

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Insights into breast cancer metastasis: understanding the roles of transcription factors, microRNAs and the tumor microenvironment

Permalink

<https://escholarship.org/uc/item/9wc3t57g>

Author

Chou, Jonathan

Publication Date

2012

Supplemental Material

<https://escholarship.org/uc/item/9wc3t57g#supplemental>

Peer reviewed|Thesis/dissertation

**Insights into breast cancer metastasis: understanding the roles of
transcription factors, microRNAs and the tumor microenvironment**

by

Jonathan Chou

DISSERTATION

Submitted in partial satisfaction of the requirements for the degree of

DOCTOR OF PHILOSOPHY

in

Biomedical Sciences

in the

GRADUATE DIVISION

of the

UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

ACKNOWLEDGEMENTS

I've had a lot of fun during these past four years in graduate school, which have passed by very quickly. There are many people I would like to thank, who have helped me at various points along my journey, given me advice and guidance, listened patiently to my frustrations, and made it possible for me to be doing the work that I enjoy.

First, I'd like to thank Zena Werb, my graduate advisor, who has given me a wonderful perspective on science, and has taught me how to read broadly and to take a step back and fit the research into a larger picture. Zena has given me a degree of scientific freedom that I don't think most graduate students have. I have appreciated her support to allow me to try different things in the lab, solely out of my own interests. She has also taught me the importance of choosing an interesting and significant problem to work on and wisely advised me that if I'm going to spend a lot of time and energy figuring something out, I might as well spend it solving something important. In addition, Zena has allowed me to pursue opportunities outside the lab, whether it is doing a clinical preceptorship, taking a class at Mission Bay, or teaching medical school small groups. Zena has provided an encouraging atmosphere for a developing young scientist, and has been incredibly supportive about helping me finish my PhD in a reasonable time so that I can return back and complete medical school.

There are many people in the Werb Lab that I have helped me tremendously. Euan Storch taught me the ropes when I first joined the lab. During my rotation, I was fortunate to have worked with Euan on an exciting project to characterize a novel metastasis-promoting gene. I learned many techniques from him that I constantly used, including 3D cell culture. Sylvain Provot taught me how to systematically approach a scientific problem and showed me all of his organizational skills. Sylvain was a wonderful collaborator and mentor on the various GATA3 projects, and always provided many helpful suggestions and criticisms about my work. He gave me a lot of day-to-day guidance, and helped me to think more logically and to write better scientific papers. Like Sylvain, Vicki Plaks was a wonderful collaborator and mentor, who taught me everything I know about working with mice. We have had fantastic scientific discussions and a fun time working on the metastatic niche project together. Vicki has been very encouraging, and always ignites the scientific spark.

Audrey Brenot has been a great collaborator and lab manager. We have had a lot of interesting discussions in the animal house procedure room, while injecting lots of mice (in total, I think we've injected around four or five hundred mice together). As the lab manager, Audrey has also kept the lab organized and running smoothly. Without Audrey, my PhD would have taken at least twice as long. She also has outstanding party planning skills, and has planned great lab get-togethers (including the lab happy hour, baby showers, goodbye parties and birthday celebrations), making the lab a fun place to work. Amy-Jo Casbon has

taught me a lot about flow cytometry and Flow-Jo analysis. In addition, I thank Amy-Jo for introducing the lab to a new level of organization: Lablife and the label maker! Marja Lohela has been a wonderful benchmate, and has let me occupy all of the space in our shared fridge and freezer without ever getting mad at me. She has given me excellent advice on the art and black magic of immunostaining. Payam Shahi, the newest member of the lab, has been wonderful to bounce ideas off of. His enthusiasm for science is truly contagious, and he is always been willing to lend a hand. I have really enjoyed our conversations on both science and non-science topics.

I have been lucky to have several outstanding UC Berkeley undergraduates work with me. I recruited Jeffrey Lin two years ago to join my project, when Jeff was just starting his junior year at Cal. He worked with me for two years and has really been instrumental in several of the projects. I have had a fun time mentoring him, and am happy to see that Jeff is applying to MD/PhD programs this year. In addition, I have had two new students join me this summer, Jennifer Tai and Joanne Dai. They have brought great energy into the lab and have helped me enormously to finish up various loose ends this summer.

Elena Atamaniuc and Ying Yu have both been wonderful technicians in the lab who have been lifesavers on more than a handful of occasions. Elena has been tremendously helpful with maintaining my mouse colony, and Ying has helped me perform bioluminescence imaging and genotyping. Both have helped me with my mice whenever I am away from the lab. In addition, Helen Capili has been a true pro at cutting all of my difficult blocks and providing me sections. Linda Prentice has also provided invaluable histology advice and has helped process all of my tissue blocks for the past four years.

The other many awesome people in the Werb Lab, both past and present, that have made it a great place to work: Kai Kessenbrock, Devon Lawson, Laurie Littlepage, Ken Takei, Annie Le, George Lemieux, Caroline Bonnans, Mikala Egeblad, Matilda Chan, Joanna Phillips, Delia Pata, Kelly Kersten and Eline van Kappel. Thank you for the laughs, the help, the suggestions and the support.

The two people that make our lab run: Rolan Gregg and Meredith Donnelly. Both Meredith and Rolan handle the most important parts of the lab—ordering all of the reagents we need and handling our reimbursements. They have been helpful with so many aspects of day-to-day life in the lab.

I would like to thank my qualifying exam and thesis committee members: Jay Debnath (my qualifying exam chair), Gabriele Bergers (my thesis committee chair), Dean Sheppard and Andrei Goga. They have provided me with invaluable feedback and criticisms about my work and research proposal, and gave encouraging advice at each thesis committee meeting. I thank them for their guidance throughout graduate school.

In addition, Dave Steiner and Lionel Lim helped me get into microRNA research. I appreciated all of our discussions and suggestions. It was great bouncing around ideas on miR-29 with Dave, and being able to exchange reagents, and Lionel helped me get jump started by generously providing reagents.

Two of my favorite classes as an undergraduate at Stanford were Cancer Biology taught by Joe Lipsick, and Developmental Biology taught by Sue McConnell and Michael Simon. These classes transformed how to think about and approach biology, and I thank these professors for sparking my interest in development and cancer nearly a decade ago. My undergraduate research advisor, Yanmin Yang, showed me the excitement of doing biomedical research and supported and encouraged me to pursue MD/PhD programs.

Several important physicians and nurses saved my life eight years ago when I was diagnosed with Hodgkin's lymphoma: Dr. Sandra Horning (my primary oncologist, who was an inspiration because of her personal battle with cancer), Dr. Richard Hoppe, Dr. Ranjana Advani, Dr. Ginna Lapport and Dr. Robert Lowsky (my BMT doctors) and Sheila Breslin, the wonderful research nurse coordinator who helped my family and me get through the most difficult period of my life. They gave me another chance at life and inspired me to pursue a career as a cancer researcher and physician.

My friends and family has been a constant source of support and love, reminding me to take care of myself and what's important in life. Chung-Hsiu Yao, my mom, who has been a strong pillar in my life, who has provided me with the opportunities, skills and confidence to pursue a medical and scientific career. She has taught me many important life lessons, and has supported me in everything that I do. Jennifer Chou, my sister, who has supported me with love, games and cupcakes. My extended family and in-laws, Chung-Li Hsu, Anna and King Choi, who flew out from Houston and Boston to attend my dissertation seminar, and my sister-in-law, Priscilla Choi, who has baked me many cinnamon rolls. Finally, my wife, Charina Choi, who is my best friend, supporter, critic, partner, exercise coach and Dominion challenger, has supported me in everything that I do. As the family coach, she makes sure that we have a healthy dinner every night and get our weekly exercise. Charina always has bold and creative ideas to pursue and challenges me to think outside the box. I thank her for taking a leap of faith to stay in the Bay Area for graduate school, allowing us to discover treasures like Berkeley Bowl, Yali's Cafe, Chez Panisse and the Cheeseboard. We have had a lot of fun together, both in and outside of the lab.

As I return back to medical school, I have the incredible privilege of interacting with patients and their families. They give me a strong purpose to continue doing research. As a cancer survivor, I have also experienced the cruelty of modern medicine, of infusing patients with cytotoxic chemotherapy. It is my hope that research into the mechanisms underlying cancer and metastasis will translate into improved survival outcomes. There is still much to learn.

ABSTRACT

Insights into breast cancer metastasis: understanding the roles of transcription factors, microRNAs and the tumor microenvironment

by
Jonathan Chou

Most cancer-related deaths are attributable to metastatic disease, a multistep process that includes invasion, angiogenesis, intravasation, extravasation, colonization and growth. Despite advances in our understanding of cancer, few therapeutics to prevent or treat metastases are currently available. The focus of my dissertation research has been to investigate the molecular and cellular events that lead to cancer metastasis, using breast cancer as a model, and to elucidate factors that may promote or suppress this process.

Here, I investigate the effects of GATA3 on tumor cell differentiation and metastasis. I show that GATA3 inhibits lung metastasis through miR-29b, a microRNA that promotes differentiation and targets pro-metastatic regulators of the tumor environment for degradation. GATA3 also inhibits bone metastasis, likely through suppression of osteopontin. In addition, I present a preliminary GATA3 ChIP-seq dataset performed in breast cancer cells to uncover novel GATA3 targets.

Because microRNAs are generally down-regulated in cancer, I investigate the consequences of globally decreasing microRNAs using a conditional knockout for *Dgcr8*. I show that *Dgcr8* controls cellular differentiation and branching morphogenesis in both normal and tumorigenic mammary epithelial cells. In contrast to other cancer studies on *DICER*, *Dgcr8* deletion in MMTV-PyMT tumors impairs growth and metastasis.

The microenvironment at distant metastatic sites is altered by the primary tumor. I describe preliminary work on the metastatic lung niche in MMTV-PyMT mice that is established early during tumor progression and associated with increases in pro-inflammatory mediators. We find increased number of MMP9-expressing CD11b+Gr1+ myeloid cells that cluster around lung metastases, which may promote colonization and growth.

Finally, I describe two projects investigating the roles of novel metastasis-promoting genes, *Zeppo1* and *Znf217*. Both these genes lie in commonly amplified regions in poor-prognostic breast cancers. *Zeppo1* promotes metastasis by repressing *E-cadherin* expression and promoting a p120-catenin isoform switch while *Znf217* promotes mesenchymal marker expression and chemotherapy resistance. I also describe the development of the tetrapod quantum dot, a novel tool to study cell-generated forces and characterize novel FVB/n cell lines that I've generated from primary mouse breast tumors, which will serve as a resource for our lab and others.

TABLE OF CONTENTS

Chapter 1: <i>Introduction</i>	1
Chapter 2: <i>GATA3 in development and cancer differentiation: cells GATA have it!</i>	12
Chapter 3: <i>GATA3 suppresses metastasis and modulates the tumor microenvironment by regulating miR-29b expression</i>	45
Chapter 4: <i>GATA3 inhibits bone metastasis and expression of Osteopontin and Rank</i>	104
Chapter 5: <i>Genome-wide identification of GATA3 binding sites in breast cancer cells by ChIP-seq reveals candidate genes involved in luminal differentiation and metastasis suppression</i>	129
Chapter 6: <i>DGCR8 controls branching morphogenesis and cell fate in mammary epithelial cells and promotes breast cancer progression</i>	213
Chapter 7: <i>Tumor-derived Systemic Factors and the Early Metastatic Niche</i> ...	262
Chapter 8: <i>Conclusions</i>	294
Appendix A: <i>Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization</i> ...	312
Appendix B: <i>The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression</i>	360
Appendix C: <i>A nanocrystal sensor for luminescence detection of cellular forces</i>	412
Appendix D: <i>Generation of novel FVB cell lines for breast cancer studies</i>	431
References	458
Library Release	489

LIST OF TABLES

Table 2.1: The functions of GATA3 in various tissues	44
Table S3.1: miRNA qPCR screen results.....	100
Table S3.2: Primer sequences for real-time qPCR and cloning.....	102
Table 5.1: GATA3 CHIP-seq in human MDA231-GATA3 breast cancer cells...	150
Table 5.1: GATA3 CHIP-seq in murine 4T1-Gata3 breast cancer cells.....	205

