miR- 30–based vectors should enable the identification and characterization of genes that modify drug responses in vivo. To test this possibility, we examined the ability of a Trp53 shRNA-mir to promote chemoresistance in Eμ-Myc lymphomas, which respond poorly to therapy in the absence of Trp53 (ref. 16).

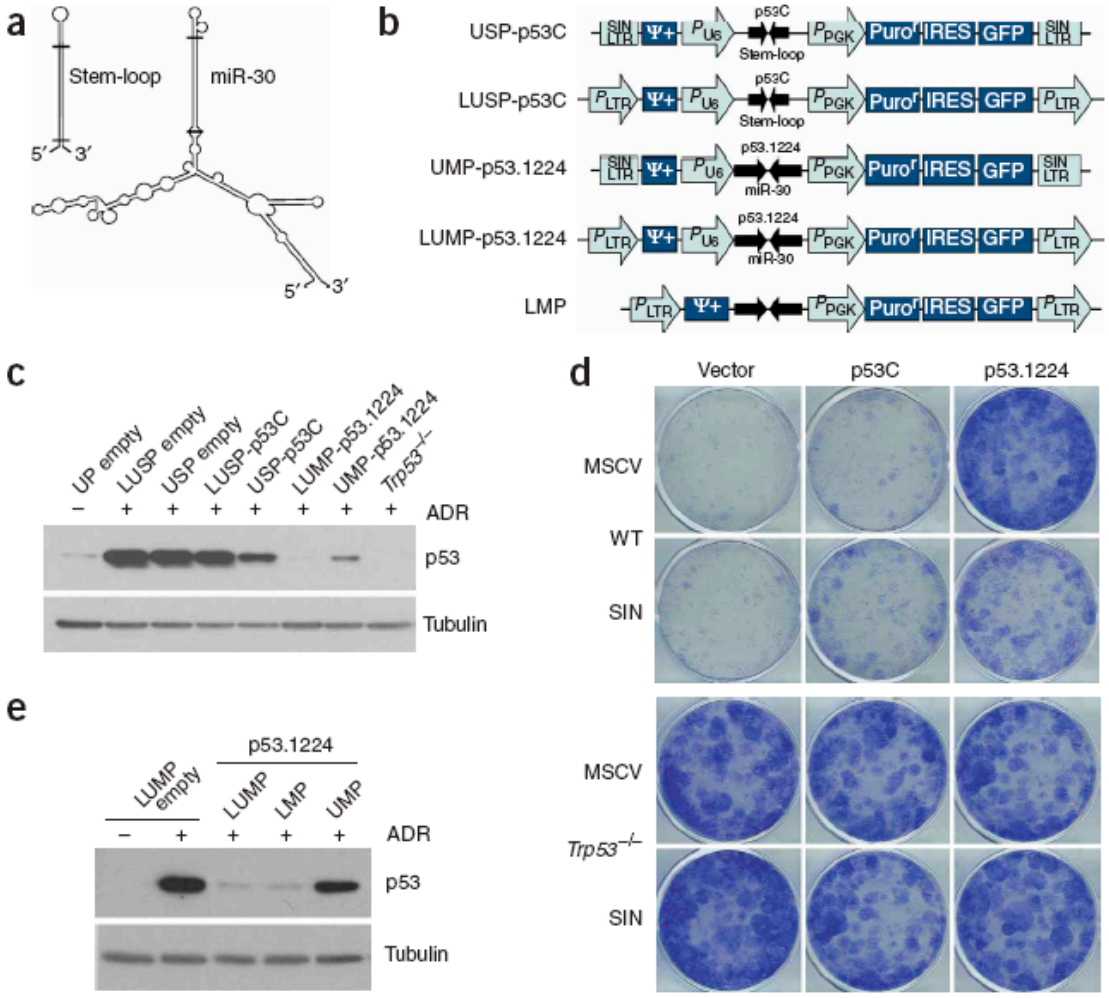


Figure 5.1 Effective knock-down by single-copy expression of miR-30–based shRNAs from a retroviral LTR promoter. (a) Schematic representation of predicted RNA folds for simple stem-loop and miR-30–based shRNAs. Folds are based on the 29-nt stem-loop p53C shRNA9 and the miR-30–based p53.1224 shRNA. Note
extensive predicted folding for the B300-nt miR-30–based RNA. Approximate Drosha and Dicer cleavage sites are indicated by horizontal lines. Folds were generated using mfold35. (b) Retroviral vectors used to deliver shRNAs to mammalian cells. Provirus layouts are shown to indicate promoter activity of the integrated virus. Active promoters are shown as open arrows, with two inverted black arrows representing shRNA stem sequences. (c) Western-blot analysis for Trp53 expression of NIH3T3 cells transduced with the retroviral vectors shown in b and selected in puromycin. A tubulin blot is shown as a loading control. ADR, adriamycin. (d) Colony formation assay for the cells shown in c. Cells were seeded in six-well plates at 2,500 cells per well and allowed to grow for 10 d before fixation. (e) Western-blot analysis for Trp53 expression in NIH3T3 cells transduced at less than 5% efficiency (assessed by GFP FACS; data not shown) with the retroviral vectors shown in b. A tubulin blot is shown as a loading control. Similar results were obtained in other cell types, including wild-type (WT) and Cdkn2a (p19ARF)- null MEFs (data not shown). ADR, adriamycin.

We introduced LMS-p53.1224 (coexpressing GFP) into chemosensitive Eμ-Myc lymphoma cells at ~10% infection efficiency and transplanted the mixed cell populations into syngeneic recipient mice. Upon lymphoma manifestation, we treated mice with the chemotherapeutic drug adriamycin and monitored their tumor response. Notably, mice carrying lymphomas expressing the p53.1224 shRNA-mir did not regress after treatment with adriamycin and had significantly lower overall survival than control tumor-bearing mice (Fig. 2c). Furthermore, the percentage of GFP-positive cells markedly increased in lymphomas carrying p53.1224 but not the control vector, indicative of a selective advantage for cells expressing p53.1224 (Fig. 2d). Despite extensive previous efforts with stem-loop shRNAs (data not shown), this is the first time to our knowledge that sufficient Trp53 knock-down to promote chemoresistance *in vivo* has been achieved. Taken together, these data indicate that miR-30–based shRNAs expressed from Pol II promoters are suitable for a variety of *in vivo* applications, perhaps including tissue specific gene knockdowns and *in vivo* forward genetic screens.

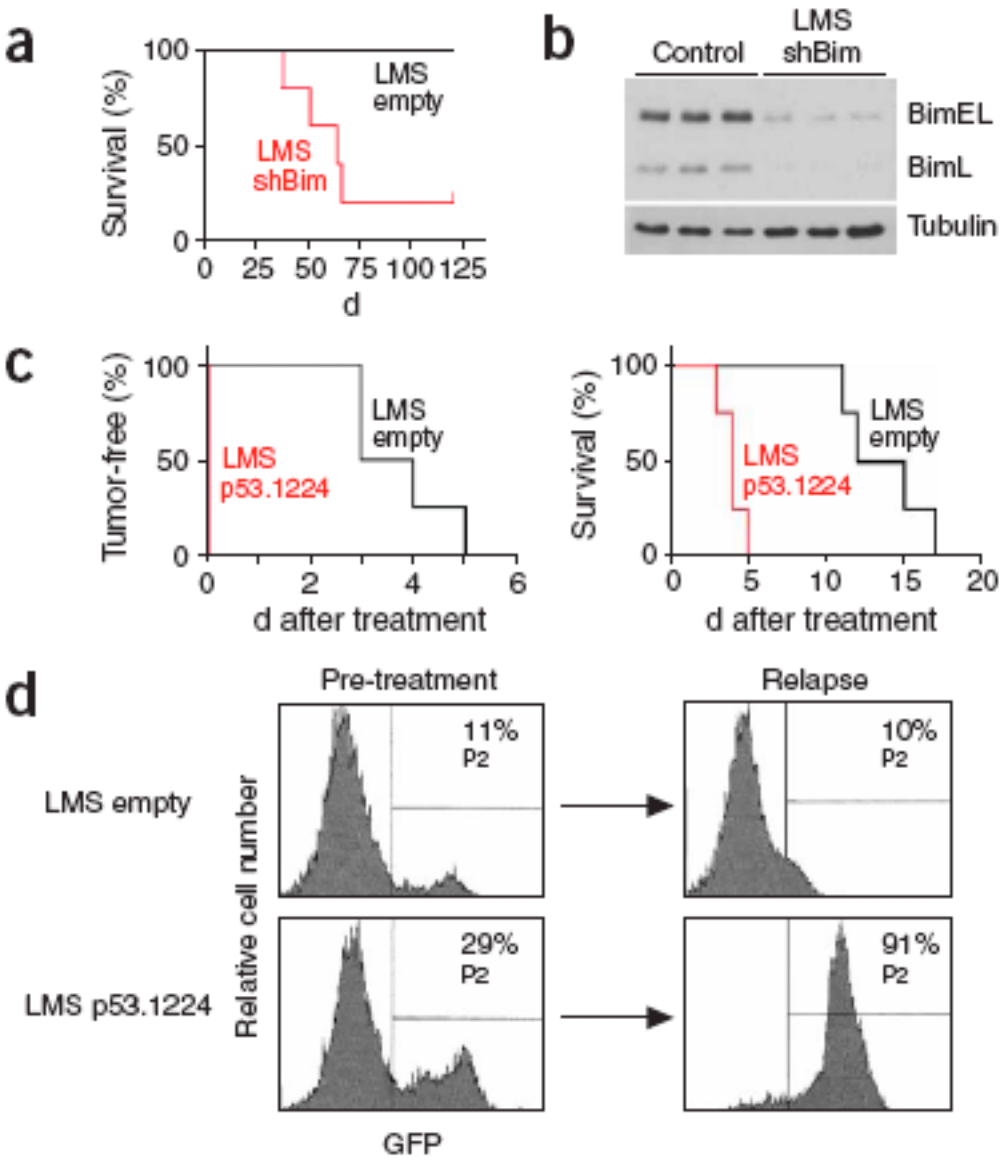


Figure 5.2 RNA Pol II–driven shRNAs can effectively promote tumorigenesis and chemotherapy resistance in vivo. (a) Kaplan-Meier curve showing mouse survival after adoptive transfer of Em-Myc HSCs infected with LTR-driven Bim shRNAs. (b) Western blot showing reduced expression of two Bim protein isoforms, BimEL and BimL, in Em-Myc lymphomas expressing Bim shRNAs. Archived tumors arising from Em-Myc HSCs (on a wild-type, Cdkn2a(p19ARF)+/! or Trp53+/! background; data not shown) were used as controls for Bim expression. (c) Kaplan-Meier curves showing tumor-free survival (left) and overall survival (right) for mice carrying Cdkn2a (p19ARF)- null lymphomas infected with either the LMP-p53.1224 retrovirus or vector control. Tumor-bearing mice were given a single dose of 10 mg adriamycin per kg body weight at day zero. (d) Flow cytometry analysis of GFP expression in lymphoma cells from the mice in c. Representative histograms show the percent of GFP-positive cells at the time of treatment (left) and after tumor relapse (right).