LIST OF FIGURES

Figure 2.1: Functional domains in the mouse GATA-binding protein 3	41
Figure 2.2: GATA3 in normal mammary gland development	42
Figure 2.3: GATA3 induces differentiation in carcinoma cells	43
Figure 3.1: GATA3 suppresses spontaneous and experimental breast cancer metastases to the lungs	86
Figure 3.2: GATA3 induces a more luminal phenotype and decreases cell migration.....	87
Figure 3.3: miR-29b is induced by GATA3, enriched in luminal, good prognostic breast cancers, and associated with reduced metastatic potential	88
Figure 3.4: miR-29b promotes luminal characteristics and loss of miR-29b induces a de-differentiated, mesenchymal phenotype	89
Figure 3.5: miR-29b targets pro-metastatic genes involved in remodeling the tumor microenvironment.....	90
Figure 3.6: miR-29b expression inhibits lung metastases and loss of miR-29b increases lung metastases	91
Figure 3.7: miR-29b knockdown increases target gene expression and miR-29b suppresses metastasis by repressing four microenvironmental targets.....	92
Figure 3.8: miR-29b is an important downstream target of GATA3 that mediates its ability to promote luminal differentiation and suppress metastasis.....	93
Figure S3.1: GATA3 suppresses primary tumor growth and experimental lung metastases in human MDA231 breast cancer cells	94
Figure S3.2: GATA3 affects cell migration in 2D and 3D culture, and promotes a luminal differentiation program	95
Figure S3.3: GATA3 induces miR-29b expression, inhibits TGFb-induced EMT, and transcriptionally dampens TGFb signaling components.....	96
Figure S3.4: miR-29b is enriched in luminal epithelial cells, and its loss results in de-differentiation and an EMT-like phenotype.....	97
Figure S3.5: miR29 suppresses pro-metastatic regulators of the tumor microenvironment, including <i>ANGPTL4</i> , <i>LOX</i> , <i>MMP9</i> , <i>VEGFA</i> , and inhibits lung metastases in human and mouse breast cancer cells	98
Figure S3.6: GATA3 promotes differentiation and suppresses metastasis by regulating miR-29b expression.....	99
Figure 4.1: Gata3 inhibits bone metastasis and decreases <i>Opn</i> expression vivo	125
Figure 4.2: Gata3 inhibits expression of <i>Opn</i> and <i>Rank</i> , two genes that promote bone metastasis, but not through a direct transcriptional mechanism	126

Figure 4.3: <i>OPN</i> , but not <i>RANK</i> , is enriched in basal type cancers and increases during tumor progression	127
Figure 4.4: <i>Opn</i> knockdown decreases experimental lung and bone metastases, without affecting primary tumor growth	128
Figure 5.1: GATA3 regulates components of the TGFb/BMP pathways	146
Figure 5.2: GATA3 regulates genes involved in differentiation, inflammation, proteolysis and cell signaling	147
Figure 5.3: Semaphorins, a family of genes identified by GATA3 ChIP-seq, are enriched in luminal breast cancer cells	148
Figure 5.4: Initial validation of hits in murine 4T1 breast cancer cells expressing <i>Gata3</i>	149
Figure 6.1: DGCR8 is expressed in both luminal and basal cells in the mammary gland	253
Figure 6.2: <i>Dgcr8</i> -null mammary epithelial cells fail to maintain a differentiated state and cannot undergo branching morphogenesis	254
Figure 6.3: <i>Dgcr8</i> -null mammary epithelial cells fail to fully reconstitute the mammary gland	255
Figure 6.4: MMTV-Cre-mediated deletion of <i>Dgcr8</i> results in a subtle delay in ductal outgrowth during puberty	256
Figure 6.5: Knockdown of miR-29b delays branching morphogenesis in vitro and but has only subtle effects on mammary reconstitution in vivo	257
Figure 6.6: Expression of DGCR8 is heterogeneous in various luminal and basal mouse breast cancers and low in poor-prognostic basal tumors	258
Figure 6.7: <i>DGCR8</i> expression in human breast cancer is generally independent of subtype and is not robustly associated with prognosis	259
Figure 6.8: <i>Dgcr8</i> -null MMTV-PyMT cells fail to undergo branching morphogenesis and cannot maintain a differentiated state	260
Figure 6.9: Deletion of <i>Dgcr8</i> in PyMT tumor cells is not tolerated and decreases lung metastasis	261
Figure 7.1: During early stages of tumor development, the lungs of MMTV-PyMT mice are more permissive to circulating tumor cells colonization and growth ...	290
Figure 7.2: Gene expression changes associated with increased inflammation, extracellular matrix (ECM) adhesion and remodeling, matrix metalloproteases, and angiogenesis are manifest early in primary adenomas and sustained during tumor progression	291
Figure 7.3: A pro-inflammatory state in the lungs of MMTV-PyMT mice is characterized by systemic changes in the serum and increased CD11b+Gr1+ cells	292

Figure 7.4: Gr1+ myeloid cells accumulate around metastases only in the lungs of MMTV-PyMT mice, and express MMP9.....	293
Figure A.1: <i>Zpo1</i> is a transcriptional repressor expressed in mammary epithelium	352
Figure A.2: <i>Zpo1</i> overexpression promotes epithelial migration and invasion ..	353
Figure A.3: <i>Zpo1</i> promotes increased proliferation and represses <i>E-cadherin</i> expression	354
Figure A.4: <i>Zpo1</i> overexpression promotes EMT and p120-catenin isoform expression and localization	355
Figure A.5: <i>Zpo1</i> -mediated changes in cell adhesion and migration are regulated by the ECM and require Rho GTPase and Src family kinase activity	356
Figure A.6: <i>Zpo1</i> overexpression increases breast cancer metastasis.....	357
Figure A.7: Model of <i>Zpo1</i> overexpression and metastasis promotion	358
Figure A.S1: Overexpression of <i>Zpo1</i> -V5 in EpH4.9 cells	359
Figure A.S2: Immunofluorescent localization of adherens junction proteins in 2D cultured EpH4.9- <i>Zpo1</i> cells	359
Figure B.1: ZNF217 overexpression is a prognostic indicator in breast cancer patients.....	401
Figure B.2: <i>Znf217</i> overexpression promotes increased cell motility and aberrant epithelial marker expression.....	402
Figure B.3: <i>Znf217</i> overexpression causes an increase in soft agar colonies and in mammosphere formation.....	403
Figure B.4: <i>Znf217</i> overexpression in vivo increases rate of tumor progression, tumor heterogeneity and differentiation state	404
Figure B.5: <i>Znf217</i> overexpression in vivo increases lung metastasis.....	405
Figure B.6: Identification of triciribine as a candidate inhibitor of ZNF217-induced growth.....	406
Figure B.7: Triciribine inhibits <i>Znf217</i> in vivo and in human cells.....	407
Figure B.S1: High ZNF217 expression in breast cancer patients is prognostic of poor survival	408
Figure B.S2: <i>Znf217</i> overexpression in mammary epithelial cells induces a scattered morphology	409
Figure B.S3: <i>Znf217</i> alters epithelial differentiation state and promotes metastasis in vivo	410
Figure B.S4: ZNF217 sensitizes MCF7 and ZR-75-1 mammary epithelial cells to heregulin.....	411
Figure C.1: Fabrication of luminescent tetrapod substrates	424

Figure C.2: Diagram of the AOTF microscope used to collect photoluminescence spectra of a 2D tetrapod array with spatial and temporal resolution	425
Figure C.3: Beating heart cells induce shifts in the luminescence color emitted by a tetrapod quantum dot array	425
Figure D.1: A cell line generated from an MMTV-PyMT-Luc/mCherry tumor ...	451
Figure D.2: A cell line generated from an MMTV-PyMT-DB tumor	452
Figure D.3: A cell line generated from an MMTV-Neu tumor	453
Figure D.4: A cell line generated from a C3(1)-Tag tumor	454
Figure D.5: Cell lines generated from spontaneous lung and breast tumors ...	455
Figure D.6: An FVB/n liver tumor cell line expressing Myc and Ras	456
Figure D.7: Analysis of epithelial and mesenchymal markers expressed by these new cell lines	457

Chapter 1: Introduction

Source: This chapter contains background to current cancer metastasis research relevant to the following chapters of this dissertation.

Contributions: This is an original review of the literature written by me, with helpful discussions with members of the Werb Laboratory.

INTRODUCTION

The problem of metastasis

According to the National Cancer Institute, it is estimated that over 1.6 million men and women will be diagnosed with some type of cancer, and nearly 600,000 men and women will die of cancer in 2012. Although tremendous progress in cancer research has been made over the past several decades, patients who fail conventional treatments, who experience a relapse or who develop metastatic disease have poor prognoses. Indeed, it is estimated that over 80% of cancer-related mortalities are attributable to metastatic tumor growth, a process by which the primary tumor spreads to distant sites throughout the body, utilizing the body's blood and lymphatic vessels as travel conduits.

Metastases represent the culmination of a multistep cell biological process, which includes recruiting blood vessels, intravasation into the circulation, scattering to distant tissues, extravasation into the parenchyma of a new organ, and subsequent colonization and growth (Nguyen et al. 2009). This ability to invade and metastasize is one of the classical "hallmarks of cancer" (Hanahan and Weinberg 2000). In breast cancer, metastases often appear in the bone, causing localized erosion of bone tissue and leading to extreme pain. Metastases also arise in the brain, compromising cognitive function, as well as in the liver and lung, leading to organ failure.

Despite advances in our understanding of cancer, few therapeutics to prevent metastases or treat existing metastases are currently clinically available. The focus of my graduate research has been to investigate the molecular and

cellular events that lead to cancer metastasis, using breast cancer as a model, and to elucidate factors that may promote or suppress this deadly process. Ultimately, the overall goal of my research is to identify new targets and pathways to inhibit or activate that will improve cancer patient survival outcomes.

Dissemination, the tumor microenvironment and the metastatic niche

Although metastasis has generally been thought of a late event during tumor progression, there is increasing evidence suggesting that dissemination has already occurred by the time a small primary tumor is first detected. Using two mouse models of breast cancer, Husemann and colleagues found that disseminated cells in the bone marrow could be detected in the premalignant phase of tumorigenesis (Husemann et al. 2008). They found this to be true in human patients with ductal carcinoma in situ (DCIS) as well. These tumor cells must accomplish a challenging task of breaking past a basement membrane barrier and invading into the adjacent normal tissue parenchyma, a process that involves remodeling the extracellular matrix, recruiting blood vessels, and acquiring motility to access the circulation. These scattered cells are essentially undetectable by current imaging modalities, forming only small tumor colonies called micrometastases that will remain dormant for many years. Unfortunately for some patients, these micrometastases will eventually start growing and become macroscopic metastases that are difficult to treat. In these new sites of metastatic growth, immune cells and other components of the tumor microenvironment play an important and contributory role, forming a niche for these tumor cells to grow and prosper (Hanahan and Coussens 2012). The

identification of the signals and cell types that initiate and maintain the metastatic niche and how to perturb it to prevent metastasis are active areas of research.

Tumor differentiation, the epithelial-to-mesenchymal transition (EMT) and cancer stem cells

Before molecular signatures for various cancers existed, pathologists had long appreciated an inverse correlation between tumor differentiation and aggressiveness. Patients with more differentiated tumors generally have better prognoses and lower risk for metastasis than those with poorly-differentiated tumors. Indeed, when a biopsy is performed for a suspected case of breast cancer, the specimen is stained with estrogen receptor (ER), progesterone receptor (PR) and Her2/Neu, a member of the EGFR superfamily. These markers of differentiation determine not only the type of therapy that a patient will be on, but also stratifies the patient into risk groups. Indeed, those patients with triple-negative breast cancer (TNBC), in which the tumor lacks all three molecular markers of differentiated mammary epithelium, respond poorly to conventional treatments and have the lowest survival rate. These observations have led to a more modern hypothesis that there exists less-differentiated stem-like cells within solid tumors, which have been referred to as cancer stem cells (CSCs) or tumor-initiating cells (TICs), capable of self-renewal and giving rise to the primary tumor and/or its metastases (Visvader and Lindeman 2008). Indeed, aggressive, poorly-differentiated basal and triple-negative breast cancers are enriched for genes normally expressed in stem cells (Oct4, c-Myc, Sox2, Nanog), suggesting

that these genes support the stem-like phenotype in tumors (Ben-Porath et al. 2008).

During the metastatic process, cancer cells can selectively turn on the expression of embryonic morphogenesis regulators to undergo an epithelial-to-mesenchymal transition (EMT) and concomitantly turn off programs that maintain their differentiated state (Yang and Weinberg 2008). This program is transcriptionally controlled by a number of transcription factors, including Twist, Snail (Snai1), Slug (Snai2), Zeb1 and Zeb2 (SIP1), many of which directly repress *E-cadherin* (*Cdh1*) expression, as well as polarity genes (Aigner et al. 2007). This allows polarized, immotile epithelial cells to convert into motile mesenchymal cells, a process important physiologically during several stages of embryonic development, and is associated with changes in differentiation markers from cell-cell junction proteins and cytokeratin intermediate filaments to vimentin filaments and fibronectin. Tumor cells that hijack this pathway can gain motility and modify cell adhesion properties, which increases metastasis (Yang et al. 2004a). In addition, the acquisition of an EMT-like state generates cells with properties of stem cells (Mani et al. 2008), thus providing a link between tumor differentiation, EMT, and cancer stem cells.