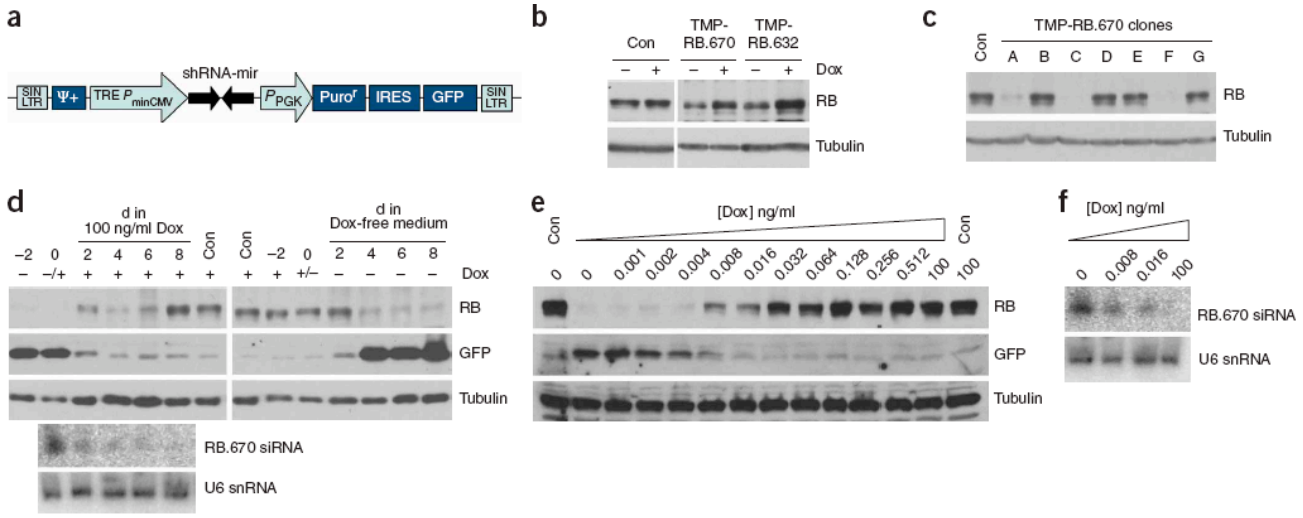


Figure 5.3 Stable and regulatable shRNA expression from a tetracycline-responsive RNA Pol II promoter. (a) Provirus layout of the SIN-TREmiR30-PIG (TMP) retroviral vector. (b) Western-blot analysis of RB1 expression in HeLa-tTA cells infected with TMP-RB.670. Cells were treated with 100 ng ml^{-1} doxycycline (Dox) for 4 d before collection. Control uninfected HeLa-tTA cells (Con) treated with doxycycline are also shown. (c) RB1 expression in homogeneous cultures derived from single-cell clones of HeLa-tTA cells infected at single-copy with TMP-RB.670. Cells were cultured in normal, doxycycline-free medium before collection. Con, control uninfected HeLa-tTA cells. (d) RB1 and RB.670 siRNA expression in HeLa-tTA clone RB.670C cells over time in response to shifting into or out of doxycycline (Dox). Cells were cultured without doxycycline (left panels) or in 100 ng ml^{-1} doxycycline (right panels) for 8 d before being shifted into 100 ng ml^{-1} doxycycline or doxycycline-free medium, respectively. Note the presence of a faint nonspecific band in the GFP immunoblot (longer exposures showed weak GFP expression, probably produced by the PGK-driven transcript). Similar results were observed for all RB.670 clones showing good RB1 knock-down in doxycycline-free medium (c, above), with some clonal variation in kinetics. In addition, similar results were obtained using other shRNAmirs targeting RB1 and Pten (data not shown). Lower panels show small RNA northern blots detecting the RB.670 siRNA, correlating with RB1 knock-down. U6 snRNA was used as a loading control. Con, control uninfected HeLa-tTA cells. (e) Doxycycline (Dox) dose-response analysis of RB1 expression in HeLa-tTA clone RB.670C. Cells were cultured for 8 d in the indicated doxycycline concentration before collection. Control uninfected HeLa-tTA cells (Con) cultured with or without doxycycline are also shown. Note the presence of a nonspecific band in the GFP immunoblot, running just below GFP. (f) RB.670 siRNA production in HeLa-tTA clone RB.670C cells cultured in the indicated concentration of doxycycline (Dox), determined by small RNA northern blotting with U6 snRNA as a loading control.

Tetracycline-regulated shRNA-mir expression

Inducible expression of protein-coding cDNAs in cultured cells and animals has revolutionized analysis of gene function, and reversible loss-of-function systems using regulated shRNA expression hold similar promise. Simple stem-loop shRNAs can be expressed conditionally from modified Pol III promoters[17–24], but the inducibility of these systems is variable and has not been shown *in vivo*. Given that low-copy Pol II promoters can effectively drive expression of shRNA-mirs, we reasoned that the widely used tetracycline-regulated Pol II promoter TRE-CMV[25] could be adapted to stably express shRNA-mirs. Using a SIN vector backbone, we cloned a shRNA-mir targeting human RB1 (RB.670) downstream of the TRE-CMV promoter, producing SIN-TREmiR30-PIG (TMP-RB.670; Fig. 3a). We transduced HeLa cells stably expressing the tetracycline transactivator protein tTA (tetracycline-off) with TMP-RB.670 at a low multiplicity of infection (<10% efficiency) and expanded them in the absence of the tetracycline analog doxycycline. Although RB1 levels were only modestly suppressed in cell populations (Fig. 3b), 6 of 13 single-cell clones isolated from these populations showed excellent RB1 knock-down (Fig. 3c and data not shown), showing that the TRE-CMV promoter can effectively drive shRNA-mir expression at low copy number.

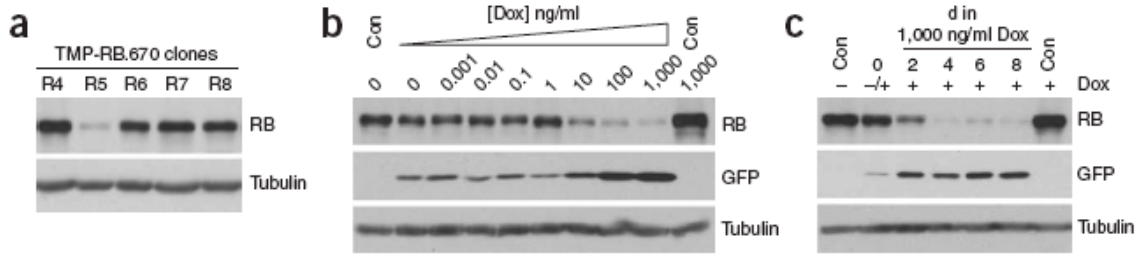


Figure 5.4 Regulated shRNA expression using a tetracycline-on system. (a) RB1 expression in homogeneous cultures derived from single-cell clones of U2OS-rtTA cells infected at ~1% efficiency with TMP-RB.670. Cells were cultured in 1,000 ng ml^{-1} doxycycline for several days before collection. (b) Doxycycline (Dox) doseresponse analysis of RB1 expression in U2OSrtTA clone RB.670R5 cells. Cells were cultured for 8 d in the indicated doxycycline concentration before collection. Control uninfected U2OS-rtTA cells (Con) cultured with or without doxycycline are also shown. (c) RB1 expression in U2OS-rtTA clone RB.670R5 cells in response to doxycycline treatment. Cells were cultured without doxycycline (Dox) for 8 d before being shifted into 1,000 ng ml^{-1} doxycycline. RB.670R5 cells express some GFP in doxycycline-free medium, and RB1 levels are slightly decreased compared with controls, indicating slightly leaky expression from the TRE-CMV promoter in this particular clone. Con, control uninfected U2OS-rtTA cells.

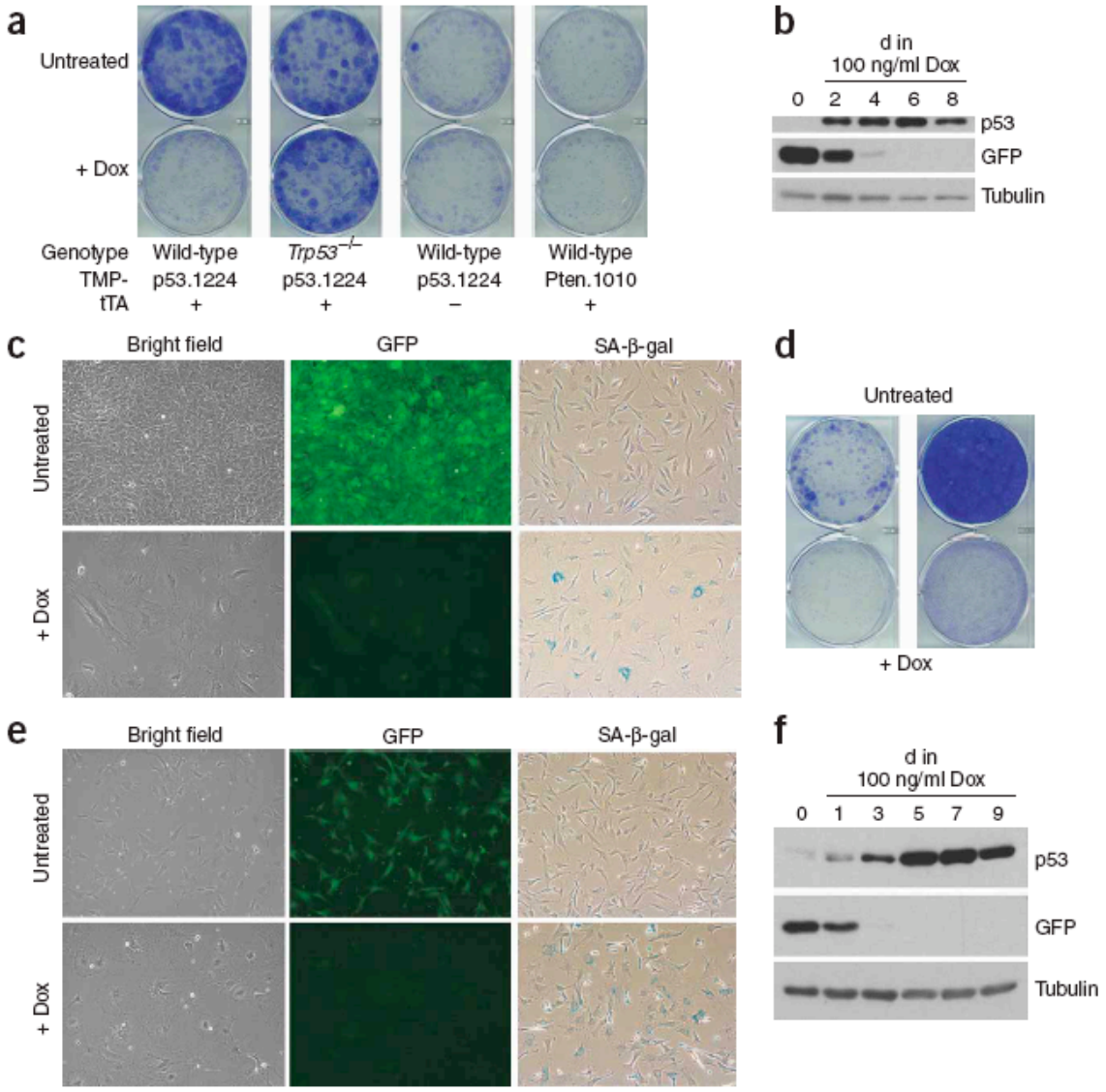


Figure 5.5 Reversible Trp53 knock-down in primary MEFs. (a) Colony formation assays of wild-type MEFs doubly infected with TMP-p53.1224 and tTA. Cells were seeded in six-well plates at 5,000 cells per well and grown for 8 d before collection. Upper wells contained doxycycline free medium, whereas lower wells contained 100 ng ml^{-1} doxycycline (Dox). Positive control Trp53-null MEFs are shown, as are negative control wild-type MEFs infected with TMPp53.1224 alone or with TMP-Pten.1010 plus tTA. (b) Western-blot analysis of Trp53 and GFP expression in cells expanded from a single cell clone of wild-type MEFs infected with TMP-p53.1224 and tTA (WtT cells). Cells were cultured in 100 ng ml^{-1} doxycycline (Dox) for various times before collection. (c) Morphology and GFP fluorescence of WtT-p53.1224 cells originally plated at colony formation density and cultured in doxycycline-free medium (upper panels) or 100 ng ml^{-1} doxycycline (Dox; lower panels). Right, SA-b-gal staining of WtT-p53.1224 cells cultured in doxycyclinefree medium (upper panel) or 100 ng ml^{-1} doxycycline (lower panel). (d) Left, colony formation assay for WtT-p53.1224 cells cultured for 8 d in 100 ng ml^{-1} doxycycline (Dox) then seeded in doxycycline-free medium (upper well) or 100 ng ml^{-1} doxycycline (lower well). Right, colony

formation assay of cells equivalent to those in the upper well of the left panel (formerly doxycycline-treated, dormant WtT-p53.1224 cells after extended culture in doxycycline-free medium). Cells were seeded and collected as described in a. (e) Morphology, GFP fluorescence and SA-b-gal staining of WtT-p53.1224 cells infected with Kras and cultured in normal medium (upper panels) or 100 ng ml^{-1} doxycycline (Dox; lower panels). (f) Western-blot analysis of Trp53 and GFP expression in WtT-p53.1224 cells infected with Kras. Cells were cultured in 100 ng ml^{-1} doxycycline (Dox) for various times before collection.