Although the idea of EMT and cancer stem cells is attractive, there is considerable controversy over whether EMT occurs in vivo. One issue is the difficulty of identifying carcinoma cells from primary human tumor samples that have passed through an EMT. The majority of human breast carcinomas, for example, express E-cadherin and maintain their epithelial morphology, yet still

metastasize, suggesting that they have disseminated without switching to a mesenchymal phenotype (Tarin et al. 2005). Proponents of the EMT hypothesis argue that this is likely due to the transient nature of the EMT program when carcinoma cells metastasize, and that this process is reversible, i.e., cells can undergo a mesenchymal-to-epithelial transition (MET). This makes it unlikely to capture cells that are in an EMT using current detection methods. Moreover, the issue is further complicated by the fact that the EMT is not a binary state. Cancer cells most likely pass through a partial EMT program rather than a complete one, taking on some characteristics of epithelial cells, and other characteristics of mesenchymal cells. This suggests that a spectrum of epithelial-mesenchymal properties exists, and that cells are plastic enough to rapidly shift between multiple states.

As a corollary, the EMT-CSC hypothesis implies that identifying factors and signals that promote differentiation is an attractive strategy for treating poorly differentiated cancers. This idea of promoting differentiation in cancer has been successfully utilized to treat patients with acute promyelocytic leukemia (APML). Indeed, differentiation therapy (either alone or in combination with cytotoxic chemotherapy) is an important alternative strategy. Recently, this approach was demonstrated for treating poorly-differentiated rhabdomyosarcoma (Mishra and Merlino 2009; Taulli et al. 2009). In addition, groups working to find agents that selectively kill CSCs have identified molecules that inhibit tumor progression by promoting differentiation (Gupta et al. 2009). Therefore, understanding the mechanisms controlling normal development and the transcriptional pathways

involved may reveal novel and more sustainable strategies to limit tumor progression and metastasis.

Molecular signatures of metastasis and organ-specific metastasis

Recent work has demonstrated that organ-specific metastasis signatures can be identified. Using the mouse as a “sorter” for organ-specific metastases, Massague and colleagues inject cancer cells into mice to identify the ones that most readily colonize specific organs such as the bones, lungs, or brain (Kang et al. 2003; Minn et al. 2005; Bos et al. 2009). They enrich for these organ-specific metastatic cells by isolating the cells from the metastatic site, and re-injecting them into mice, thereby isolating a pool of cells with enhanced metastatic properties. To identify the genes involved, they perform microarrays to find differentially expressed genes between the parental and enhanced lines, and compare this to microarray data from large cohorts of primary breast cancers. This strategy reveals a subset of site-specific metastasis genes, some of which enhance primary tumorigenicity and others that augment organ-specific metastasis. Other metastasis genes fulfill functions specialized for the microenvironment of the metastatic site and are consequently not selected for in primary tumors. This approach of utilizing a xenograft model to discover genes that mediate organ-specific patterns of metastatic colonization has yielded insight into gene expression programs that govern metastasis, even though most of the lines used are late-stage, already-metastatic cells. The intrinsic genomic instability of cancer cells increases the frequency of alterations necessary to acquire metastatic traits, such as chromosomal gains, losses, and

rearrangements, a mechanism by which tumor cells gain metastasis-promoting genes or lose metastasis-inhibiting genes.

Overview of thesis

In this thesis, I review in **Chapter 2** the existing the literature on a pro-differentiation transcription factor, GATA3, and its role in mammary development, breast cancer and tumor dissemination. I then demonstrate in **Chapter 3** that GATA3 inhibits lung metastasis in mouse models of breast cancer, in part through regulation of a downstream microRNA, miR-29b. I show that miR-29b promotes luminal characteristics, is associated with good prognostic luminal breast cancers, and inhibits metastasis by targeting pro-metastatic regulators of the tumor environment for degradation. Interestingly, loss of miR-29b promotes metastasis and results in the acquisition of an EMT/progenitor-like state, suggesting that differentiation and the microenvironment are coordinately controlled by a GATA3-miR-29b axis to suppress metastasis.

In **Chapter 4**, I describe preliminary evidence showing that GATA3 also inhibits bone metastasis and decreases *Opn/Spp1* and *Rank* expression, two genes previously shown to be important in bone metastasis. Preliminary results show that *Opn* knockdown decreases bone and lung metastasis in vivo. In **Chapter 5**, I describe unpublished data from a collaborative project utilizing chromatin immunoprecipitation followed by massive parallel sequencing (ChIP-Seq) to uncover novel GATA3 targets that may play roles in tumor differentiation and metastasis suppression. This approach reveals that the TGF β pathway is

controlled by GATA3, and implicates other molecules involved in stemness, growth signaling and inflammatory pathways.

Because microRNAs (miRNAs) have been found to be generally down-regulated in cancer, I investigate the consequences of globally decreasing miRNA expression in the mammary gland in **Chapter 6**. Utilizing a conditional knockout model for *Dgcr8*, I show that *Dgcr8* controls cellular differentiation and branching morphogenesis in both normal and tumor mammary epithelial cells. I find that DGCR8 protein is heterogeneously expressed in several mouse models of breast cancer, and significantly down-regulated in the C3(1)-Tag mouse model of basal breast cancer. I also examine the effects on tumor growth and differentiation after *Dgcr8* deletion in the MMTV-PyMT mouse model.

In **Chapter 7**, I describe preliminary experiments highlighting the existence of an early metastatic niche in the lungs of the MMTV-PyMT mouse model. This systemic effect is manifested early during tumor progression, and is associated with increased expression of several pro-inflammatory mediators in the primary tumor, including *S100a8*, *S100a9*, *Saa3*, *Il17*, and *Spp1*. In addition, early during tumor development, the MMTV-PyMT lungs contain an increased number of CD11b+Gr1+ myeloid cells, which have been traditionally called myeloid-derived suppressor cell (MDSCs). These CD11b+Gr1+ cells express MMP9 and cluster around lung metastases.

I conclude in **Chapter 8** with my thoughts and perspectives on metastasis research, and where I hope the field will move in the coming years. The answers to the many exciting and important questions in metastasis biology will hopefully

open up new possibilities and strategies for therapeutic intervention and lead to better patient survival outcomes.

In addition, I have included four additional chapters in the Appendix. In **Appendix A**, I describe a project investigating the role of a novel metastasis-promoting gene, *ZPO1*, which lies in an 8p11–12 amplicon commonly found in luminal B type breast cancers. This amplification is associated with increased proliferation, higher tumor grade and reduced metastasis-free survival. We show that *Zpo1* promotes metastasis in part by repressing *E-cadherin* expression and promoting a p120-catenin isoform switch, which promotes cell migration. In **Appendix B**, I describe a project investigating the role of another metastasis-promoting gene, *ZNF217*, which lies in a 20q13 amplicon found in 20–30% of primary human breast cancers, as well as in colon, ovarian, brain, head and neck and pancreatic cancers. *ZNF217* overexpression promotes metastasis and mesenchymal marker expression, and is associated with chemotherapy resistance and poor prognosis. Interestingly, we identified a nucleoside analog, triciribine, which selectively inhibits *ZNF217*-induced tumor growth, providing an opportunity for tumor-specific therapy.

Although cellular forces play an important role in biology and disease pathogenesis (including cell migration, a prerequisite to metastasis), there are few ways to measure these forces, especially in more complex geometries. In **Appendix C**, I describe the development of a novel nanocrystal tool, the tetrapod quantum dot, which is a collaborative project with my wife, Charina Choi. We describe a proof-of-principle experiment showing that contracting cardiomyocytes

can change the emission wavelength of tetrapod quantum dots, and calculate the force of contraction based on this fluorescence. Our calculated value agrees well with previous values predicted in the literature. We anticipate that this tool can be used to study forces generated in more physiologic environments, such as cell movement in collagen or branching morphogenesis in Matrigel.

Finally, I describe in **Appendix D** the establishment and some initial characterization of novel cell lines that I've generated from primary mouse breast tumors on FVB/n background, which will serve as a resource for our lab and other labs studying breast cancer in syngeneic, immunocompetent mice. These include lines from the MMTV-PyMT, MMTV-PyMT-DB, MMTV-Neu/Her2, and C3(1)-Tag mice. I also describe some initial characterization of metastatic potential of several FVB/n lines established by other investigators, including a Myc/Ras-driven liver tumor (LT2MR), a spontaneous lung adenocarcinoma (LAP0297) and a spontaneous breast adenocarcinoma (MaP0008).

Chapter 2: GATA3 in development and cancer differentiation: cells GATA have it!

Source: The following chapter was published as a mini-review from: Chou, J., Provot, S., and Werb, Z. (2010). GATA3 in development and cancer differentiation: cells GATA have it! *J Cell Physiol* 222, 42-49.

Contributions: This manuscript is an original review of the literature on GATA3 and its role in mammary development and breast cancer. I wrote the manuscript, with contributions from Sylvain Provot and Zena Werb. Zena Werb supervised the project.

GATA3 in development and cancer differentiation: cells GATA have it!

Jonathan Chou^{1,2}, Sylvain Provot¹, and Zena Werb^{1,2,*}

¹Department of Anatomy, and, ²Program in Biomedical Sciences, University of California, San Francisco, CA 94143-0452

*Corresponding Author:

Zena Werb, Ph.D.

Department of Anatomy, Box 0452

University of California, San Francisco

513 Parnassus Avenue

San Francisco, CA 94143-0452, USA

Tel: (415) 476-4622

Fax: (415) 476-4565

email: zena.werb@ucsf.edu

ABSTRACT

There is increasing evidence that the numerous mechanisms that regulate cell differentiation during normal development are also involved in tumorigenesis. In breast cancer, differentiation markers expressed by the primary tumor are routinely profiled to guide clinical decisions. Indeed, numerous studies have shown that the differentiation profile correlates with the metastatic potential of tumors. The transcription factor GATA3 has emerged recently as a strong predictor of clinical outcome in human luminal breast cancer. In the mammary gland, GATA3 is required for luminal epithelial cell differentiation and commitment, and its expression is progressively lost during luminal breast cancer progression as cancer cells acquire a stem cell-like phenotype. Importantly, expression of GATA3 in GATA3-negative, undifferentiated breast carcinoma cells is sufficient to induce tumor differentiation and inhibits tumor dissemination in a mouse model. These findings demonstrate the exquisite ability of a differentiation factor to affect malignant properties, and raise the possibility that GATA3 or its downstream genes could be used in treating luminal breast cancer. This review highlights our recent understanding of GATA3 in both normal mammary development and tumor differentiation.

INTRODUCTION

One of the classical “hallmarks of cancer” is the ability for the tumor to invade and metastasize (Hanahan and Weinberg 2000). This complex process includes several steps, including recruiting blood vessels, intravasation into the circulation, scattering to distant tissues, extravasation into the parenchyma of a new organ, and subsequent colonization and growth (Nguyen et al. 2009). Pathologists have long recognized the intimate connection between tumor progression and its differentiation status. Well-differentiated tumors are generally less advanced and carry a better prognosis whereas poorly differentiated tumors are generally more aggressive and carry a worse prognosis (Kufe and Bast 2003). It has been postulated that cancer cells selectively turn on the expression of embryonic morphogenesis regulators to undergo the epithelial-mesenchymal transition (EMT) and concomitantly turn off programs that maintain their differentiated state (Yang and Weinberg 2008). Interestingly, in addition to gaining motility, the acquisition of an EMT-like state generates cells with properties of stem cells (Mani et al. 2008). In recent years, these embryonic factors have been found to confer malignant traits such as invasiveness and resistance to apoptosis to neoplastic cells (Gupta et al., 2009). This observation has led to the hypothesis that there might exist less-differentiated stem-like cells within solid tumors, which have been referred to as cancer stem cells or tumor-initiating cells capable of self-renewal and giving rise to the entire tumor (Visvader and Lindeman 2008). These considerations illustrate the fact that the key regulatory mechanisms controlling normal embryonic development (EMT, stem cell differentiation and

others) are critical players during tumor progression. They also underscore the importance of identifying the overlapping molecular programs that are shared in these two cellular processes in order to understand how cancers develop and metastasize.