For all clones showing RB1 knock-down in the absence of doxycycline, RB1 levels returned to normal after approximately one week of treatment with doxycycline (Fig. 3d and data not shown). RB1 expression was dose-dependent, with maximum RB.670 siRNA production and RB1 knock-down achieved in low doxycycline concentrations and vice versa (Fig. 3e,f). In all cases, GFP and RB.670 siRNA production were positively correlated (Fig. 3d,f), whereas GFP and RB1 levels were inversely correlated (Fig. 3d,e), implying that, in this system, GFP expression serves as a surrogate marker of shRNA production. As expected, when we introduced TMP-RB.670 into U20S cells stably expressing the reverse tTA (rtTA; tetracycline-on) protein[26], doxycycline concentration and RB1 knock-down were positively correlated in a subset of single-cell clones (Fig. 4). Therefore, expression of shRNA-mir cassettes from the TRE-CMV promoter allows regulated gene knock-down in tetracycline-on or -off configurations.

We went on to test regulated gene knock-down in primary cells. Inactivation of the tumor suppressor Trp53 immortalizes wild-type MEFs and transforms MEFs overexpressing oncogenic Kras27. We cotransduced early-passage MEFs with TMP-p53.1224 and a retrovirus expressing the tTA (tetracycline-off) protein. We reasoned that many doubly infected MEFs (called wild-type/tTA/TMP-p53.1224 or

WtT-p53.1224) would show stable Trp53 knock-down when cultured in doxycycline-free medium. Indeed, WtT-p53.1224 cells plated at low density formed colonies as efficiently as Trp53-null MEFs (Fig. 5a), suggesting that Trp53 was functionally inactivated in most cells. In stark contrast, colony formation of WtT-p53.1224 cells cultured in doxycycline in parallel was similar to that of control cells (Fig. 5a). Growth of Trp53-null MEFs was unaffected by doxycycline, ruling out nonspecific effects (Fig. 5a). We observed similar results in single-cell clones, where doxycycline treatment produced a rapid reactivation of Trp53 and loss of GFP expression and clonogenic potential (Fig. 5b,c). The cell cycle arrest produced by Trp53 reactivation was associated with a flattened morphology and senescence-associated β-galactosidase (SA-b-gal) staining (Fig. 5c), indicating that loss of Trp53 is required for maintaining the immortalized state.

Because clonogenic assays are highly sensitive to 'leakiness' in the system, the data described above indicate that the shRNA-mir construct was under very tight control. Accordingly, WtT-p53.1224 clones maintained in doxycycline remained dormant for many weeks, retaining high Trp53 and low GFP expression (data not shown). Nevertheless, these senescent cells reacquired GFP expression and resumed proliferating shortly after removal of doxycycline[28] (Fig. 5d), a state that could be completely reversed by re-addition of doxycycline (Fig. 5d). Even WtT-p53.1224 cells transformed by coexpression of oncogenic Kras (Fig. 5e) became morphologically senescent and SA-β-gal–positive when treated with doxycycline (Fig. 5e), with Trp53 and GFP expression changes similar to those observed in parental WtT cells (Fig. 5f). These data indicate that cells can be reversibly switched between cycling and senescent states simply by regulating Trp53 knock-down and that ongoing Trp53 inactivation is required for

immortalization and transformation of MEFs *in vitro*. Because complete Trp53 inactivation is strongly selected for in these models, our ability to re-establish senescence in entire cell populations using the tetracycline-controlled p53.1224 vector highlights the efficiency and tight regulation of the TRE-CMV shRNA-mir system.

Control of tumor growth by regulating Trp53 knock-down

Tetracycline-regulated systems expressing protein-coding cDNAs have been used to study the requirement for continued oncogene expression in tumor maintenance[29]. To test whether tetracycline-controlled expression of shRNA-mir cassettes might enable a similar analysis of tumor-suppressor genes, we examined the requirement for Trp53 inactivation in tumors derived from WtT-p53.1224/Kras MEFs. We injected these transformed cells subcutaneously into the flanks of nude mice, where they formed visible, rapidly growing and strongly GFP positive tumors after B2 weeks (Fig. 6a). In mice treated with doxycycline in their drinking water, however, tumor GFP intensity was markedly and rapidly diminished compared with untreated mice, and after 4 d, most tumors were GFP-negative (Fig. 6a). Notably, tumor growth slowed upon doxycycline treatment, and mice treated with doxycycline for 10 d showed tumor regression and eventually became tumor-free (Fig. 6b).

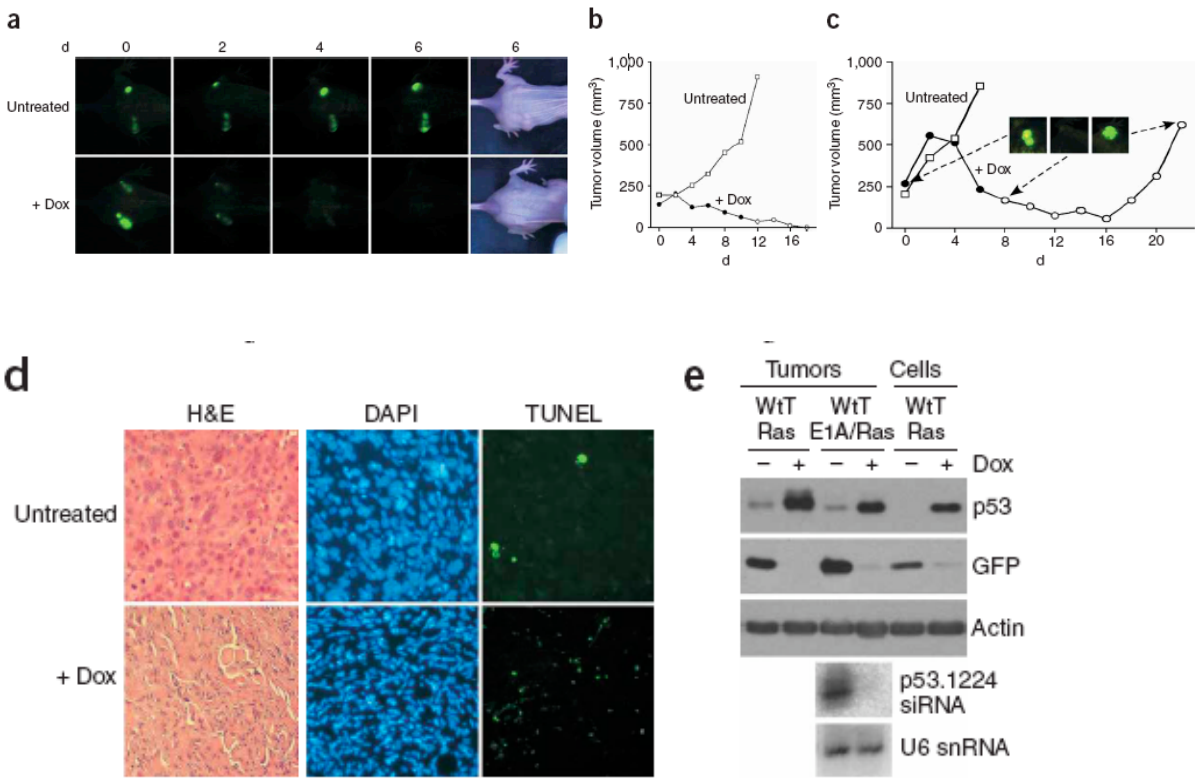


Figure 5.6 Regulated Trp53 knock-down in tumors. (a) GFP and standard imaging of representative tumor-bearing nude mice, with doxycycline (Dox) treatment commencing at day 0 (lower panels). Untreated controls are also shown (upper panels). (b) Representative tumor growth curves for WtT-p53.1224/Kras tumors in an untreated mouse (open squares) or a mouse treated for 10 d with doxycycline (Dox; filled circles indicate doxycycline treatment) commencing at day 0. Each data point is the average volume of two tumors for a single mouse. Similar results were obtained for eight different WtT-p53.1224/Kras clones, with slightly differing kinetics. (c) Representative tumor growth curves for WtT-p53.1224/E1A/Kras tumors in an untreated mouse (open squares) or a mouse treated for 7 d with doxycycline (Dox; filled circles indicate doxycycline treatment) commencing at day 0. Each data point is the average volume of two tumors for a single mouse. Insets show GFP status of a single tumor at various times. (d) Histological analysis of cell morphology and apoptosis in representative nude mouse tumors from untreated mice or mice treated with doxycycline (Dox) for several days. H&E, hematoxylin and eosin. (e) Western-blot analysis of Trp53 and GFP expression in representative WtTp53.1224/Kras and WtT-p53.1224/E1A/Kras tumors from untreated mice or mice treated with doxycycline (Dox) for several days. Cultured WtT-p53.1224/ Kras cells treated with doxycycline are shown as a control.

Tumors obtained from doxycycline-treated mice contained cells with unusually compact nuclei and had widespread apoptosis compared with doxycycline-free controls

(Fig. 6d). Their regression was Trp53-dependent, as Trp53 levels were rapidly restored after doxycycline addition (Fig. 6e), and tumors derived from Trp53-null MEFs infected with tTA, TMP-53.1224 and Kras lost GFP expression but continued to grow upon doxycycline administration (data not shown). Northern-blot analysis of RNA isolated from tumors showed tight regulation of the p53.1224 siRNA by doxycycline *in vivo* (Fig. 6e). In many cases, often when the initial tumors were small, doxycycline-treated mice became tumor-free and remained so for weeks, even upon doxycycline removal (data not shown). But removing doxycycline from mice with larger tumors or after brief doxycycline treatment often allowed renewed GFP expression and tumor growth (Fig. 6c). We obtained similar results for several WtTp53.1224/Kras clones and WtT-p53.1224 cells transformed by adenovirus E1A and Kras, with variable tumor growth rates and regression kinetics (data not shown). Therefore, despite the genomic instability that accompanies Trp53 loss30, ongoing Trp53 inactivation is a requirement for sustained tumor growth in these models. These findings highlight the potential of this technology for the study of many aspects of biology, including identification or validation of potential drug targets in animal models.

DISCUSSION

We generated an integrated system that exploits synthetic primary microRNAs expressed from Pol II promoters to stably and conditionally knock-down genes of interest in mammalian cells. This potent gene knock-down system can be used to study a variety of *in vitro* and *in vivo* processes, including cancer phenotypes such as tumorigenesis

and treatment response. Furthermore, extremely tight conditional gene knock-down can be achieved by expressing shRNA-mir cassettes using tetracycline-responsive systems, including lentiviral systems[31]. As shown here, this facilitates analysis of the role of tumor-suppressor gene expression in the maintenance of immortalized or transformed states and in continued tumor growth *in vivo*. Finally, stable and regulatable gene knock-down can occur when a single copy of the Pol II promoter–shRNA-mir cassette is present in the genome, which is essential for complex library screens and should facilitate many *in vivo* applications. Our data suggest that gene knock-down by expression of shRNAmirs may, in practice, be very similar to overexpression of protein-coding cDNAs. Therefore, expression systems allowing targeted, regulated and tissue-specific expression, which have traditionally been limited to gene over-expression studies, may now be adapted for loss of function analyses. As all vectors described in this work are compatible with genome-wide, sequence-verified banks of miR-30–based shRNAs targeting human and mouse genes[10], they create an important resource for diverse, large-scale RNAi studies in mammalian systems.

METHODS

Vectors. The retroviral vector MSCV-PIG[11] has an EcoRI site in the polylinker and another between the PuroR cassette and the IRES sequence. To facilitate cloning into the polylinker, we destroyed the second site using a PCR-based strategy. We generated a PCR product using MSCV-PIG template (primer sequences available on request), digested it with BglII and MfeI and cloned it into MSCV-PIG digested with BglII and EcoRI, yielding MSCV-PIGdRI. We

made MSCV-U6miR30-PIG by ligating the 762-bp BamHI-MfeI U6-miR30 context cassette from pSM2 (refs. 10,32) into MSCV-PIGdRI digested with BglII and EcoRI. We made MSCV-LTRmiR30-PIG by ligating the 256-bp SalIMfeI miR30 context cassette from pSM2 into MSCV-PIGdRI digested with XhoI and EcoRI. We made MSCV-LTRmiR30-SV40GFP (LMS) in two steps. First, we ligated an EcoRI-ClaI SV40GFP fragment of ~1.2 kb from pBabe GFP into MSCV-puro (Clontech) digested with EcoRI and ClaI, forming MSCVSV40GFP. We then digested this with XhoI and EcoRI and inserted the 256-bp SalI-MfeI miR30 context cassette from pSM2, forming MSCV-LTRmiR30- SV40GFP. We made SIN-PIGdRI by ligating the 2,524-bp EcoRI-SalI fragment from MSCV-PIGdRI into pQCXIX (Clontech) digested with EcoRI and XhoI. We made SIN-TREmiR30-PIG in two steps. First, we generated a PCR product spanning the TRE-CMV promoter using template plasmid pQTXIX (a gift from A. Malina, McGill University, Montreal, Canada), generated by cloning the XbaI-EcoRI TRE-CMV promoter fragment from pUHD10.3 (ref. 25) into pQCXIX (Clontech) digested with XbaI and EcoRI (primer sequences available on request). We then digested this PCR product with BglII and MfeI and ligated it into SIN-PIGdRI digested with BglII and EcoRI (removing its CMVpromoter), forming SIN-TRE-PIG. We completed SIN-TREmiR30-PIG by ligating the 256-bp SalI-MfeI miR30 context cassette from pSM2 into SINTRE-PIG digested with XhoI and EcoRI. We generated DNA fragments encoding various shRNA-mir folds using oligonucleotide template PCR as described previously[32] or subcloned them as 110-bp XhoI-EcoRI fragments from the pSM2 shRNA-mir library[10]. Template sequences are available on request. We designed oligonucleotides using RNAi Central. We digested PCR products with XhoI and EcoRI

and ligated them into the unique XhoI and EcoRI sites in the miR30 context in the vectors described above.

Cell culture and expression analysis. We grew cells in Dulbecco's modified Eagle medium containing 10% fetal bovine serum at 37 1C with 7.5% CO2. We dissolved doxycycline (Clontech) in water and generally used it at a final concentration of 100 ng ml^{-1}. We refreshed medium containing doxycycline every 2 d. We carried out infections and colony formation assays as previously described11.We detected SA-b-gal activity as previously described[27], with sample equilibration and X-gal staining done at pH 5.5. For western-blot analysis, we separated Laemmli buffer protein lysates by SDS-PAGE and transferred them to Immobilon PVDF membrane (Millipore). We used antibodies against p53 (1:1,000 IMX25, Vector Laboratories), Pten (1:1,000 486, a gift from M. Myers, Cold Spring Harbor Laboratory, New York), GFP (1:5,000, Clontech), tubulin (1:5,000 B-5-1-2, Sigma), actin (1:5,000, Sigma) and RB (1:1,000 G3-245, Becton Dickinson with 1:100 XZ-55 and C36 hybridoma supernatants). To isolate total RNA, we lysed cultured cells in 1 ml of Trizol (Invitrogen). We separated 30 mg of RNA on a denaturing acrylamide gel, transferred it to Hybond N+ membrane (Amersham) and probed it with end-labeled oligonucleotides complementary to predicted siRNAs.

Lymphoma studies. We carried out Eμ-Myc lymphomagenesis and drug treatment studies as previously described[11,33]. We isolated chemosensitive lymphoma cells from tumors arising in mice transplanted with Eμ-Myc Cdkn2a(p19ARF)+/- HSCs, which invariably lose the wild-type Cdkn2a allele but retain wild-type Trp53 (ref. 34). The Cold Spring Harbor Laboratory Animal Care and Use Committee approved all mouse experiments described in this work.

Nude mouse tumor analysis. We injected B106 transformed cells subcutaneously into the two rear flanks of nude mice. We treated mice with 0.2 mg ml^{-1} doxycycline in a 0.5% sucrose solution in light-proof bottles, refreshed every 4 d. We calculated tumor volume (mm3) as length X $width^2$ X $\pi/6$. For analysis of protein expression, we snap-froze tumors and pulverized them in liquid nitrogen using a mortar and pestle. We lysed powdered tumor in Laemmli buffer or Trizol for western or northern blotting, respectively. For histology, we fixed tumor tissue for 24 h in 4% formaldehyde in phosphate buffered saline before embedding and sectioning it. We measured apoptosis by TUNEL assay (In situ Cell Death Detection Kit, POD; Roche).

URLs. RNAi Codex is available at http://codex.cshl.edu/. Oligonucleotides were designed using RNAi Central (http://katahdin.cshl.org/siRNA/RNAi.cgi?type?shRNA).

ACKNOWLEDGMENTS

We thank M. Narita for advice on generating vectors; R. Sachidanandam and N. Sheth for shRNA design; A. Malina for the pQTXIX vector; A. Denli, W. Keyes, D. Burgess and A. Bric for experimental assistance; F. Stegmeier and S. Elledge for communicating unpublished results; members of the laboratory of S.W.L. for advice and discussions; and L. Bianco and Cold Spring Harbor Laboratory animal house staff for their assistance. This study was supported by a Mouse Models of Human Cancer Consortium grant and a DNA Tumor Virus grant from the National Cancer Institute. This study was also supported by the Leukemia Research Foundation (R.A.D.), the Helen Hay Whitney Foundation (M.T.H.) and a Ruth L. Kirschstein NRSA (J.T.Z.). D.R.S. is a Beckman Foundation scholar of the Watson School of Biological Sciences. G.J.H. is supported by an Innovator Award from the US Army Breast Cancer Research Program. S.W.L. is an AACR-NCFR Research Professor.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturegenetics/

Literature Cited

1. Brummelkamp, T.R., Bernards, R. & Agami, R. A system for stable expression of shortinterfering RNAs in mammalian cells. *Science* **296**, 550–553 (2002).
2. Paddison, P.J., Caudy, A.A., Bernstein, E., Hannon, G.J. & Conklin, D.S. Short hairpin RNAs (shRNAs) induce sequence-specific silencing in mammalian cells. *Genes Dev*.**16**, 948–958 (2002).
3. Xia, H., Mao, Q., Paulson, H.L. & Davidson, B.L. siRNA-mediated gene silencing in vitro and in vivo. *Nat. Biotechnol*. **20**, 1006–1010 (2002).
4. Unwalla, H.J. et al. Negative feedback inhibition of HIV-1 by TAT-inducible expression of siRNA. *Nat. Biotechnol*. **22**, 1573–1578 (2004).
5. Ling, X. & Li, F. Silencing of antiapoptotic survivin gene by multiple approaches of RNA interference technology. *Biotechniques* **36,** 450–454, 456–460 (2004).
6. Song, J. et al. Gene silencing in androgen-responsive prostate cancer cells from the tissue-specific prostate-specific antigen promoter. *Cancer Res*. **64**, 7661–7663 (2004).
7. Zeng, Y., Wagner, E.J. & Cullen, B.R. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol. Cell* **9**, 1327–1333 (2002).
8. Zeng, Y., Cai, X. & Cullen, B.R. Use of RNA polymerase II to transcribe artificial microRNAs. *Methods Enzymol.* **392**, 371–380 (2005).
9. Boden, D. et al. Enhanced gene silencing of HIV-1 specific siRNA using microRNA designed hairpins. Nucleic Acids Res. **32**, 1154–1158 (2004).
10. Silva, J.M. et al. Second-generation shRNA libraries covering the mouse and human genomes. Nat. Genet. advance online publication 2 October 2005 (doi:10.1038/ng1650).
11. Hemann, M.T. et al. An epi-allelic series of p53 hypomorphs created by stable RNAi produces distinct tumor phenotypes in vivo. *Nat. Genet.* **33**, 396–400 (2003).
12. Lee, Y. et al. MicroRNA genes are transcribed by RNA polymerase II. *EMBO J.* **23**, 4051–4060 (2004).
13. Cai, X., Hagedorn, C.H. & Cullen, B.R. Human microRNAs are processed from capped,polyadenylated transcripts that can also function as mRNAs. *RNA* **10**, 1957–1966 (2004).
14. Rubinson, D.A. et al. A lentivirus-based system to functionally silence genes in primary mammalian cells, stem cells and transgenic mice by RNA interference. *Nat. Genet.***33**, 401–406 (2003).
15. Egle, A., Harris, A.W., Bouillet, P. & Cory, S. Bim is a suppressor of Myc-induced mouse B cell leukemia. *Proc. Natl. Acad. Sci. USA* **101**, 6164–6169 (2004).
16. Schmitt, C.A., McCurrach, M.E., de Stanchina, E.,Wallace-Brodeur, R.R. & Lowe, S.W. INK4a/ARF mutations accelerate lymphomagenesis and promote chemoresistance by disabling p53. *Genes Dev.* **13**, 2670–2677 (1999).
17. van de Wetering, M. et al. Specific inhibition of gene expression using a stably integrated, inducible small-interfering-RNA vector. *EMBO Rep.* **4**, 609–615 (2003).
18. Miyagishi, M., Sumimoto, H., Miyoshi, H., Kawakami, Y. & Taira, K. Optimization of an siRNA-expression system with an improved hairpin and its significant suppressive effects in mammalian cells. *J. Gene Med.* **6**, 715–723 (2004).