A fundamental aspect of development is the specification and maintenance of differentiated cell types arising from multipotent progenitor cells. The specification of cell fate is mediated in part by hierarchical networks of transcription factors and *cis*-regulatory elements that control their expression. Transcription factors are often organized in multi-gene families and play essential roles in activating target genes of specific cell fates and in repressing target genes of alternate cell fates. The GATA family of transcription factors, of which there are six in mammals, are such master regulators. The GATA factors share common features: they contain two transactivation domains at the amino terminus, two zinc fingers at the carboxyl terminus and a conserved basic region that is located immediately after each zinc finger motif (Fig. 2.1). GATA family members bind to a consensus DNA sequence (A/T)GATA(A/G) in the promoters of genes to directly activate or repress expression of target genes. In addition, GATA factors recruit chromatin remodeling complexes to remodel gene loci (Takemoto et al. 2002; Zhou and Ouyang 2003; Yamashita et al. 2004). At the amino acid level, the family members share varying degrees of homology. For example, while GATA2 and GATA3 are about 55% homologous, GATA3 and GATA4 are only 20% homologous. However, the zinc finger motifs are about 80% homologous among all the six members.

Each specific GATA member is also highly conserved across vertebrate species. GATA3 homologs are found in human, chimpanzee, dog, mouse, rat, chicken, zebrafish, frog, and fruit fly. Between human and mouse, GATA3 shares 97% amino acid identity. In addition, structural analysis of GATA3 suggests that the GATA members can homo- or heterodimerize to bind DNA, or directly bind DNA via its two zinc fingers (Bates et al. 2008).

GATA factors are expressed in a tissue-specific manner. GATA1 and GATA2 are expressed primarily in hematopoietic cells, whereas GATA4, GATA5 and GATA6 are expressed in mesoderm- and endoderm-derived tissues such as the heart, liver and intestines. GATA3 is present in both hematopoietic (e.g., T cells) and non-hematopoietic tissues, including the kidney, central nervous system, skin and mammary gland (Kaufman et al. 2003; Kouros-Mehr et al. 2006; Grote et al. 2008). Although these transcription factors have a restricted expression pattern, their functions are partially interchangeable. For example, the *Gata1* null defects in erythroid cells can be partially rescued by knocking in *Gata3* into the endogenous *Gata1* locus (Tsai et al. 1998; Takahashi et al. 2000). In addition, GATA1, -2, -3 and -4 can all activate expression of interleukin-4 (IL-4) and IL-5, a GATA3 target gene in T cells, as well as repress interferon- γ (IFN γ) (Ranganath and Murphy 2001). This limited capacity for GATA factors to substitute functionally for each other suggests that the cellular context in which each GATA factor is expressed is important. However, since GATA3 is unable to fully rescue the *Gata1*-null phenotype, it suggests that the redundancy is

incomplete, and that despite being highly homologous proteins, each GATA factor retains distinct functions.

Numerous studies using loss- and gain-of-function approaches have shown that GATA family members are crucial transcriptional regulators during the development of a variety of tissues. With the exception of *Gata5*, null mutations for each of the *Gata* genes results in embryonic lethality in mice, underscoring their pivotal roles during development. Furthermore, it was recently shown that GATA4, in combination with another transcription factor Tbx5 and a chromatin-remodeling protein Baf60c, can direct ectopic differentiation of mouse mesoderm into beating cardiomyocytes (Takeuchi and Bruneau 2009). These studies all point to the importance of GATA family members in various aspects of development and cell differentiation.

In this review, we highlight the recent understanding of GATA3 both in mammary gland development as well as in breast cancer differentiation and metastasis. While we will briefly touch on aspects of GATA3 in other systems such as the immune system and skin, we refer the reader to other recently published reviews for a more in depth discussion on those topics (Ho and Pai 2007; Ho et al. 2009)

GATA3 contributes to the normal development of a variety of tissues

GATA3 is a critical regulator in both mouse and human development. *Gata3* null embryos die between E11 and E12 due to internal bleeding, and display growth retardation, deformities in the brain and spinal cord, and gross aberrations in fetal liver hematopoiesis, suggesting that this gene is important in the development of various systems (Pandolfi et al. 1995). Haploinsufficiency of *GATA3* results in Barakat syndrome in humans, characterized by familial hypoparathyroidism, sensorineural deafness and renal dysplasia (also known as HDR syndrome), and can be caused by mutations in *GATA3* that render it physically or functionally inactive. Indeed, several mutations causing Barakat syndrome have been mapped to the critical zinc fingers and adjacent regions that mediate binding to DNA (Van Esch et al. 2000). Interestingly, mutations that abrogate the DNA-binding ability of *GATA3* are also found in human breast cancer specimens (Usary et al. 2004).

GATA3 can undergo several post-translational modifications. The KRRLSA motif found in between the two zinc fingers has sites for acetylation and phosphorylation. The finding that the hypoacetylated KRR mutant of *GATA3* functions as a hypomorph underscores the importance of these residues. Additionally, the Ras-ERK MAPK cascade regulates *GATA3* stability in T cells through inhibition of the Mdm2 E3 ligase and the ubiquitin-proteasome system (Yamashita et al. 2005).

For *GATA3* to regulate gene expression, it must translocate from the cytoplasm into the nucleus to access its target genes. *GATA3* contains a

classical nuclear import signal, and is transported into the nucleus by importin- α (Yang et al. 1994). The affinity of GATA3 to importin- α is regulated by phosphorylation, which is mediated by p38 mitogen-activated protein kinase (MAPK) and serves to enhance nuclear transport (Goldfarb et al. 2004; Maneechotesuwan et al. 2007). Interestingly, corticosteroids, which are commonly used to treat allergic disease, have a potent inhibitory effect on GATA3 in T cells by competing for importin- α , and by inducing the expression of a p38 MAPK inhibitor (Maneechotesuwan et al. 2009).

Given that GATA3 is expressed in such a wide variety of tissues, it is not surprising that the context of GATA3 expression is critical. This is in part mediated by interactions with other protein partners, which may help direct GATA3 to cell-specific targets or modify GATA3 function. These proteins include Smad3, a component of the TGF β signaling pathway, which is important for GATA3 to regulate the expression of T helper cell 2 (T_H2) cytokines (Blokzijl et al. 2002). In addition, in thymocytes, GATA3 binds to the Friend-of-GATA proteins 1 and 2 (FOG1 and FOG2) and PU.1 (Zhou et al. 2001; Nesbit et al. 2004; Chang et al. 2005). These partners may function as co-factors to facilitate GATA3 activity, or function to sequester GATA3 from binding DNA. Because many of these studies have been conducted in T cells, it is unknown whether these regulatory mechanisms and binding partners can be generalized to other cell types, or whether other tissue-specific binding partners exist.

The function of GATA3 has been most extensively studied in T cell development. These studies demonstrate that GATA3 is involved in various

aspects of thymocyte development, and emphasize the notion that GATA3 levels are carefully titrated throughout thymocyte development (Ho et al. 2009). Levels that are too low result in developmental failure, while levels that are too high are cytotoxic. Furthermore, inappropriate expression of GATA3 during thymocyte development can divert the development of thymocyte progenitors into alternative lineages (i.e. mast cells) in a Notch-dependent manner (Taghon et al. 2007).

During their development, thymocytes are specified to be single-positive CD4⁺ (helper T cells, or T_H cells) or CD8⁺ (cytotoxic T cells). The T_H cell population is then further subdivided into T_H1 and T_H2 cells (although other effector cell types such as T_H17 cells have recently been characterized). Importantly, the specification of T_H2 fate depends on GATA3, whereas the T box family transcription factor T-bet specifies the T_H1 lineage (Grogan and Locksley 2002). Recently, *Gata3* was also shown to be a downstream target of Notch; Notch-mediated differentiation of T_H2 cells depends on GATA3. In fact, in the absence of GATA3, Notch turns from a T_H2 inducer into a powerful inducer of T_H1 differentiation (Amsen et al. 2007; Fang et al. 2007). GATA3 also plays a role in chromatin remodeling to allow T_H2 transcriptional factors access to the DNA locus (Lee et al. 2001; Avni et al. 2002). In addition, GATA3 and T-bet cross-antagonize each other to repress the alternate cell fate. For example, GATA3 inhibits the expression of IL-12-induced *Ifn* γ , a classical T_H1 cytokine, and directly transactivates *Il-5* and *Il-13* to reinforce the T_H2 cell choice (Ouyang et al. 1998; Ansel et al. 2006). These studies suggest that GATA3 has a

fundamental role in thymocyte developmental lineage choice, survival and maintenance.

Enforced expression of *Gata3* during T cell development induces CD4⁺CD8⁺ double-positive (DP) T cell lymphoma. The malignant transformation involves cooperation with c-Myc. The lymphoma cells also exhibit activating Notch1 mutations, which result in high expression of Notch targets. Therefore, *Gata3* over-expression converts DP thymocytes into a pre-malignant state, characterized by high c-Myc expression, whereby subsequent induction of Notch1 signaling leads to fully transformed thymocytes (van Hamburg et al. 2008). This example illustrates the dual role of GATA3 in both physiologic development and cancer biology, and underscores the importance of carefully titrating GATA3 levels during development.

GATA3 also plays important roles in skin and hair follicle development. In the skin, epithelial cells undergo an upward differentiation process to give rise to the different hair follicle lineages such as the medulla, cortex and cuticle of the hair shaft and the inner root sheath (IRS) (Fuchs 2007). GATA3 promotes differentiation of the IRS cell lineage, and loss of GATA3 in the skin using a lacZ knock-in that disrupts the *Gata3* allele results in an expansion of IRS precursors and a paucity of differentiated IRS cells (Kaufman et al, 2003). Two other groups have deleted *Gata3* specifically in the epidermis and hair follicles using keratin-14-Cre (K14-Cre). These conditional knockout (CKO) mice display delayed hair growth and maintenance, abnormal hair follicle organization, and defects in skin differentiation (de Guzman Strong et al. 2006; Kurek et al. 2007). The *Gata3*-

CKO mice have desiccated skin that leads to a defective skin barrier and also exhibit an increase in basal epidermal cell proliferation, even though the mice are bald. Transcriptional profiling of *Gata3*-CKO mice shows a defect in lipid biosynthesis due to loss of the lipid acyltransferase gene *Agpat5*, which was identified as a direct GATA3 target.

In addition to its roles in T cell and skin development, GATA3 has different functions in several other tissues. In particular, GATA3 also plays important roles in the development of the nervous system (Lim et al. 2000; Moriguchi et al. 2006; Hong et al. 2008; Zhao et al. 2008; Jones and Warchol 2009), kidney (Grote et al. 2008), lens fiber cells in the eye, and the mammary gland. For example, inactivation of *Gata3* in the nephric duct leads to ectopic ureter budding during development, resulting in a spectrum of urogenital malformations. In addition, embryonic lethality due to loss of *Gata3* has been hypothesized to be caused by noradrenaline deficiency in the sympathetic nervous system (Lim et al. 2000). These diverse roles of GATA3 are summarized in Table 1. Its role in the development of the mammary gland is discussed in detail in the following section.

GATA3 is critical for the normal development of the mammary gland

The mammary gland is composed of mammary epithelial cells, as well as various stromal cells such as adipocytes, fibroblasts, macrophages and mast cells. The epithelium consists of a dual layer of epithelial cells that originate from a common progenitor but are specified by distinct pathways, similarly to the T_H1/T_H2 system.

The luminal epithelial cells line the ductal epithelium, secrete milk proteins and express GATA3. These cells are surrounded by a basal layer of myoepithelial cells, which do not express GATA3 (Fig. 2.2A). These differentiated cell types arise from a multipotent progenitor population that has been recently characterized (Asselin-Labat et al. 2006; Shackleton et al. 2006). Interestingly, the GATA3-negative basal cell population contains the progenitor cell pool, among other cell types (Asselin-Labat et al. 2007). This is consistent with the model that less committed progenitors do not express GATA3, a factor that promotes luminal cell differentiation and maintenance.

In addition to the epithelial cells, GATA3 is expressed in the white adipocyte precursors present in the mammary gland. But in contrast to the positive role of GATA3 in thymocyte and epithelial cell differentiation, down-regulation of GATA3 in these cells leads to adipocyte differentiation, while constitutive GATA3 expression suppresses adipocyte differentiation and prevents cells from developing beyond the preadipocyte stage. This is mediated in part through direct suppression of peroxisome proliferator-activated receptor γ (*Ppar γ*). Thus, GATA3 regulates the preadipocyte-adipocyte developmental transition (Tong et al. 2000).

Prior to puberty, the mammary gland is a rudimentary organ consisting of a primitive network of ductal epithelium. Shortly after the onset of puberty, specialized structures known as terminal end buds (TEBs) develop at the invading epithelial tips of the mammary epithelium (Fig. 2.2A). TEBs contain an outer layer of cap cells, which are believed to be myoepithelial progenitors, and a

multilayered inner core of body cells, which contains the luminal cell progenitors. The TEBs proliferate, bifurcate, and invade into the fatty stroma of the mammary gland in a process known as branching morphogenesis. In the mouse, this process continues until 10 – 12 weeks of age, when the mature ductal tree is established and completely fills the length of the fat pad (Smalley and Ashworth 2003; Sternlicht et al. 2006).