19. Czauderna, F. et al. Inducible shRNA expression for application in a prostate cancer mouse model. *Nucleic Acids Res.* **31**, e127 (2003).
20. Chen, Y., Stamatoyannopoulos, G. & Song, C.Z. Down-regulation of CXCR4 by inducible small interfering RNA inhibits breast cancer cell invasion in vitro. *Cancer Res.* **63**, 4801–4804 (2003).
21. Wiznerowicz, M. & Trono, D. Conditional suppression of cellular genes: lentivirus vector-mediated drug-inducible RNA interference. *J. Virol.* **77**, 8957–8961 (2003).
22. Matsukura, S., Jones, P.A. & Takai, D. Establishment of conditional vectors for hairpin siRNA knockdowns. *Nucleic Acids Res.* **31**, e77 (2003).
23. Hosono, T. et al. Adenovirus vector-mediated doxycycline-inducible RNA interference. *Hum. Gene Ther.* **15**, 813–819 (2004).
24. Gupta, S., Schoer, R.A., Egan, J.E., Hannon, G.J. & Mittal, V. Inducible, reversible, and stable RNA interference in mammalian cells. *Proc. Natl. Acad. Sci. USA* **101**, 1927–1932 (2004).
25. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl. Acad. Sci. USA* **89**, 5547–5551 (1992).
26. Gossen, M. et al. Transcriptional activation by tetracyclines in mammalian cells. Science 268, 1766–1769 (1995).
27. Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D. & Lowe, S.W. Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16INK4a. *Cell* **88**, 593–602 (1997).
28. Dirac, A.M. & Bernards, R. Reversal of senescence in mouse fibroblasts through lentiviral suppression of p53. J. Biol. Chem. 278, 11731–11734 (2003).
29. Felsher, D.W. Reversibility of oncogene-induced cancer. *Curr. Opin. Genet. Dev.* **14**, 37–42 (2004).
30. Harvey, M. et al. In vitro growth characteristics of embryo fibroblasts isolated from p53-deficient mice. *Oncogene* **8**, 2457–2467 (1993).
31. Stegmeier, F., Hu, G., Rickles, R., Hannon, G.J. & Elledge, S.J. A lentiviral microRNAbased system for single copy Pol II regulated RNAi in mammalian cells. *Proc. Natl.Acad. Sci. USA* **102**, 13212–13217 (2005).
32. Paddison, P.J. et al. Cloning of short hairpin RNAs for gene knockdown in mammalian cells. *Nat. Methods* **1**, 163–167 (2004).
33. Schmitt, C.A., Rosenthal, C.T. & Lowe, S.W. Genetic analysis of chemoresistance in primary murine lymphomas. *Nat. Med.* **6**, 1029–1035 (2000).
34. Schmitt, C.A. et al. A senescence program controlled by p53 and p16INK4a contributes to the outcome of cancer therapy. *Cell* **109**, 335–346 (2002).
35. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* **31**, 3406–3415 (2003).

Chapter 6

Conclusions and Perspectives

6.1. Experimental Conclusions and their Context in the Field

Small RNAs are important regulators of numerous cellular and developmental processes in mammals. The profiling of small RNAs from mammary progenitor cells has allowed us to assess their miRNA composition, providing insights into the potential cellular mechanisms and regulatory networks required for normal mammary gland development. The aforementioned analysis has also allowed us to co-opt these miRNAs as developmental sensors, thus enabling *in vivo* tracking of a particular cellular compartment. The generation of the miR-sensor mouse model may potentially provide a means of marking stem and/or progenitor populations in the mammary gland, and perhaps be an invaluable tool allowing the assessment of stem/progenitor cell contribution in cancer development.

Stem cells are critical for tissue homeostasis in a microenvironment that is constantly challenged by developmental progression, genetic predispositions and macro-environmental effects (such as diet and lifestyle). The existence of stem/progenitor cells within various tissue compartments such as intestine, skin, brain, lung, pancreas and bone marrow have been identified. From these studies, it is becoming clear that mechanisms governing self-renewal and differentiation programs are shared across these multipotent tissue-specific stem/progenitor cells. In fact, gene expression analyses of ES cells, HSC, and neuronal precursors have demonstrated that they utilize and regulate similar signaling pathways (Phillips et al. 2000), (Terskikh et al. 2001) (Ivanova et al. 2002). One example is the highly conserved Wnt and Notch signaling pathways, which have been shown to be critical in the maintenance of many stem cell types. Extending this philosophy to cancer,

it is proposed that malignant stem cells will likely share similar molecular mechanisms in maintaining self-renewal, cell-fate determination, and apoptosis.

In light of the cancer stem cell hypothesis the study of stem/progenitor cells within the normal mammary epithelium has become an intense area of research. In early 2004 mammary stem/progenitor cell markers were limited to Sca-1 and SP phenotype. By 2006, CD-49f, CD-29 and CD-24, were identified and characterized in primary murine mammary gland. In addition, the immortalized cell line Comma-Dβ (Deugnier et al. 2006)) was characterized providing accesses to cells amendable to 2D and 3D cultures with reconstitution potential. Using this system we have added ALDH to the murine MESC marker toolkit. As we approach 2010, cell surface and functional markers that clearly identify the MESC compartment during mammary development remain to be discovered.

From studies using Comma-Dβ cells, we learned that the basic rules that govern the maintenance and differentiation process of mammary epithelial cells are preserved even in an artificial 2-D culture condition with defined media. The progenitor populations within this cell line are receptive to Wnt-1 signaling, and require Notch-4 for their survival, similar to what others have observed *in vivo* (Shackelton et al. 2006). We can manipulate the expansion or depletion of these progenitor cells through the enforced expression of conserved developmental pathways or miRs, providing us a system where we can alter the cellular compartments and observe their affects on the tissue. Since relatively little is known about the molecular mechanisms that govern self-renewal programs in stem/progenitor cells or 'tumor stem cells' (Reya et al. 2001), Comma-Dβ cells, which are amenable to 2-D and 3-D cultures, constitutes a beautiful system to

identify novel self-renewal regulators. Furthermore, transplantation experiments with manipulated cells allow us to validate our candidate self-renewal regulator genes *in vivo*. Using Comma-Dβ cells, we were able to correlate elevated levels of ALDH in murine mammary cells with cellular resistance to mafosfamide, similar to what is seen in HSC (see chapter 2). From small RNA expression profiling, we were able to identify miRNAs that can be used as sensors for marking cellular compartments in mixed cultures. Using miR-sensors, we were able to enrich for progenitors within Comma-Dβ cells, which allowed us to assess their proliferative and differentiation potential *in vitro*.

At the conclusion of my first paper, I had preliminary data with primary tumor samples provided by Max Wicha's group showing that let-7 family members were down-regulated in ALDH+ cells in 2 out of 3 primary tumor samples. Thus, we profiled human breast cancer cell lines and their tumor-initiating populations (ALDH +cells) to assess the (1) universality of let-7 down-regulation in tumor-initiating cells (2) the potential use of other miR-sensors to mark breast cancer lines that contained a tumorigenic population but lacked ALDH (3) the biological outcome of enforced expression of down-regulated miRs such as let-7 and miR-93, the results of which are described in chapter 3. These miRNA profiles should aid in the identification of genes that play an important role in metastasis and resistance to chemotherapy.

From the let-7 sensor mice described in chapter 4, we were able to show that (1) let-7 expression patterns can be assessed *in vitro* and *in vivo* using miR-sensors, (2) let-7 is differentially expressed in various tissues and that its down-regulation, correlating with higher GFP expression, can be used to enrich for HSC/progenitor cells, and (3) let-7 expression is dynamically regulated during mammary development. These results

provided us with a better understanding of the mammary stem/progenitor cell populations, and enabled us to follow their developmental fates at various stages of mammary gland maturation.

Though the expression patterns of miRNAs have been assessed in many tissues using northern blot analysis, Q-PCR and miRNA- array platforms, their functions in mammalian systems remains largely unknown. Dynamic, cell-specific imaging of miRNA expression and/or activity patterns, along with their target mRNAs and downstream effector genes, is crucial to place small RNAs into their spatial and temporal contexts during development. The sensor mice described in this thesis demonstrated that this is feasible, and could be easily adapted for use in other organ or tissue systems in the mouse.

The expression of let-7 in differentiated cell-types may provide a way to isolate less committed cell-types. Using the let-7 sensor system, cells could be sorted on the basis of GFP expression, and using various *in vitro* assays, it will be possible to assess their differentiation potentials from various tissues or organs. When coupled with other cellular markers associated with multipotent adult stem cells, it might enable a better characterization of the stem cell population in tissues that have thus far been resistant to these analyses.

The relationship between let-7 expression and estrogen receptor status within the mammary epithelium remains unresolved. It has been shown that the body and luminal epithelial cells express estrogen receptors. Interestingly, these cells have a low level of GFP reporter expression, indicating the presence and activity of let-7. In contrast, cap cells, which lack estrogen receptors, down-regulate let-7 expression. Whether estrogen

response elements exist within the primary transcript or promoter of let-7 remains to be seen. Moreover, the observation that E-cadherin and let-7 expression are down-regulated during involution is interesting, and merits further investigation. This correlation might place let-7 in a regulatory network that maintains cellular architectural integrity.

6.2 Future Directions:

The stem cell/progenitor niche of the mammary gland remains undefined, unlike the *Drosophila* ovaries. In this system, the niche imposes a spatiotemporal regulation of gene expression; with the cells in the hub expressing the Notch receptors while the Delta ligands are expressed in the stromal compartment. Interactions between the two cell types dictate the modes of division in the stem cells and their differentiated progeny.

The terminal end bud of mammary glands appears to fit the criteria of a niche. Knockout mouse models of EGFR (epidermal growth factor receptor) and its ligand amphiregulin revealed that stromal-epithelial interactions are essential for mammary development (Sternlicht et al. 2005). EGFR was shown to be required only in the stroma, while amphiregulin is exclusively expressed in cap cells and required in the epithelium. Amphiregulin must be proteolytically shed from the membrane epithelial cap cells surface in order to activate EGFR on nearby stromal cells. Subsequently, downstream targets of EGFR signal back to epithelial compartment in a feedback loop. In addition, the terminal end bud is viewed as a dynamic structure (Zhang and Macara 2008), (Parsa et al. 2008) in which, proliferative cells are situated adjacent to cells receiving apoptotic cues. The differential and spatially distinct expression patterns of let-7 in cap cells and body cells might shed light on the role of stromal-epithelial interactions, as well as differential interpretation of these signals to induce proliferation versus apoptosis.

In addition, the role of let-7 during pregnancy and involution is still unclear. It will be critical to correlate key regulators such as estrogen receptor, leukemic inhibitory factor, prolactin receptor and progesterone with let-7 levels at different mammary developmental stages. We have yet to uncover the auxiliary factors that regulate miRNA maturation, and it will be interesting to understand whether miRNA biogenesis is under the control of these factors in response to hormonal signals.

It is suggestive, though inconclusive, that let-7 is downregulated in other stem/progenitor compartments. High GFP expression at the crypt base of the intestine leads us to speculate let-7 may be involved in the proliferative capacity of intestinal epithelium. The let-7 sensor mice, in combination with other cellular markers, might further refine the stem cell population in various tissues during development.

The role of miRNAs in tumorigenesis is also intriguing, especially with the correlation of EMT with miRNA expression or repression. By crossing the miRNA sensor mice to misregulating key oncogenic/tumor suppressive pathways, we would be able to assess their contribution to tumor initiation and progression, as well as providing insights into novel oncogenes. The identification of the onco-miRs and p53-regulated miR-34 suggest that this is a reasonable avenue of research, and with a better understanding of miRNAs in potential tumorigenic cells, we will be able to shed light on the controversial cancer stem cell theory.

Small RNAs are integral components of numerous immensely complicated gene networks. By harnessing their ability to suppress multiple genes, these may prove to be a viable therapeutic option in the near future. With major important questions still remaining, it is truly an exciting time for this growing field, and I believe that the

findings in the thesis research are merely the precursors to other seminal discoveries to be made in the fields of small RNA and stem cell biology.

Literature Cited

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**(7): 3983-3988.

Alvi, A.J., Clayton, H., Joshi, C., Enver, T., Ashworth, A., Vivanco, M.M., Dale, T.C., and Smalley, M.J. 2003. Functional and molecular characterisation of mammary side population cells. *Breast Cancer Res* **5**(1): R1-8.

Asselin-Labat, M.L., Vaillant, F., Shackleton, M., Bouras, T., Lindeman, G.J., and Visvader, J.E. 2008. Delineating the epithelial hierarchy in the mouse mammary gland. *Cold Spring Harb Symp Quant Biol* **73**: 469-478.