Microarray profiling of the TEBs versus mature epithelial ducts versus the stroma shows that GATA3 is the most highly expressed transcription factor in the mammary epithelium (Kouros-Mehr and Werb 2006). Using a mammary epithelium-specific knockout of *Gata3*, we and others have shown that GATA3 is necessary for mammary gland development (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007). Early deletion of *Gata3* specifically in the mammary epithelium using the murine mammary tumor virus (MMTV) promoter-Cre recombinase (MMTV-Cre) shows that prior to puberty, the rudimentary mammary gland is similar to wild-type littermates. However, with the onset of puberty, the mammary glands of *Gata3*-CKO mice fail to develop TEBs. In addition, the epithelium fails to invade into the stroma (Fig. 2.2B). When examined five weeks later, the outgrowths display gross structural defects, including irregular luminal diameters and deficiencies in side branching. Taken together, this suggests a role for GATA3 in ductal elongation and branching. Furthermore, the ductal epithelium that forms contains regions that lack luminal cells and regions that contain a multilayered luminal epithelium. Surprisingly, these outgrowths, which express Cre recombinase and have undergone recombination, retain a functional *Gata3*

allele, suggesting that there is a selective pressure to maintain GATA3 expression during development (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007).

Deletion of *Gata3* in the adult mammary gland after development has taken place using a doxycycline-inducible system reveals severe cellular defects in the luminal epithelium, including de-differentiation of the luminal cells, disorganization of the duct, a decrease in cell-cell adhesion and an increase in cell proliferation. These characteristics are strongly reminiscent of cancer cell properties important for cell invasion and metastasis. In this case, cell detachment into the lumen is followed by widespread cell death, likely due to the lack of basement membrane-derived survival signals. However, acute loss of *Gata3* leads to an expansion of a de-differentiated luminal cell population prior to cell death. Further analysis shows that these *Gata3*-deleted cells retain luminal character and do not transdifferentiate into myoepithelial cells. Long-term loss of *Gata3*, however, leads to caspase-mediated luminal cell death and lactational insufficiency. Remarkably, introduction of *Gata3* into a purified mammary progenitor-enriched population induces luminal cell differentiation. This suggests that GATA3 is necessary in the adult mammary gland to maintain the integrity and function of the luminal epithelium and sufficient to specify the luminal cell fate (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007).

How acute loss of *Gata3* in the mammary gland might lead to expansion of a de-differentiated epithelial cell population remains unknown, but this feature is also observed in the skin and the lens (Kurek et al. 2007; Maeda et al. 2009).

These studies suggest that there is a complex relationship between differentiation and regulation of the cell cycle. One might expect that GATA3 normally represses the cell cycle so that acute loss of GATA3 results in increased proliferation. However, a recent study shows that in contrast, GATA3 normally represses the cyclin inhibitor *p18^{INK4C}* (Pei et al. 2009). Low GATA3-expressing cells such as luminal progenitors express a high level of *p18^{INK4C}*, which functions to restrain progenitor cell proliferation. Thus, the mechanisms of how a GATA3-negative pool of cells expands and how GATA3 might regulate aspects of the cell cycle remain open questions.

Several direct downstream targets of GATA3 in the luminal epithelium have been identified. They include genes such as *Foxa1*, an important regulator of estrogen receptor (ER) expression, mucin, which may play a role in epithelial polarity, and the cyclin inhibitor *p18^{INK4C}*, which has been suggested to restrain luminal cell progenitors (Abba et al. 2006; Kouros-Mehr et al. 2006; Pei et al. 2009). Bioinformatic analysis of a mammary epithelial microarray dataset reveals additional epithelial-specific transcription factors that may cooperate with GATA3 in its gene regulatory network, which include MSX2, FOXP4, TRPS1, ELF5, EHF, RUNX1 (Kouros-Mehr and Werb 2006).

GATA3 in tumor differentiation and metastasis

Given the fundamental role of GATA3 in maintaining the differentiation and adhesion of the luminal epithelial cell, we and others hypothesized that loss of *Gata3* is causally involved in the pathogenesis of breast cancer. In the clinical

setting, breast tumors are commonly subdivided into estrogen receptor positive (ER+) and negative (ER-) tumors. Whereas ER+ tumors tend to be morphologically well differentiated and exhibit a relatively good prognosis, ER- tumors are poorly differentiated and exhibit a poor prognosis. To date, in a series of eleven independent microarray gene expression profiling studies of ER+ and ER- breast tumors, GATA3 emerges as a strong prognostic indicator of breast cancer. Low GATA3 expression is strongly associated with higher histologic grade, poor differentiation, positive lymph nodes, ER- and progesterone receptor (PR) negative status, and HER2/neu overexpression, all indicators of poor prognosis (Perou et al. 2000; Sorlie et al. 2001; Janssen et al. 2002; Sorlie et al. 2003; Mehra et al. 2005). While two tissue microarray studies have suggested that GATA3 status has independent prognostic significance in breast cancer (Mehra et al. 2005; Dolled-Filhart et al. 2006), this matter is still controversial. One study shows that while GATA3 expression has a strong association with ER, in a multivariate analysis of over 3,100 cases of invasive ductal carcinoma, GATA3 lacks independent prognostic value (Voduc et al. 2008), despite being associated with ER+ expression (Hoch et al. 1999). Nonetheless, the highest levels of GATA3 are observed in the “luminal A” subtype of breast cancer, which expresses luminal differentiation markers such as ER/PR and has the best prognostic outcome. Given the various subtypes of cancer, it is possible that GATA3 plays a different role in basal versus luminal type tumors.

GATA3 levels in various human breast cancer cell lines inversely correlate with their metastatic capability. Metastatic cell lines such as the MDA-MB-231

cells have low GATA3 levels, whereas non-metastatic cell lines such as the MCF7 cells have high GATA3 levels (Kouros-Mehr et al. 2008). Furthermore, in several mouse models, including the MMTV-PyMT (polyoma middle T antigen) and MMTV-Neu models which develop luminal breast cancer, loss of GATA3 correlates with loss of differentiation genes, the transition from adenoma to early carcinoma and the onset of tumor dissemination. The MMTV-PyMT model recapitulates many characteristics of human disease including the progression to metastasis (Lin et al. 2003). Late carcinomas and metastases invariably do not express GATA3. Interestingly, using a tumor transplant model, our laboratory showed that re-introduction of *Gata3* using a retrovirus into MMTV-PyMT late carcinoma cells that are then transplanted orthotopically into the mammary fat pad is sufficient to differentiate the tumor cells (Fig. 2.3). The tumor cells not only form lumens, but also express β -casein and basement membrane components such as perlecan, suggesting that the cells exhibit apical-basal polarity. Strikingly, re-introduction of *Gata3* suppresses tumor metastases to the lungs by over 25-fold, linking GATA3 to both tumor differentiation and metastasis (Kouros-Mehr et al. 2008). In addition, mutations in the zinc finger domains of GATA3, which diminish or abolish the ability of GATA3 to bind DNA, have also been identified in a subset of human breast cancers, further underscoring the importance of GATA3-regulated genes in breast cancer (Usary et al. 2004).

Whether GATA3 suppresses organ-specific metastasis is another topic of interest. We and others have shown that GATA3 suppresses pulmonary metastases from mouse and human mammary tumors (Kouros-Mehr et al. 2008;

Dydensborg et al. 2009). To extend these studies from the MMTV-PyMT mouse model to human cells, Dydensborg and colleagues expressed *GATA3* in the human-derived MDA-MB-231 LM2 cell line, which was generated by repeated rounds of in vivo selection for lung tropism (Minn et al. 2005). These authors demonstrate that while expression of *GATA3* does not affect the number of spontaneous lung metastases when the cells are implanted orthotopically in the mammary fat pad, there is a reduction in the number of cells that survive in the lungs when the cells are injected intravenously (i.v.) into the tail vein. In accordance with our previous work, this suggests that *GATA3* affects tumor cell survival at distant sites such as the lung. However, in this experimental model, the authors did not notice an increase in differentiation genes. This is possibly due to the use of MDA-MB-231 cells, which are an undifferentiated, aneuploid, triple negative (ER-/PR-/HER2-) cell line. To further understand the effects of *GATA3* on the LM2 cells, the authors performed microarray analysis to identify genes that are up- and down-regulated in the MDA-MB-231 control tumors versus the *GATA3*-expressing tumors. Interestingly *ID1* and *ID3*, which previously were shown to promote breast cancer metastasis by facilitating sustained proliferation during early stages of metastatic colonization (Gupta et al. 2007), are significantly down-regulated in *GATA3*-expressing cells. Whether *GATA3* directly controls the expression of these genes, however, remains to be determined. Additional work will be required to determine if *GATA3* suppresses metastases to other organs commonly affected in humans such as the bone or brain.

Does the loss of GATA3 expression in breast cancer induce tumor de-differentiation so that cells expand and acquire EMT-like characteristics? Since the loss of GATA3 in the mammary gland causes luminal cell proliferation and basement-membrane detachment, similar mechanisms may exist in breast cancer to accelerate malignant conversion and gain metastatic capability. However, when *Gata3* is deleted in early, well-differentiated tumors in the MMTV-PyMT model, the cells undergo caspase-mediated cell death, similar to deletion in normal epithelium. This suggests that premature loss of GATA3 is not sufficient to promote malignant progression, and is not tolerated in early tumors. Instead, a GATA3-negative, stem cell-like tumor population persisting in early tumors expands during tumor progression and is likely responsible for the transition to the GATA-negative state. Consistent with this notion, adenoma and carcinoma cells are progressively enriched for cell-surface markers of mammary stem-like cells. In addition, this stem-like population appears to be more motile when compared to the differentiated cells. Thus, although progression to a GATA3-negative state underlies the onset of tumor dissemination, further events are likely necessary for successful formation of metastases from disseminated cells. These data indicate that GATA3 is a crucial regulator of tumor differentiation and suppressor of tumor metastasis.

In luminal progenitor cells, GATA3 directly represses $p18^{INK4C}$ to regulate the cell cycle, and therefore, the levels of GATA3 must be carefully titrated in these cells (Pei et al. 2009). Low $p18^{INK4C}$ and high *GATA3* expression, which would result in increased cell proliferation, are simultaneously observed in

luminal A breast cancer. Mice deficient for *p18^{INK4C}* have an expanded luminal progenitor population throughout life and develop ER+ luminal tumors at a high penetrance. Because GATA3 is expressed, however, these cells differentiate into luminal epithelial cells, resulting in a well-differentiated tumor with a good prognosis (Pei et al. 2009). In contrast, the absence of GATA3 in a tumor results in a slowly proliferating but undifferentiated cancer. Thus, GATA3 is required not only for differentiation, but also to regulate cell proliferation in the mammary gland. However, the relationship between cell proliferation and GATA3 is probably more complex, and may depend on the differentiation status of the cell. Interestingly, luminal progenitors have also been suggested to be the candidate target cell population for basal tumor development in *BRCA1* mutation carriers (Lim et al. 2009).

In addition to regulating cell differentiation, adhesion and proliferation, GATA3 may influence tumor progression and metastasis in an indirect mechanism by affecting the microenvironment. The microenvironment plays important roles during both normal and cancer development (Coussens and Werb 2002). For example, macrophages found in the stroma immediately adjacent to the TEB help form and are often associated with collagen fibers to facilitate ductal branching; ablation of macrophages delays outgrowth and branching. Macrophages are also found within the TEB, where they clear apoptotic epithelial cells as the lumen is formed (Pollard 2009). In several mouse models of breast cancer including the MMTV-PyMT model, the loss of GATA3 coincides with the onset of angiogenesis and an increased number of tumor-

associated macrophages (TAMs). Analysis of MMTV-PyMT mammary tumors reveals a progressive increase in macrophages during tumor development. These macrophages regulate the angiogenic switch to promote tumor angiogenesis (Lin et al. 2006; DeNardo et al. 2009). These observations are interesting given the pro-tumorigenic roles of TAMs in a number of cancers; a high density of TAMs in human tumors correlates with poor prognosis in more than 80% of cases (Bingle et al. 2002). In accordance with this, experimental evidence shows that suppressing macrophage growth factors such as CSF1 can delay tumor progression and inhibit metastasis (Lin et al. 2001). In light of these studies, and the observation that GATA3 may repress components of the interferon response signature associated with metastasis, one could speculate that GATA3 may affect tumorigenesis and metastasis by influencing the tumor microenvironment. However, an understanding of whether this is a primary or secondary effect (e.g., due to inducing differentiation) awaits further study.