Bach, S.P., Renehan, A.G., and Potten, C.S. 2000. Stem cells: the intestinal stem cell as a paradigm. *Carcinogenesis* **21**(3): 469-476.

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**(7229): 608-611.

Bartel, D.P. 2004. MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* **116**(2): 281-297.

Bashirullah, A., Pasquinelli, A.E., Kiger, A.A., Perrimon, N., Ruvkun, G., and Thummel, C.S. 2003. Coordinate regulation of small temporal RNAs at the onset of Drosophila metamorphosis. *Dev Biol* **259**(1): 1-8.

Beard, C., Hochedlinger, K., Plath, K., Wutz, A., and Jaenisch, R. 2006. Efficient method to generate single-copy transgenic mice by site-specific integration in embryonic stem cells. *Genesis* **44**(1): 23-28.

Bernstein, E., Caudy, A.A., Hammond, S.M., and Hannon, G.J. 2001. Role for a bidentate ribonuclease in the initiation step of RNA interference. *Nature* **409**(6818): 363-366.

Bernstein, E., Kim, S.Y., Carmell, M.A., Murchison, E.P., Alcorn, H., Li, M.Z., Mills, A.A., Elledge, S.J., Anderson, K.V., and Hannon, G.J. 2003. Dicer is essential for mouse development. *Nat Genet* **35**(3): 215-217.

Bethke, A., Fielenbach, N., Wang, Z., Mangelsdorf, D.J., and Antebi, A. 2009. Nuclear hormone receptor regulation of microRNAs controls developmental progression. *Science* **324**(5923): 95-98.

Bhattacharya, S., Jackson, J.D., Das, A.V., Thoreson, W.B., Kuszynski, C., James, J., Joshi, S., and Ahmad, I. 2003. Direct identification and enrichment of retinal stem cells/progenitors by Hoechst dye efflux assay. *Invest Ophthalmol Vis Sci* **44**(6): 2764-2773.

Bohnsack, M.T., Czaplinski, K., and Gorlich, D. 2004. Exportin 5 is a RanGTP-dependent dsRNA-binding protein that mediates nuclear export of pre-miRNAs. *Rna* **10**(2): 185-191.

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**(7): 730-737.

Borchert, G.M., Lanier, W., and Davidson, B.L. 2006. RNA polymerase III transcribes human microRNAs. *Nat Struct Mol Biol* **13**(12): 1097-1101.

Boulanger, C.A., Wagner, K.U., and Smith, G.H. 2005. Parity-induced mouse mammary epithelial cells are pluripotent, self-renewing and sensitive to TGF-beta1 expression. *Oncogene* **24**(4): 552-560.

Bracken, C.P., Gregory, P.A., Kolesnikoff, N., Bert, A.G., Wang, J., Shannon, M.F., and Goodall, G.J. 2008. A double-negative feedback loop between ZEB1-SIP1 and the microRNA-200 family regulates epithelial-mesenchymal transition. *Cancer Res* **68**(19): 7846-7854.

Brennecke, J., Hipfner, D.R., Stark, A., Russell, R.B., and Cohen, S.M. 2003. bantam encodes a developmentally regulated microRNA that controls cell proliferation and regulates the proapoptotic gene hid in Drosophila. *Cell* **113**(1): 25-36.

Cai, X., Hagedorn, C.H., and Cullen, B.R. 2004. Human microRNAs are processed from capped, polyadenylated transcripts that can also function as mRNAs. *Rna* **10**(12): 1957-1966.

Chan, J.A., Krichevsky, A.M., and Kosik, K.S. 2005. MicroRNA-21 is an antiapoptotic factor in human glioblastoma cells. *Cancer Res* **65**(14): 6029-6033.

Charafe-Jauffret, E., Ginestier, C., Iovino, F., Wicinski, J., Cervera, N., Finetti, P., Hur, M.H., Diebel, M.E., Monville, F., Dutcher, J., Brown, M., Viens, P., Xerri, L., Bertucci, F., Stassi, G., Dontu, G., Birnbaum, D., and Wicha, M.S. 2009. Breast cancer cell lines contain functional cancer stem cells with metastatic capacity and a distinct molecular signature. *Cancer Res* **69**(4): 1302-1313.

Charafe-Jauffret, E., Ginestier, C., Monville, F., Wicinski, J., Bertucci, F., Viens, P., Xerri, L., Cabaud, O., Dontu, G., Wicha, M.S., and Birnbaum, D. 2008. [Stem cells and epithelial cancers: the example of breast cancer]. *Ann Pathol* **28 Spec No 1**(1): S30-32.

Chen, C., Ridzon, D.A., Broomer, A.J., Zhou, Z., Lee, D.H., Nguyen, J.T., Barbisin, M., Xu, N.L., Mahuvakar, V.R., Andersen, M.R., Lao, K.Q., Livak, K.J., and Guegler, K.J. 2005. Real-time quantification of microRNAs by stem-loop RT-PCR. *Nucleic Acids Res* **33**(20): e179.

Cheng, A.M., Byrom, M.W., Shelton, J., and Ford, L.P. 2005a. Antisense inhibition of human miRNAs and indications for an involvement of miRNA in cell growth and apoptosis. *Nucleic Acids Res* **33**(4): 1290-1297.

Cheng, N., Bhowmick, N.A., Chytil, A., Gorksa, A.E., Brown, K.A., Muraoka, R., Arteaga, C.L., Neilson, E.G., Hayward, S.W., and Moses, H.L. 2005b. Loss of TGF-beta type II receptor in fibroblasts promotes mammary carcinoma growth and invasion through upregulation of TGF-alpha-, MSP- and HGF-mediated signaling networks. *Oncogene* **24**(32): 5053-5068.

Cheung, A.M., Wan, T.S., Leung, J.C., Chan, L.Y., Huang, H., Kwong, Y.L., Liang, R., and Leung, A.Y. 2007. Aldehyde dehydrogenase activity in leukemic blasts defines a subgroup of acute myeloid leukemia with adverse prognosis and superior NOD/SCID engrafting potential. *Leukemia* **21**(7): 1423-1430.

Corti, S., Locatelli, F., Papadimitriou, D., Donadoni, C., Salani, S., Del Bo, R., Strazzer, S., Bresolin, N., and Comi, G.P. 2006. Identification of a primitive brain-derived neural stem cell population based on aldehyde dehydrogenase activity. *Stem Cells* **24**(4): 975-985.

Cox, C.V., Evely, R.S., Oakhill, A., Pamphilon, D.H., Goulden, N.J., and Blair, A. 2004. Characterization of acute lymphoblastic leukemia progenitor cells. *Blood* **104**(9): 2919-2925.

Daniel, C.W., De Ome, K.B., Young, J.T., Blair, P.B., and Faulkin, L.J., Jr. 1968. The in vivo life span of normal and preneoplastic mouse mammary glands: a serial transplantation study. *Proc Natl Acad Sci U S A* **61**(1): 53-60.

Denli, A.M., Tops, B.B., Plasterk, R.H., Ketting, R.F., and Hannon, G.J. 2004. Processing of primary microRNAs by the Microprocessor complex. *Nature* **432**(7014): 231-235.

Deugnier, M.A., Faraldo, M.M., Teuliere, J., Thiery, J.P., Medina, D., and Glukhova, M.A. 2006. Isolation of mouse mammary epithelial progenitor cells with basal characteristics from the Comma-Dbeta cell line. *Dev Biol* **293**(2): 414-425.

Doench, J.G., Petersen, C.P., and Sharp, P.A. 2003. siRNAs can function as miRNAs. *Genes Dev* **17**(4): 438-442.

du Toit, R.S., Locker, A.P., Ellis, I.O., Elston, C.W., Nicholson, R.I., and Blamey, R.W. 1989. Invasive lobular carcinomas of the breast--the prognosis of histopathological subtypes. *Br J Cancer* **60**(4): 605-609.

Eggan, K., Akutsu, H., Loring, J., Jackson-Grusby, L., Klemm, M., Rideout, W.M., 3rd, Yanagimachi, R., and Jaenisch, R. 2001. Hybrid vigor, fetal overgrowth, and viability of mice derived by nuclear cloning and tetraploid embryo complementation. *Proc Natl Acad Sci U S A* **98**(11): 6209-6214.

Eggan, K., Rode, A., Jentsch, I., Samuel, C., Hennek, T., Tintrup, H., Zevnik, B., Erwin, J., Loring, J., Jackson-Grusby, L., Speicher, M.R., Kuehn, R., and Jaenisch, R. 2002. Male and female mice derived from the same embryonic stem cell clone by tetraploid embryo complementation. *Nat Biotechnol* **20**(5): 455-459.

Elston, C.W., Gresham, G.A., Rao, G.S., Zebro, T., Haybittle, J.L., Houghton, J., and Kearney, G. 1982. The cancer research campaign (King's/Cambridge trial for early breast cancer: clinico-pathological aspects. *Br J Cancer* **45**(5): 655-669.

Esau, C., Davis, S., Murray, S.F., Yu, X.X., Pandey, S.K., Pear, M., Watts, L., Booten, S.L., Graham, M., McKay, R., Subramaniam, A., Propp, S., Lollo, B.A., Freier, S., Bennett, C.F., Bhanot, S., and Monia, B.P. 2006. miR-122 regulation of lipid metabolism revealed by in vivo antisense targeting. *Cell Metab* **3**(2): 87-98.

Esquela-Kerscher, A., Trang, P., Wiggins, J.F., Patrawala, L., Cheng, A., Ford, L., Weidhaas, J.B., Brown, D., Bader, A.G., and Slack, F.J. 2008. The let-7 microRNA reduces tumor growth in mouse models of lung cancer. *Cell Cycle* **7**(6): 759-764.

Evans, M.J. and Kaufman, M.H. 1981. Establishment in culture of pluripotential cells from mouse embryos. *Nature* **292**(5819): 154-156.

Fialkow, P.J., Gartler, S.M., and Yoshida, A. 1967. Clonal origin of chronic myelocytic leukemia in man. *Proc Natl Acad Sci U S A* **58**(4): 1468-1471.

Fialkow, P.J., Singer, J.W., Adamson, J.W., Berkow, R.L., Friedman, J.M., Jacobson, R.J., and Moohr, J.W. 1979. Acute nonlymphocytic leukemia: expression in cells restricted to granulocytic and monocytic differentiation. *N Engl J Med* **301**(1): 1-5.

Fialkow, P.J., Singer, J.W., Adamson, J.W., Vaidya, K., Dow, L.W., Ochs, J., and Moohr, J.W. 1981. Acute nonlymphocytic leukemia: heterogeneity of stem cell origin. *Blood* **57**(6): 1068-1073.

Flanagan, L., Van Weelden, K., Ammerman, C., Ethier, S.P., and Welsh, J. 1999. SUM-159PT cells: a novel estrogen independent human breast cancer model system. *Breast Cancer Res Treat* **58**(3): 193-204.

Forster, C., Makela, S., Warri, A., Kietz, S., Becker, D., Hultenby, K., Warner, M., and Gustafsson, J.A. 2002. Involvement of estrogen receptor beta in terminal differentiation of mammary gland epithelium. *Proc Natl Acad Sci U S A* **99**(24): 15578-15583.

Foudi, A., Hochedlinger, K., Van Buren, D., Schindler, J.W., Jaenisch, R., Carey, V., and Hock, H. 2009. Analysis of histone 2B-GFP retention reveals slowly cycling hematopoietic stem cells. *Nat Biotechnol* **27**(1): 84-90.