Identifying GATA3-regulated targets

An important goal to understanding the diverse roles of GATA3 on various cell types remains to identify downstream targets that mediate the effects of GATA3. How does GATA3 control cell fate decisions and what downstream targets are important for it to do so? One important step will be to conduct a comprehensive analysis of GATA3 binding sites by chromatin-immunoprecipitation (ChIP) followed either by microarray profiling (ChIP on chip) or by deep sequencing (ChIP-Seq).

One potential set of GATA3 targets is microRNAs (miRNAs). miRNAs are small, non-coding RNAs that serve to modulate gene expression post-transcriptionally by either inhibiting translation of their specific targets or causing degradation of the target mRNAs. miRNAs play a role in the maintenance of mouse mammary epithelial progenitor cells, and miRNAs such as let-7 promote mammary differentiation (Ibarra et al. 2007). Not surprisingly, miRNAs are also important regulators of tumor progression and metastasis, both in promoting and suppressing metastasis. For example, miR-10b is highly expressed in metastatic breast cancer cells and promotes cell migration and invasion. Overexpression of miR-10b in otherwise non-metastatic breast tumors initiates robust invasion and metastasis, and the level of miR-10b expression in primary breast carcinomas correlates with clinical progression (Ma et al. 2007). Other studies have identified miRNAs that inhibit breast cancer metastasis (Ma and Weinberg 2008; Tavazoie et al. 2008; Valastyan et al. 2009)).

Recently, two studies have shown that GATA1 regulates erythropoiesis by controlling miRNAs (Dore et al. 2008; Pase et al. 2009). Using a microarray screening approach, miR-144 and miR-451 were identified as GATA1-regulated miRNAs. The authors demonstrate that these two miRNAs are important for erythrocyte development, and show that in the zebrafish, one important function of miR-451 is to down-regulate *gata2*. These studies reveal that miRNAs lie directly downstream of GATA factors, uncovering a new mechanism by which GATA factors specify cell fate, and that regulation of miRNA loci may be key components to how these transcription factors function. Future studies

investigating the link between GATA3 and miRNAs may provide valuable new insight into mammary gland development and the pathogenesis of breast cancer, which may ultimately reveal novel therapeutic strategies.

CONCLUSION

The role of GATA3 in the differentiation of the mammary luminal cell adds to the growing body of evidence implicating the GATA family of transcription factors as key regulators of cell fate specification and maintenance. GATA3 promotes the differentiation of luminal cells, while repressing other cell types in the mammary gland such as adipocytes. A better understanding of how GATA3 regulates luminal cell differentiation will be important in breast cancer therapy and shed further light on its role as a prognostic factor. GATA3 defines a distinct class of cancer genes that are differentiation factors rather than conventional tumor suppressor genes, which affect the malignant phenotype by enforcing differentiation. This concept forms the basis for using high-dose retinoic acid therapy to restore tumor differentiation and achieve remission in acute promyelocytic leukemia patients, who carry a translocation that blocks myeloid differentiation. Therefore, uncovering paracrine or juxtacrine signals that activate GATA3 expression during luminal cell specification, as well as understanding the downstream targets of GATA3 will be critical to our understanding of tumor differentiation in breast cancer.

ACKNOWLEDGEMENTS

This study was supported by grants from the National Institutes of Health (CA129523 and ES012801 to Z.W.) and funds from the UCSF Medical Scientist Training Program and a California Breast Cancer Research Program Pre-doctoral Fellowship (to J.C.). We thank Charina Choi for helpful discussion and critical reading of this manuscript, and Pengfei Lu and Kai Kessenbrock for help with figures.

FIGURE LEGENDS

Fig. 2.1: Functional domains in the mouse GATA-binding protein 3. GATA3 is composed of 443 amino acids, and contains two amino terminal transactivation domains, TA1 and TA2, and two zinc-finger motifs, ZF1 and ZF2, which are followed immediately by a conserved basic region. The distal zinc-finger motif (ZF2) binds to DNA containing the canonical GATA motif, (A/T)GATA(A/G). The proximal zinc-finger motif (ZF1) seems to have broader specificity. Mutation of the amino acids KRR, which lies in the first basic region in between ZF1 and ZF2, confers dominant negative or hypomorphic function. The second basic region contains two important motifs: the YxKxHxxxRP motif (in which x denotes any amino acid) mediates DNA binding and the NRPL motif forms the interface between two GATA molecules, indicating that GATA3 may homo- or heterodimerize.

Fig. 2.2: GATA3 in normal mammary gland development. (A) A schematic representation of the mammary epithelium and stroma during mammary gland development. The luminal epithelial cells, highlighted in yellow, express GATA3 while the myoepithelial cells, highlighted in red, express very low levels of GATA3. The terminal end bud (TEB) invades through the stroma during pubertal development, and consists of both GATA3-negative cap cells and GATA3-positive cells. The stroma consists of a heterogeneous cell population of fibroblasts, adipocytes, macrophages, eosinophils, and mast cells, and plays an important role in facilitating branching and ductal elongation. Arrow points to the

direction of TEB migration. (B) Whole-mount carmine red staining of mouse mammary glands from five-week old wild-type (left) and GATA3 conditional knock-out (CKO) (right) mice outlines the epithelium. In the wild-type mammary gland, the invasion process into the stroma has begun (from left to right) past the lymph node (LN), with multiple TEBs formed. Inset shows bifurcating TEBs. In contrast, the GATA3-CKO mammary gland shows a defect in epithelial invasion into the stroma, without prominent TEBs formed. Inset shows a lack of TEBs at the distal end of the epithelium. Scale bar corresponds to 3 mm. (Figure 2A is modified from Lu and Werb, 2008 and Figure 2B is reprinted from Cell, 127, Kouros-Mehr et al., GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland, 2006, with permission from Elsevier.).

Fig. 2.3: GATA3 is sufficient to induce differentiation in carcinoma cells.

Primary cultures of adenocarcinomas from MMTV-PyMT mice were transduced with retrovirus containing either empty vector (control) or GATA3 and transplanted into wild-type mice. Tumors were grown for six weeks. H&E staining (A and B) and immunocytochemistry for β -casein (C and D) show that tumor cells expressing GATA3 differentiate and form milk proteins. Schematic shows that GATA3 tumors not only form lumens (E and F), but also express differentiation markers and basement membrane components such as perlecan (marked in blue). (Figure 3 A – D is reprinted from Cancer Cell, 13, Kouros-Mehr et al., GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model, 2008, with permission from Elsevier).

Table 2.1: The function of GATA3 in various tissues. GATA3 contributes to the differentiation and normal development of epithelial and non-epithelial tissues. Studies in T cells and the mammary gland show that GATA3 also plays an important role in tumorigenesis.