Ginestier, C., Hur, M.H., Charafe-Jauffret, E., Monville, F., Dutcher, J., Brown, M., Jacquemier, J., Viens, P., Kleer, C.G., Liu, S., Schott, A., Hayes, D., Birnbaum, D., Wicha, M.S., and Dontu, G. 2007. ALDH1 is a marker of normal and malignant human mammary stem cells and a predictor of poor clinical outcome. *Cell Stem Cell* **1**(5): 555-567.

Goodell, M.A., Brose, K., Paradis, G., Conner, A.S., and Mulligan, R.C. 1996. Isolation and functional properties of murine hematopoietic stem cells that are replicating in vivo. *J Exp Med* **183**(4): 1797-1806.

Gregory, P.A., Bert, A.G., Paterson, E.L., Barry, S.C., Tsykin, A., Farshid, G., Vadas, M.A., Khew-Goodall, Y., and Goodall, G.J. 2008. The miR-200 family and miR-205 regulate epithelial to mesenchymal transition by targeting ZEB1 and SIP1. *Nat Cell Biol* **10**(5): 593-601.

Gregory, R.I., Yan, K.P., Amuthan, G., Chendrimada, T., Doratotaj, B., Cooch, N., and Shiekhattar, R. 2004. The Microprocessor complex mediates the genesis of microRNAs. *Nature* **432**(7014): 235-240.

Grimson, A., Farh, K.K., Johnston, W.K., Garrett-Engele, P., Lim, L.P., and Bartel, D.P. 2007. MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol Cell* **27**(1): 91-105.

Gupta, G.P. and Massague, J. 2006. Cancer metastasis: building a framework. *Cell* **127**(4): 679-695.

Han, J., Lee, Y., Yeom, K.H., Kim, Y.K., Jin, H., and Kim, V.N. 2004. The Drosha-DGCR8 complex in primary microRNA processing. *Genes Dev* **18**(24): 3016-3027.

He, L. and Hannon, G.J. 2004. MicroRNAs: small RNAs with a big role in gene regulation. *Nat Rev Genet* **5**(7): 522-531.

He, L., Thomson, J.M., Hemann, M.T., Hernando-Monge, E., Mu, D., Goodson, S., Powers, S., Cordon-Cardo, C., Lowe, S.W., Hannon, G.J., and Hammond, S.M. 2005. A microRNA polycistron as a potential human oncogene. *Nature* **435**(7043): 828-833.

Heimann, R., Lan, F., McBride, R., and Hellman, S. 2000. Separating favorable from unfavorable prognostic markers in breast cancer: the role of E-cadherin. *Cancer Res* **60**(2): 298-304.

Heo, I., Joo, C., Cho, J., Ha, M., Han, J., and Kim, V.N. 2008. Lin28 mediates the terminal uridylation of let-7 precursor MicroRNA. *Mol Cell* **32**(2): 276-284.

Hess, D.A., Wirthlin, L., Craft, T.P., Herrbrich, P.E., Hohm, S.A., Lahey, R., Eades, W.C., Creer, M.H., and Nolta, J.A. 2006. Selection based on CD133 and high aldehyde dehydrogenase activity isolates long-term reconstituting human hematopoietic stem cells. *Blood* **107**(5): 2162-2169.

Hochedlinger, K., Yamada, Y., Beard, C., and Jaenisch, R. 2005. Ectopic expression of Oct-4 blocks progenitor-cell differentiation and causes dysplasia in epithelial tissues. *Cell* **121**(3): 465-477.

Holmes, C. and Stanford, W.L. 2007. Concise review: stem cell antigen-1: expression, function, and enigma. *Stem Cells* **25**(6): 1339-1347.

Humphreys, R.C., Krajewska, M., Krnacik, S., Jaeger, R., Weiher, H., Krajewski, S., Reed, J.C., and Rosen, J.M. 1996. Apoptosis in the terminal endbud of the murine mammary gland: a mechanism of ductal morphogenesis. *Development* **122**(12): 4013-4022.

Hutvagner, G. and Zamore, P.D. 2002. A microRNA in a multiple-turnover RNAi enzyme complex. *Science* **297**(5589): 2056-2060.

Ibarra, I., Erlich, Y., Muthuswamy, S.K., Sachidanandam, R., and Hannon, G.J. 2007. A role for microRNAs in maintenance of mouse mammary epithelial progenitor cells. *Genes Dev* **21**(24): 3238-3243.

Ivanova, N.B., Dimos, J.T., Schaniel, C., Hackney, J.A., Moore, K.A., and Lemischka, I.R. 2002. A stem cell molecular signature. *Science* **298**(5593): 601-604.

Jackson-Fisher, A.J., Bellinger, G., Breindel, J.L., Tavassoli, F.A., Booth, C.J., Duong, J.K., and Stern, D.F. 2008. ErbB3 is required for ductal morphogenesis in the mouse mammary gland. *Breast Cancer Res* **10**(6): R96.

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., Sawyers, C.L., and Weissman, I.L. 2004. Granulocyte-macrophage progenitors as candidate leukemic stem cells in blast-crisis CML. *N Engl J Med* **351**(7): 657-667.

Jones, C., Mackay, A., Grigoriadis, A., Cossu, A., Reis-Filho, J.S., Fulford, L., Dexter, T., Davies, S., Bulmer, K., Ford, E., Parry, S., Budroni, M., Palmieri, G., Neville, A.M., O'Hare, M.J., and Lakhani, S.R. 2004. Expression profiling of purified normal human luminal and myoepithelial breast cells: identification of novel prognostic markers for breast cancer. *Cancer Res* **64**(9): 3037-3045.

Kawasaki, H., Mizuseki, K., Nishikawa, S., Kaneko, S., Kuwana, Y., Nakanishi, S., Nishikawa, S.I., and Sasai, Y. 2000. Induction of midbrain dopaminergic neurons from ES cells by stromal cell-derived inducing activity. *Neuron* **28**(1): 31-40.

Kenney, N.J., Smith, G.H., Lawrence, E., Barrett, J.C., and Salomon, D.S. 2001. Identification of Stem Cell Units in the Terminal End Bud and Duct of the Mouse Mammary Gland. *J Biomed Biotechnol* **1**(3): 133-143.

Khvorova, A., Reynolds, A., and Jayasena, S.D. 2003. Functional siRNAs and miRNAs exhibit strand bias. *Cell* **115**(2): 209-216.

Kiel, M.J., Yilmaz, O.H., Iwashita, T., Yilmaz, O.H., Terhorst, C., and Morrison, S.J. 2005. SLAM family receptors distinguish hematopoietic stem and progenitor cells and reveal endothelial niches for stem cells. *Cell* **121**(7): 1109-1121.

Kim, V.N. 2005. Small RNAs: classification, biogenesis, and function. *Mol Cells* **19**(1): 1-15.

Kina, T., Ikuta, K., Takayama, E., Wada, K., Majumdar, A.S., Weissman, I.L., and Katsura, Y. 2000. The monoclonal antibody TER-119 recognizes a molecule associated with glycophorin A and specifically marks the late stages of murine erythroid lineage. *Br J Haematol* **109**(2): 280-287.

Kloosterman, W.P. and Plasterk, R.H. 2006. The diverse functions of microRNAs in animal development and disease. *Dev Cell* **11**(4): 441-450.

Kloosterman, W.P., Wienholds, E., Ketting, R.F., and Plasterk, R.H. 2004. Substrate requirements for let-7 function in the developing zebrafish embryo. *Nucleic Acids Res* **32**(21): 6284-6291.

Kordon, E.C. and Smith, G.H. 1998. An entire functional mammary gland may comprise the progeny from a single cell. *Development* **125**(10): 1921-1930.

Korpal, M., Lee, E.S., Hu, G., and Kang, Y. 2008. The miR-200 family inhibits epithelial-mesenchymal transition and cancer cell migration by direct targeting of E-cadherin transcriptional repressors ZEB1 and ZEB2. *J Biol Chem* **283**(22): 14910-14914.

Krutzfeldt, J., Kuwajima, S., Braich, R., Rajeev, K.G., Pena, J., Tuschl, T., Manoharan, M., and Stoffel, M. 2007. Specificity, duplex degradation and subcellular localization of antagomirs. *Nucleic Acids Res* **35**(9): 2885-2892.

Krutzfeldt, J., Rajewsky, N., Braich, R., Rajeev, K.G., Tuschl, T., Manoharan, M., and Stoffel, M. 2005. Silencing of microRNAs in vivo with 'antagomirs'. *Nature* **438**(7068): 685-689.

Lagos-Quintana, M., Rauhut, R., Yalcin, A., Meyer, J., Lendeckel, W., and Tuschl, T. 2002. Identification of tissue-specific microRNAs from mouse. *Curr Biol* **12**(9): 735-739.

Lee, Y., Ahn, C., Han, J., Choi, H., Kim, J., Yim, J., Lee, J., Provost, P., Radmark, O., Kim, S., and Kim, V.N. 2003. The nuclear RNase III Drosha initiates microRNA processing. *Nature* **425**(6956): 415-419.

Lee, Y., Kim, M., Han, J., Yeom, K.H., Lee, S., Baek, S.H., and Kim, V.N. 2004. MicroRNA genes are transcribed by RNA polymerase II. *Embo J* **23**(20): 4051-4060.

Lerner, C. and Harrison, D.E. 1990. 5-Fluorouracil spares hemopoietic stem cells responsible for long-term repopulation. *Exp Hematol* **18**(2): 114-118.

Lindeman, G.J., Visvader, J.E., Smalley, M.J., and Eaves, C.J. 2008. The future of mammary stem cell biology: the power of in vivo transplants. *Breast Cancer Res* **10**(3): 402; author reply 403.

Lund, E., Guttinger, S., Calado, A., Dahlberg, J.E., and Kutay, U. 2004. Nuclear export of microRNA precursors. *Science* **303**(5654): 95-98.

Ma, L., Teruya-Feldstein, J., and Weinberg, R.A. 2007. Tumour invasion and metastasis initiated by microRNA-10b in breast cancer. *Nature* **449**(7163): 682-688.

Manz, M.G., Miyamoto, T., Akashi, K., and Weissman, I.L. 2002. Prospective isolation of human clonogenic common myeloid progenitors. *Proc Natl Acad Sci U S A* **99**(18): 11872-11877.

Martin, G.R. 1981. Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci U S A* **78**(12): 7634-7638.

McCreath, K.J., Howcroft, J., Campbell, K.H., Colman, A., Schnieke, A.E., and Kind, A.J. 2000. Production of gene-targeted sheep by nuclear transfer from cultured somatic cells. *Nature* **405**(6790): 1066-1069.

Morrison, S.J., Uchida, N., and Weissman, I.L. 1995. The biology of hematopoietic stem cells. *Annu Rev Cell Dev Biol* **11**: 35-71.

Newman, M.A., Thomson, J.M., and Hammond, S.M. 2008. Lin-28 interaction with the Let-7 precursor loop mediates regulated microRNA processing. *Rna* **14**(8): 1539-1549.

Nguyen, D.X., Bos, P.D., and Massague, J. 2009. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* **9**(4): 274-284.

Oka, H., Shiozaki, H., Kobayashi, K., Inoue, M., Tahara, H., Kobayashi, T., Takatsuka, Y., Matsuyoshi, N., Hirano, S., Takeichi, M., and et al. 1993. Expression of E-cadherin cell adhesion molecules in human breast cancer tissues and its relationship to metastasis. *Cancer Res* **53**(7): 1696-1701.

Olsen, P.H. and Ambros, V. 1999. The lin-4 regulatory RNA controls developmental timing in Caenorhabditis elegans by blocking LIN-14 protein synthesis after the initiation of translation. *Dev Biol* **216**(2): 671-680.

Orimo, A., Gupta, P.B., Sgroi, D.C., Arenzana-Seisdedos, F., Delaunay, T., Naeem, R., Carey, V.J., Richardson, A.L., and Weinberg, R.A. 2005. Stromal fibroblasts present in invasive human breast carcinomas promote tumor growth and angiogenesis through elevated SDF-1/CXCL12 secretion. *Cell* **121**(3): 335-348.

Orom, U.A., Nielsen, F.C., and Lund, A.H. 2008. MicroRNA-10a binds the 5'UTR of ribosomal protein mRNAs and enhances their translation. *Mol Cell* **30**(4): 460-471.

Parizotto, E.A., Dunoyer, P., Rahm, N., Himber, C., and Voinnet, O. 2004. In vivo investigation of the transcription, processing, endonucleolytic activity, and functional relevance of the spatial distribution of a plant miRNA. *Genes Dev* **18**(18): 2237-2242.

Park, S.M., Gaur, A.B., Lengyel, E., and Peter, M.E. 2008. The miR-200 family determines the epithelial phenotype of cancer cells by targeting the E-cadherin repressors ZEB1 and ZEB2. *Genes Dev* **22**(7): 894-907.

Parsa, S., Ramasamy, S.K., De Langhe, S., Gupte, V.V., Haigh, J.J., Medina, D., and Bellusci, S. 2008. Terminal end bud maintenance in mammary gland is dependent upon FGFR2b signaling. *Dev Biol* **317**(1): 121-131.

Pasquinelli, A.E., McCoy, A., Jimenez, E., Salo, E., Ruvkun, G., Martindale, M.Q., and Baguna, J. 2003. Expression of the 22 nucleotide let-7 heterochronic RNA throughout the Metazoa: a role in life history evolution? *Evol Dev* **5**(4): 372-378.

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., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O., and Botstein, D. 2000. Molecular portraits of human breast tumours. *Nature* **406**(6797): 747-752.

Phillips, R.L., Ernst, R.E., Brunk, B., Ivanova, N., Mahan, M.A., Deanehan, J.K., Moore, K.A., Overton, G.C., and Lemischka, I.R. 2000. The genetic program of hematopoietic stem cells. *Science* **288**(5471): 1635-1640.

Reinhart, B.J., Slack, F.J., Basson, M., Pasquinelli, A.E., Bettinger, J.C., Rougvie, A.E., Horvitz, H.R., and Ruvkun, G. 2000. The 21-nucleotide let-7 RNA regulates developmental timing in Caenorhabditis elegans. *Nature* **403**(6772): 901-906.

Reya, T., Morrison, S.J., Clarke, M.F., and Weissman, I.L. 2001. Stem cells, cancer, and cancer stem cells. *Nature* **414**(6859): 105-111.

Rossi, D.J., Jamieson, C.H., and Weissman, I.L. 2008. Stems cells and the pathways to aging and cancer. *Cell* **132**(4): 681-696.

Rybak, A., Fuchs, H., Smirnova, L., Brandt, C., Pohl, E.E., Nitsch, R., and Wulczyn, F.G. 2008. A feedback loop comprising lin-28 and let-7 controls pre-let-7 maturation during neural stem-cell commitment. *Nat Cell Biol* **10**(8): 987-993.

Schulman, B.R., Esquela-Kerscher, A., and Slack, F.J. 2005. Reciprocal expression of lin-41 and the microRNAs let-7 and mir-125 during mouse embryogenesis. *Dev Dyn* **234**(4): 1046-1054.

Schwarz, D.S., Hutvagner, G., Du, T., Xu, Z., Aronin, N., and Zamore, P.D. 2003. Asymmetry in the assembly of the RNAi enzyme complex. *Cell* **115**(2): 199-208.

Sempere, L.F., Dubrovsky, E.B., Dubrovskaya, V.A., Berger, E.M., and Ambros, V. 2002. The expression of the let-7 small regulatory RNA is controlled by ecdysone during metamorphosis in Drosophila melanogaster. *Dev Biol* **244**(1): 170-179.

Sempere, L.F., Sokol, N.S., Dubrovsky, E.B., Berger, E.M., and Ambros, V. 2003. Temporal regulation of microRNA expression in Drosophila melanogaster mediated by hormonal signals and broad-Complex gene activity. *Dev Biol* **259**(1): 9-18.

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**(7072): 84-88.

Siitonen, S.M., Kononen, J.T., Helin, H.J., Rantala, I.S., Holli, K.A., and Isola, J.J. 1996. Reduced E-cadherin expression is associated with invasiveness and unfavorable prognosis in breast cancer. *Am J Clin Pathol* **105**(4): 394-402.

Singh, S.K., Clarke, I.D., Terasaki, M., Bonn, V.E., Hawkins, C., Squire, J., and Dirks, P.B. 2003. Identification of a cancer stem cell in human brain tumors. *Cancer Res* **63**(18): 5821-5828.

Smalley, M. and Ashworth, A. 2003. Stem cells and breast cancer: A field in transit. *Nat Rev Cancer* **3**(11): 832-844.

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**(1): 21-31.

Smith, G.H. and Medina, D. 1988. A morphologically distinct candidate for an epithelial stem cell in mouse mammary gland. *J Cell Sci* **90 (Pt 1)**: 173-183.

Sokol, N.S. and Ambros, V. 2005. Mesodermally expressed Drosophila microRNA-1 is regulated by Twist and is required in muscles during larval growth. *Genes Dev* **19**(19): 2343-2354.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown,

P.O., Botstein, D., Eystein Lonning, P., and Borresen-Dale, A.L. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**(19): 10869-10874.

Sternlicht, M.D., Sunnarborg, S.W., Kouros-Mehr, H., Yu, Y., Lee, D.C., and Werb, Z. 2005. Mammary ductal morphogenesis requires paracrine activation of stromal EGFR via ADAM17-dependent shedding of epithelial amphiregulin. *Development* **132**(17): 3923-3933.

Stingl, J., Eaves, C.J., Kuusk, U., and Emerman, J.T. 1998. Phenotypic and functional characterization in vitro of a multipotent epithelial cell present in the normal adult human breast. *Differentiation* **63**(4): 201-213.

Stingl, J., Eaves, C.J., Zandieh, I., and Emerman, J.T. 2001. Characterization of bipotent mammary epithelial progenitor cells in normal adult human breast tissue. *Breast Cancer Res Treat* **67**(2): 93-109.

Stingl, J., Eirew, P., Ricketson, I., Shackleton, M., Vaillant, F., Choi, D., Li, H.I., and Eaves, C.J. 2006. Purification and unique properties of mammary epithelial stem cells. *Nature* **439**(7079): 993-997.

Takamizawa, J., Konishi, H., Yanagisawa, K., Tomida, S., Osada, H., Endoh, H., Harano, T., Yatabe, Y., Nagino, M., Nimura, Y., Mitsudomi, T., and Takahashi, T. 2004. Reduced expression of the let-7 microRNAs in human lung cancers in association with shortened postoperative survival. *Cancer Res* **64**(11): 3753-3756.

Terskikh, A.V., Easterday, M.C., Li, L., Hood, L., Kornblum, H.I., Geschwind, D.H., and Weissman, I.L. 2001. From hematopoiesis to neuropoiesis: evidence of overlapping genetic programs. *Proc Natl Acad Sci U S A* **98**(14): 7934-7939.

Thiery, J.P. 2002. Epithelial-mesenchymal transitions in tumour progression. *Nat Rev Cancer* **2**(6): 442-454.

Tlsty, T.D. 2001. Stromal cells can contribute oncogenic signals. *Semin Cancer Biol* **11**(2): 97-104.

Tsai, Y.C., Lu, Y., Nichols, P.W., Zlotnikov, G., Jones, P.A., and Smith, H.S. 1996. Contiguous patches of normal human mammary epithelium derived from a single stem cell: implications for breast carcinogenesis. *Cancer Res* **56**(2): 402-404.

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**(5656): 359-363.

Uchida, N. and Weissman, I.L. 1992. Searching for hematopoietic stem cells: evidence that Thy-1.1lo Lin- Sca-1+ cells are the only stem cells in C57BL/Ka-Thy-1.1 bone marrow. *J Exp Med* **175**(1): 175-184.

Urlinger, S., Baron, U., Thellmann, M., Hasan, M.T., Bujard, H., and Hillen, W. 2000. Exploring the sequence space for tetracycline-dependent transcriptional activators: novel mutations yield expanded range and sensitivity. *Proc Natl Acad Sci U S A* **97**(14): 7963-7968.

Valastyan, S., Reinhardt, F., Benaich, N., Calogrias, D., Szasz, A.M., Wang, Z.C., Brock, J.E., Richardson, A.L., and Weinberg, R.A. 2009. A pleiotropically acting microRNA, miR-31, inhibits breast cancer metastasis. *Cell* **137**(6): 1032-1046.

Van Zant, G. 1984. Studies of hematopoietic stem cells spared by 5-fluorouracil. *J Exp Med* **159**(3): 679-690.

Viswanathan, S.R., Daley, G.Q., and Gregory, R.I. 2008. Selective blockade of microRNA processing by Lin28. *Science* **320**(5872): 97-100.

Wang, Y., Medvid, R., Melton, C., Jaenisch, R., and Blelloch, R. 2007. DGCR8 is essential for microRNA biogenesis and silencing of embryonic stem cell self-renewal. *Nat Genet* **39**(3): 380-385.

Weissman, I.L., Anderson, D.J., and Gage, F. 2001. Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu Rev Cell Dev Biol* **17**: 387-403.

Welm, B.E., Tepera, S.B., Venezia, T., Graubert, T.A., Rosen, J.M., and Goodell, M.A. 2002. Sca-1(pos) cells in the mouse mammary gland represent an enriched progenitor cell population. *Dev Biol* **245**(1): 42-56.

Williams, J.M. and Daniel, C.W. 1983. Mammary ductal elongation: differentiation of myoepithelium and basal lamina during branching morphogenesis. *Dev Biol* **97**(2): 274-290.

Wood, S.A., Allen, N.D., Rossant, J., Auerbach, A., and Nagy, A. 1993. Non-injection methods for the production of embryonic stem cell-embryo chimaeras. *Nature* **365**(6441): 87-89.

Wulczyn, F.G., Smirnova, L., Rybak, A., Brandt, C., Kwidzinski, E., Ninnemann, O., Strehle, M., Seiler, A., Schumacher, S., and Nitsch, R. 2007. Post-transcriptional regulation of the let-7 microRNA during neural cell specification. *Faseb J* **21**(2): 415-426.

Yang, J., Mani, S.A., Donaher, J.L., Ramaswamy, S., Itzykson, R.A., Come, C., Savagner, P., Gitelman, I., Richardson, A., and Weinberg, R.A. 2004. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**(7): 927-939.

Yi, R., Doehle, B.P., Qin, Y., Macara, I.G., and Cullen, B.R. 2005. Overexpression of exportin 5 enhances RNA interference mediated by short hairpin RNAs and microRNAs. *Rna* **11**(2): 220-226.

Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., and Song, E. 2007a. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* **131**(6): 1109-1123.

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

Zeng, Y., Wagner, E.J., and Cullen, B.R. 2002. Both natural and designed micro RNAs can inhibit the expression of cognate mRNAs when expressed in human cells. *Mol Cell* **9**(6): 1327-1333.

Zhang, H. and Macara, I.G. 2008. The PAR-6 polarity protein regulates dendritic spine morphogenesis through p190 RhoGAP and the Rho GTPase. *Dev Cell* **14**(2): 216-226.

Zhou, S., Schuetz, J.D., Bunting, K.D., Colapietro, A.M., Sampath, J., Morris, J.J., Lagutina, I., Grosveld, G.C., Osawa, M., Nakauchi, H., and Sorrentino, B.P. 2001. The ABC transporter Bcrp1/ABCG2 is expressed in a wide variety of stem cells and is a molecular determinant of the side-population phenotype. *Nat Med* **7**(9): 1028-1034.