FIGURES



Figure 2.1

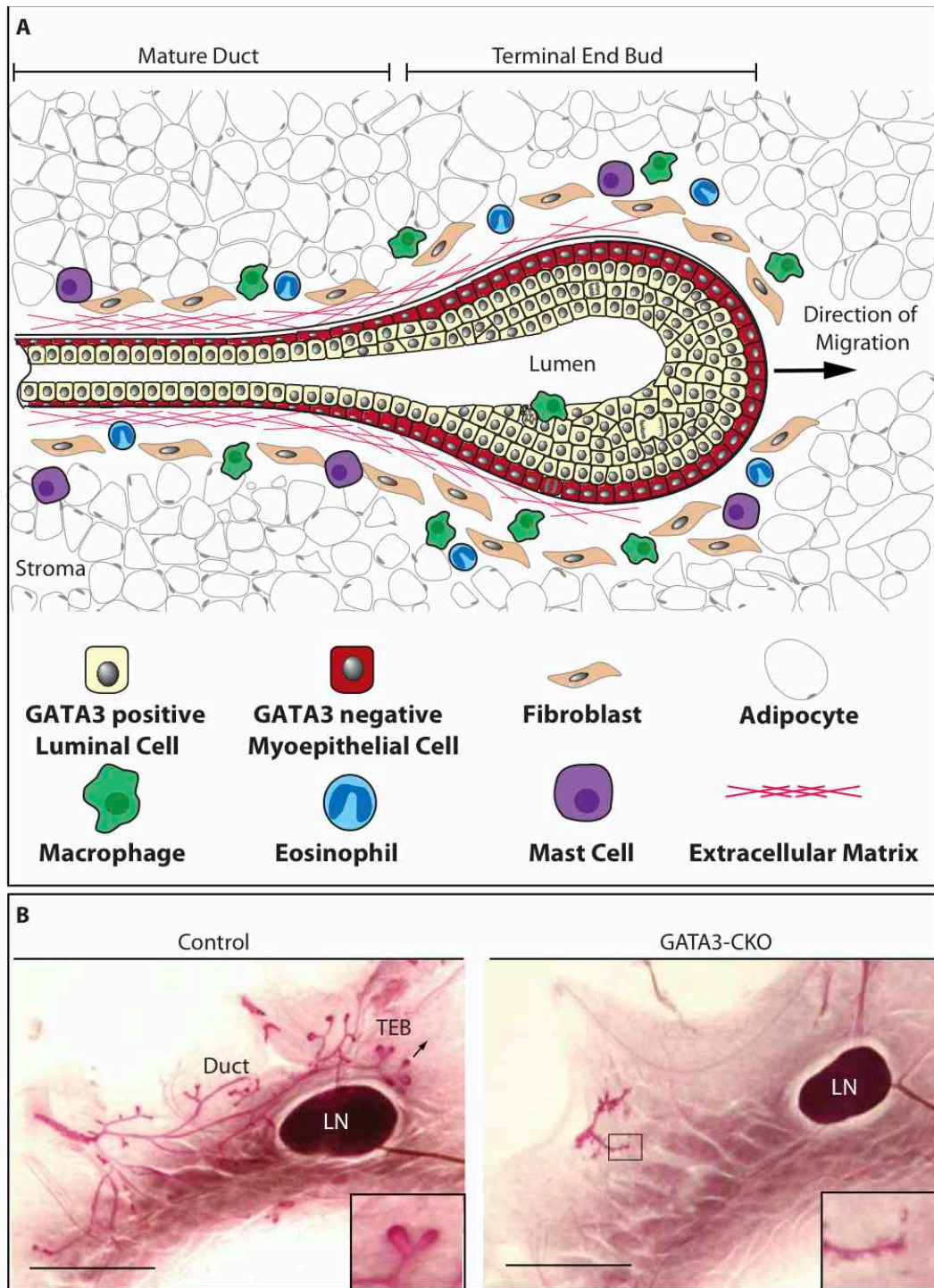


Figure 2.2

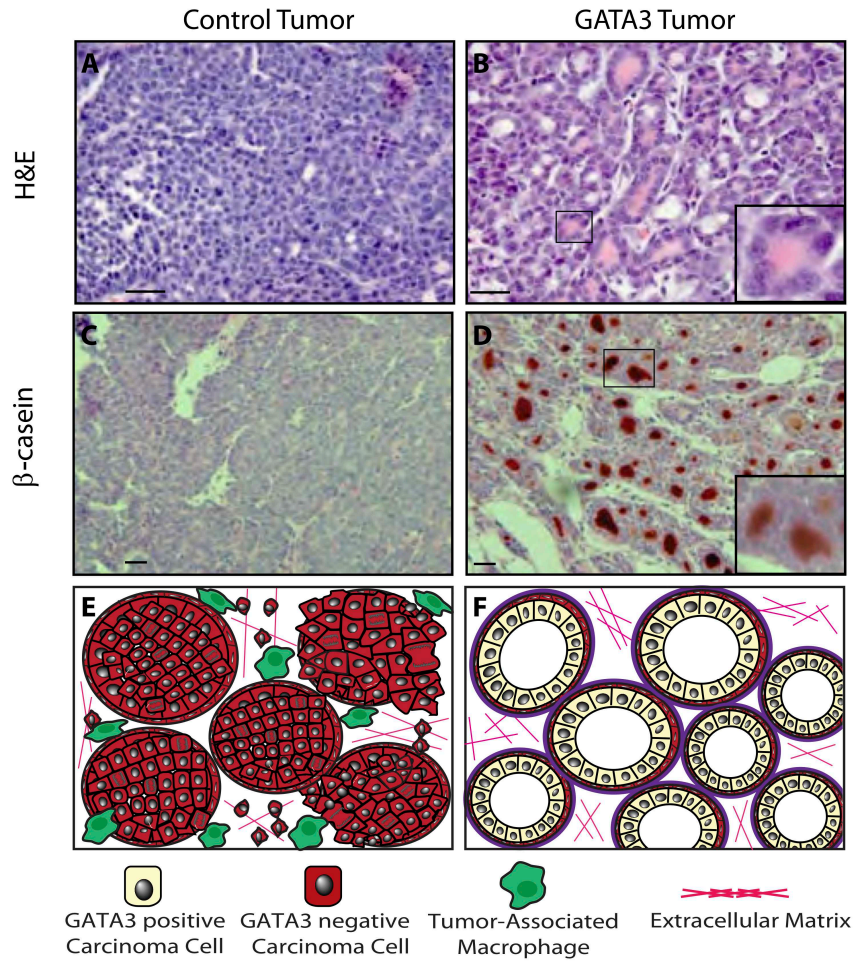


Figure 2.3

TABLE

Tissue	Functions of GATA3	Reference
T cell	<ul style="list-style-type: none"> • T_H2 cell specification • Cell survival during β-selection • Transforms T cells in cooperation with c-Myc and Notch 	(Grogan and Locksley, 2002; Ho et al., 2009; van Hamburg et al., 2008)
Skin	<ul style="list-style-type: none"> • Differentiation of the inner root sheath precursors • Decreases basal cell proliferation • Role unknown in skin cancer 	(de Guzman Strong et al., 2006; Kaufman et al., 2003; Kurek et al., 2007)
Mammary Gland	<ul style="list-style-type: none"> • Specifies and maintains differentiation, adhesion, proliferation of luminal cells • Low GATA3 expression correlates with worse breast cancer prognosis • Re-introduction into carcinoma cells suppresses metastasis to lung 	(Abba et al., 2006; Dydensborg et al., 2009; Kouros-Mehr et al., 2008; Kouros-Mehr et al., 2006; Pei et al., 2009)
Kidney	<ul style="list-style-type: none"> • Prevents ectopic metanephric kidney duct formation • Haploinsufficiency results in renal dysplasia • Role unknown in kidney cancer 	(Grote et al., 2008)
Lens	<ul style="list-style-type: none"> • Differentiation of lens fiber cells • Regulates cell cycle to suppress cellular proliferation 	(Maeda et al., 2009)
Neurons	<ul style="list-style-type: none"> • Maintains noradrenaline levels during development by up-regulating dopamine β-hydroxylase • Differentiation of sympathoadrenal and auditory ganglion neurons • Expressed in multiple areas of the brain, though function yet unknown 	(Hong et al., 2008; Jones and Warchol, 2009; Lim et al., 2000; Moriguchi et al., 2006; Zhao et al., 2008)
Adipocyte	<ul style="list-style-type: none"> • Antagonizes adipocyte differentiation • Loss of function associated with obesity 	(Tong et al., 2000)

Table 2.1

Chapter 3: GATA3 suppresses lung metastasis and modulates the tumor microenvironment by regulating miR-29b expression

Source: The following chapter is a manuscript currently in revision.

Contributions: For this project, I performed most of the experiments, with help from Jeffrey Lin, Audrey Brenot, and Sylvain Provot. Some initial observations were made by Jung-whan Kim and Sylvain Provot. I initiated experiments to investigate the role of microRNAs in regulating the tumor microenvironment. Zena Werb and I wrote the manuscript with editorial input from all authors. Zena Werb supervised the project.

GATA3 suppresses metastasis and modulates the tumor microenvironment by regulating miR-29b expression

Jonathan Chou^{1,2}, Jeffrey Lin^{1,3}, Audrey Brenot^{1,3}, Jung-whan Kim^{1,5}, Sylvain Provot^{1,6} and Zena Werb^{1,2,4}

¹Department of Anatomy, and ²Biomedical Sciences Program, University of California, San Francisco CA 94143-0452

³These authors contributed equally to this work.

⁴Corresponding Author:

Zena Werb, Ph.D.

Department of Anatomy, HSW 1323

University of California

San Francisco, CA 94143-0452 USA

Tel.: (415) 476-4622

Fax: (415) 476-4565

zena.werb@ucsf.edu

Keywords:

GATA3; metastasis; miR-29b; tumor microenvironment; tumor differentiation

Present Address:

⁵ Salk Institute for Biological Studies

10010 N. Torrey Pines Road

La Jolla, CA 92037

⁶ INSERM U606 Hôpital Lariboisière

2 Rue Ambroise Paré

75010 Paris, France

ABSTRACT

Despite advances in our understanding of breast cancer, patients with metastatic disease have poor prognoses. GATA3 is a transcription factor that specifies and maintains mammary luminal epithelial cell fate, and its expression is lost during breast cancer progression, correlating with worse prognosis in human patients. Here, using human and mouse breast cancer cells, we show that GATA3 promotes differentiation, suppresses metastasis and alters the tumor microenvironment by inducing miR-29b expression. Consistent with miR-29b being downstream of GATA3, miR-29b is enriched in luminal breast cancers and loss of miR-29b, even in GATA3-expressing cells, increases metastasis and promotes a mesenchymal phenotype. miR-29b inhibits metastasis directly by targeting a network of pro-metastatic regulators involved in angiogenesis, collagen remodeling and proteolysis, including *VEGF*, *ANGPTL4*, *PDGF*, *LOX* and *MMP9*. This discovery that regulating tumor microenvironmental genes through a GATA3-miR-29b axis inhibits cancer metastasis opens up new possibilities for therapeutic intervention in breast cancer.

INTRODUCTION

One of the classical “hallmarks of cancer” is the ability for tumor cells to invade and metastasize (Hanahan and Weinberg 2011). Metastasis is a multi-stage process that includes extracellular matrix remodeling, blood vessel recruitment, tumor cell entry and exit from circulation, and survival at a distant organ (Valastyan and Weinberg 2011). The tumor microenvironment is increasingly recognized as an important contributor to malignant progression and metastasis (Hanahan and Coussens 2012; Lu et al. 2012). In addition to remodeling the microenvironment to facilitate metastasis, cancer cells also turn on embryonic morphogenesis regulators to undergo the epithelial-to-mesenchymal transition (EMT) and turn off programs that maintain their differentiated state (Yang and Weinberg 2008). This acquisition of an EMT-like state allows cancer cells to gain motility, modify cell adhesion, and acquire stem-like properties (Mani et al. 2008). These findings support the concept that mechanisms controlling normal development play critical roles during tumor progression, and underscore the importance of identifying shared molecular programs.

A fundamental aspect of development is the specification and maintenance of differentiated cell types, which is mediated in part by hierarchical networks of transcription factors. The GATA family of transcription factors are such master regulators during the development of numerous cell types (Chou et al. 2010). We and others previously showed that GATA3 is the master transcription factor that specifies and maintains luminal cell fate in the mouse mammary gland (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007). Loss of

GATA3 is causally involved in the pathogenesis of breast cancer (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007; Kouros-Mehr et al. 2008). Expression profiling studies of breast cancer patients established that low *GATA3* is associated with poor prognosis (Mehra et al. 2005; Jacquemier et al. 2009; Yoon et al. 2010). In several mouse models of breast cancer, including the mammary tumor virus (MMTV) promoter-driven polyoma middle T antigen (PyMT) and MMTV-Neu models, *Gata3* inversely correlates with tumor progression and metastasis: loss of *GATA3* coincides with loss of differentiation, the transition from adenoma to carcinoma and the onset of tumor dissemination. Re-introduction of *Gata3* into MMTV-PyMT carcinomas transplanted orthotopically induces differentiation, suppresses dissemination (Kouros-Mehr et al. 2008) and reduces tumor-initiating capacity (Asselin-Labat et al. 2011). Furthermore, mutations in the zinc fingers of *GATA3*, which diminish or abolish DNA binding ability, have been identified in a subset of human breast cancers, emphasizing the importance of *GATA3*-regulated genes (Usary et al. 2004). However, how *GATA3* induces tumor differentiation and inhibits dissemination, and what molecular and cellular events lie downstream of *GATA3* are largely unknown.

MicroRNAs (miRNAs) are small, non-coding RNAs that modulate gene expression post-transcriptionally, by either inhibiting translation or causing degradation by binding the 3' untranslated regions (UTRs) of target mRNAs (Bartel 2009). miRNAs can control the expression of hundreds of genes, and are both positive and negative regulators of cancer metastasis (Ma et al. 2007; Tavazoie et al. 2008; Valastyan et al. 2009). Precedence for GATA-mediated

miRNA regulation has recently been established: GATA1 promotes erythrocyte differentiation through miR-451, suggesting that GATA factors utilize miRNAs to make cell fate decisions (Dore et al. 2008). In the mammary gland, miRNAs such as *let-7* promote mammary differentiation and regulate self-renewal (Ibarra et al. 2007; Yu et al. 2007). However, miRNAs downstream of GATA3 have yet to be investigated.

In this study, we hypothesized that GATA3 coordinates gene expression networks involved in metastasis through miRNA-mediated mechanisms. We investigated the molecular pathways by which GATA3 regulates cancer differentiation and metastasis, and identified miR-29b, a downstream miRNA that modulates the tumor microenvironment, metastasis and epithelial plasticity.

RESULTS

GATA3 suppresses lung metastases from human and mouse breast cancer

Based on expression array analysis, human breast cancers are classified into several subtypes that are prognostic of clinical outcome (Sorlie et al. 2001). Clinically, basal, triple negative breast cancers are aggressive and poorly differentiated (Foulkes et al. 2010). In human breast cancer cell lines (Neve et al. 2006), GATA3 is expressed higher in luminal versus basal A and basal B subtypes (**Fig. 3.1a**), consistent with its luminal localization (**Supplemental Fig. S3.1a**).

To determine how GATA3 influences cancer progression, we over-expressed GATA3 in 4T1 cells, a metastatic, murine, triple-negative, basal line, to evaluate metastasis in an immunocompetent model, as well as in MDA-MB-231 cells (referred to as MDA231), a claudin-low human line with low GATA3 expression (**Fig. 3.1b, Supplemental Fig. S3.1b**). Orthotopic transplant of 4T1-Gata3 cells in Balb/C mice gave rise to tumors similar in size to 4T1-Control cells, but showed a significant decrease in the number and size of spontaneous lung metastases (**Fig. 3.1c–3.1e**). We did not observe any difference in proliferation in vivo or in culture (**Fig. 3.1f, 3.1i**), suggesting that GATA3 cells have no intrinsic defect in proliferation, and that the difference in metastasis may depend on in vivo interactions with the microenvironment. Indeed, 4T1-Gata3 tumors exhibited a significant decrease in tumor vasculature and macrophage infiltrates (**Fig. 3.1g, 3.1h**), suggesting that GATA3 affects the tumor microenvironment. Consistent with this, we found a two-fold reduction in serum

VEGF levels in mice bearing orthotopic 4T1-Gata3 mice tumors (**Fig. 3.1j**), suggesting that GATA3 regulates VEGF in vivo.

Human MDA231-GATA3 cells injected into nude mice formed smaller tumors than MDA231-Control cells, with decreased tumor vasculature and macrophage infiltrates. MDA231-GATA3 cells also secreted less VEGF and showed no difference in proliferation in vitro (**Supplemental Fig. S3.1c–S3.1j**). These data support the concept that in vivo interactions with the microenvironment are important in GATA3-mediated metastasis suppression.

We next asked if GATA3 only limits the ability of tumor cells to disseminate from the primary site, or if it also affects the later stages of metastasis, e.g., colonization or survival at the new organ. Accordingly, we inoculated the cells directly into the circulation by intravenous injection (i.v., via tail vein) to form experimental metastases. For both 4T1 and MDA231 cells, GATA3 decreased lung metastases (**Fig. 3.1k, Supplemental Fig. S3.1k**). Thus, GATA3 also affects the late steps of metastasis.

To determine whether GATA3-expressing cells can compete with control cells, we labeled 4T1 cells with RFP (control) or Gata3-GFP, mixed them in a 1:1 ratio, and co-injected them i.v. After 2 weeks, we found that 4T1-Gata3-GFP cells accounted for ~20% of the total metastatic cells in the lung, whereas 4T1-RFP cells accounted for ~80% (**Fig. 3.1l**). These results indicate that compared to 4T1-Control cells, 4T1-Gata3 cells are at a competitive disadvantage in vivo.

GATA3 promotes differentiation and limits cell migration

Since GATA3 regulates luminal fate, we first sought to determine whether the mechanism by which GATA3 suppresses metastasis involves cell differentiation. In two-dimensional cell culture, MDA231-Control cells had a more spindle-shaped, mesenchymal morphology, whereas MDA231-GATA3 cells displayed a more epithelial phenotype (**Fig. 3.2a**). In a wound-scratch assay, MDA231-Control cells migrated as single cells to close the wound, whereas MDA231-GATA3 cells migrated as a collective sheet (**Supplemental Fig. S3.2a**). In three-dimensional (3D) Matrigel culture, which better mimics physiological conditions (Debnath and Brugge 2005; Lee et al. 2007), MDA231-GATA3 cells were less invasive than MDA231-Control cells (**Fig. 3.2b, Supplemental Movies 1 and 2**). We confirmed these results in another basal breast cancer line, Hs578T (**Supplemental Fig. S3.2b–S3.2d**).

We next used quantitative PCR (qPCR) to determine the expression of genes associated with luminal differentiation and the EMT program. MDA231-GATA3 cells had increased expression of luminal genes and decreased expression of EMT, pro-metastatic, stemness and inflammatory genes (**Fig. 3.2c; Supplemental Fig. S3.2e**). Using CD49f and EpCAM to determine the degree of differentiation along the mammary hierarchy (Visvader 2009), we found that the majority of MDA231-GATA3 cells were CD49f^{med}/EpCAM^{pos}, and had decreased CD49f (integrin α 6), indicative of a more luminal phenotype, whereas MDA231-control cells were CD49f^{hi}/EpCAM^{neg} (**Fig. 3.2d–3.2f**). MDA231-GATA3 cells also formed fewer and smaller tumor spheres from single cells (**Fig. 3.2g, 3.2h**). Taken together, these results indicate that GATA3 promotes a more

differentiated expression program and opposes the EMT/stem-like state.

GATA3 promotes miR-29b expression

In addition to the cell-intrinsic effects of GATA3 on differentiation, we were intrigued by the observation that GATA3-expressing tumors had reduced serum VEGF and tumor vasculature. However, the *Vegfa* promoter lacked clear GATA3-binding sites, and co-transfection of GATA3 did not decrease a luciferase reporter containing the *Vegfa* promoter (data not shown). We hypothesized that GATA3 regulates *Vegfa* expression and perhaps other pro-metastatic genes indirectly via miRNAs, and conducted a screen in MDA231 cells \pm GATA3 using qPCR miRNA arrays. Of 88 miRNAs evaluated, miR-29b was the most up-regulated miRNA in MDA231-GATA3 cells (**Fig. 3.3a, 3.3b, Supplemental Table S3.1**). GATA3 also increased miR-29b in Hs578T-GATA3 cells (**Supplemental Fig. S3.3a**).

The miR29a/b1 promoter, which was previously identified (Mott et al. 2010), contains two GATA3 binding sites (**Supplemental Fig. S3.3b**). In MDA231 cells, GATA3 increased the activity of a miR29a/b1-promoter reporter, further demonstrating that GATA3 induces miR-29b expression (**Fig. 3.3c**). Significantly, the 3' UTR of *Vegfa* contains a conserved binding site for miR-29b (**Fig. 3.5a**).

Since previous studies showed that miR-29b is negatively regulated by TGF β and NF- κ B (Wang et al. 2008a; Mott et al. 2010; Winbanks et al. 2011), we asked whether GATA3 inhibits the TGF β and NF- κ B pathways, thereby also

regulating miR-29b indirectly. Indeed, GATA3 inhibited TGF β and NF- κ B reporter activities and suppressed TGF β -induced EMT by inhibiting components of the TGF β pathway transcriptionally (**Fig. 3.3d, 3.3e, Supplemental Fig. S3.3c–S3.3e**). Moreover, stimulation with TGF β , TNF α or sRANKL decreased miR-29b expression in control, but not GATA3 cells (**Supplemental Fig. S3.3f** and data not shown). Thus, GATA3 induces miR-29b expression directly (by binding the GATA sites) and indirectly (by inhibiting the TGF β and NF- κ B pathways).

miR-29b expression is higher in more differentiated breast cancer and normal luminal epithelial cells

We then asked if miR29 correlates with more differentiated breast cancers, which exhibit better patient outcomes (Mehra et al. 2005). The miR29 family consists of three members with the same seed sequence: miR29a, miR-29b, and miR29c (**Supplemental Fig. S3.3g**). In a miRNA dataset of 99 primary human breast tumors (Blenkiron et al. 2007), all three miR29 members were higher in luminal, and estrogen receptor-positive (ER+) tumors (**Fig. 3.3f–3.3g**). Moreover, miR29c was associated with more favorable prognoses in a meta-analysis of over 1000 breast cancers (Buffa et al. 2011), and with luminal differentiation in an independent dataset of over 100 human breast tumors (Enerly et al. 2011).

We confirmed these findings in mouse models of breast cancer using a miRNA dataset of mouse breast tumors (Zhu et al. 2011). All miR29 members were expressed higher in luminal models compared to basal models (**Fig. 3.3h**).

In normal cells isolated from the mouse mammary gland, all miR29 members were higher in the luminal population (**Supplemental Fig. S3.4a–S3.4c**).

We next asked whether miR-29b is inversely related to metastatic ability. Using a dataset of syngeneic cell lines with varying metastatic capabilities (Aslakson and Miller 1992; Dykxhoorn et al. 2009), we found that expression of miR-29b decreased in cells with the highest metastatic capacities (**Fig. 3.3i, Supplemental Fig. S3.4d**). In the MMTV-PyMT model, which mimics progressive stages of human luminal breast cancer (Lin et al. 2003), adenoma and carcinoma cells had decreased levels of *Gata3* and miR-29b compared to normal epithelium (**Fig. 3.3j**), with the lowest levels at the carcinoma stage coinciding with metastasis. Together, these results demonstrate that miR29 expression correlates with more favorable outcomes in human patients, more differentiated phenotypes in normal and cancer cells, and reduced metastatic potential.

miR-29b promotes and maintains luminal differentiation

Since GATA3 induces miR29 expression and miR29 correlates with luminal differentiation, we asked whether miR29 itself promotes luminal characteristics. MMTV-PyMT luminal tumors become less differentiated during tumor progression (Lin et al. 2003). Over-expression of miR-29b in a primary cell line (PyMT) derived from a late-stage MMTV-PyMT carcinoma increased luminal marker expression including *Gata3* and repressed basal markers (**Fig. 3.4a**). In

addition, miR-29b promoted branching of PyMT cell aggregates in 3D Matrigel (**Fig. 3.4b**), a feature of normal branching morphogenesis (Ewald et al. 2008).

We next asked if loss of miR-29b induces a de-differentiated, mesenchymal phenotype. We generated miR-29b knockdown cells using miRZip29b lentiviruses, which express anti-sense hairpins that provide stable knockdown of endogenous miRNAs. While miR-29b levels decreased as expected, we also observed decreases in miR29a and miR29c levels, likely due to their similar sequences (**Supplemental Fig. S3.4e**). MDA231-Zip29 cells were more elongated, protrusive and spindle-like than MDA231-Control cells and associated with increases in CD49f and CD29 (**Fig. 3.4c, 3.4d**), two markers of the basal/stem cell population, both of which have miR-29b binding sites in their 3' UTRs. We also observed changes consistent with de-differentiation in mouse tumor cells after miR29 knockdown (**Supplemental Fig. S3.4f, S3.4g**).

We extended our findings using normal human mammary cell line (HMLE), which displays phenotypic plasticity in vitro (Mani et al. 2008). HMLE-Zip29 cells were more elongated and spindle-like and expressed higher levels of mesenchymal markers and lower levels of differentiation markers (**Fig. 3.4e, 3.4g**). Because the EMT program has been linked to stem-like traits (Mani et al. 2008), we examined whether miR29 knockdown altered expression of stem cell markers. We found that miR29 loss increased the CD44^{hi}/CD24^{low} stem cell-enriched compartment (**Fig. 3.4f**). These results indicate that similarly to GATA3, miR-29b promotes differentiation, and that loss of miR-29b causes de-

differentiation, accompanied by an increase in mesenchymal markers characteristic of a progenitor-like state.

miR-29b down-regulates pro-metastatic genes that modify the tumor microenvironment

We used prediction algorithms to generate a list of candidate miR-29b targets (Krek et al. 2005; Lewis et al. 2005; Betel et al. 2008). Disappointingly, genes associated with differentiation and the EMT program did not have miR-29b binding sites in their 3' UTRs. However, a number of microenvironmental genes involved in angiogenesis, collagen remodeling, matrix degradation, and ECM signaling did contain miR29 binding sites in their 3' UTRs (**Fig. 3.5a**). Many of these genes, including *ANGPTL4*, *LOX*, *MMP2*, *MMP9*, *PDGF*, *TGFB* and *VEGFA*, have been implicated previously in promoting tumor metastasis by remodeling the microenvironment to be more permissive for invasion, angiogenesis, and metastasis (Padua et al. 2008; Lu et al. 2012). To determine if miR29 modulates these genes directly, we cloned their 3' UTRs into luciferase reporters and co-transfected cells with these reporters and either a miR-29b or a control mimic. For all 3' UTRs tested, but not the empty control vector, miR-29b decreased luciferase activity by 40–80% (**Fig. 3.5b**).

We then transduced 4T1 and MDA231 cells with a miR-29b retrovirus and confirmed miR-29b over-expression. In both 4T1 and MDA231 cells, miR-29b repressed endogenous mRNA expression of microenvironmental targets (**Fig. 3.5c, 3.5d, Supplemental Fig. S3.5a–S3.5b**). We validated three targets (VEGF,

LOX, and ANGPTL4) by Western blot and found decreased protein levels in miR-29b over-expressing cells (**Fig. 3.5e**). These results suggest that miR-29b down-regulates a cohort of pro-metastatic genes involved in modifying the tumor microenvironment.

miR-29b expression suppresses lung metastasis

To test whether miR-29b suppresses metastasis, we injected mice with 4T1-Control or 4T1-miR-29b cells orthotopically. While there was no difference in primary tumor size, 4T1-miR-29b tumors had significantly fewer blood vessels and decreased fibrillar collagen (**Fig. 3.6a–3.6c**). Importantly, mice with 4T1-miR-29b tumors had fewer and smaller metastases (**Fig. 3.6d**). Furthermore, miR-29b significantly decreased experimental lung metastases (**Fig. 3.6e**). Similarly, miR29 reduced metastases in mice injected with MDA231-miR-29b cells (**Supplemental Fig. S3.5c–S3.5e**).

Because miR-29b levels decrease during tumor progression in MMTV-PyMT mice, we asked whether any direct targets suppressed by miR29 increase in parallel. At the carcinoma stage when miR-29b levels were lowest, *Vegfa* and *Lox* mRNA increased (**Fig. 3.6f**). Moreover, increasing miR-29b expression in PyMT cells decreased VEGF and LOX, and significantly reduced lung metastasis (**Fig. 3.6g–3.6i, Supplemental Fig. S3.5f, S3.5g**). Conversely, loss of miR29 in PyMT, 4T1 and MDA231 cells increased lung metastases (**Fig. 3.6j, 3.6k, Supplemental Fig. S3.5h, S3.5i**). Taken together, these results show that miR29 functions as a metastasis suppressor in human and mouse breast cancer.

miR-29b suppresses metastasis by repressing microenvironmental targets

To gain further insight into how miR-29b suppresses metastasis, we examined which targets increase after miR29 knockdown. Most of the miR-29b target genes we identified were up-regulated after miR29 loss, including *ANGPTL4*, *LOX*, *MMP9*, *VEGF* and CD49f and CD29 (**Fig. 3.7a, 3.4d**). Moreover, miR-29b knockdown increased the 3' UTR luciferase reporters of miR-29b targets (**Fig. 3.7b**), further validating these as bona fide miR-29b targets.

We focused on four genes down-regulated by miR-29b with known roles in angiogenesis and ECM remodeling to test whether restoring their expression would increase metastasis. We generated 4T1-miR-29b cells re-expressing *ANGPTL4*, *LOX*, *Mmp9*, or *Vegf*. While miR-29b alone inhibited metastasis, re-expression of each target restored lung metastasis (**Fig. 3.7c, 3.7d**), indicating that miR-29b suppresses metastasis, at least in part, by regulating these microenvironmental genes.

Regulation of miR-29b mediates the ability of GATA3 to suppress metastasis and promote differentiation

To determine whether GATA3-mediated metastasis suppression and differentiation requires miR-29b, we knocked-down miR-29b in MDA231-GATA3 cells. While GATA3 over-expression promoted epithelial clustering, concomitant loss of miR-29b abrogated GATA3's effects, resulting in an elongated, spindle-like morphology (**Fig. 3.8a**). Similarly, loss of miR29 in T47D luminal breast

cancer cells that express high levels of endogenous GATA3 caused a mesenchymal morphology, accompanied by an acquisition of EMT markers (**Supplemental Fig. S3.6a, S3.6b**). Thus, tumor cells gain mesenchymal characteristics in the absence of miR-29b, regardless of GATA3 expression.

We then asked whether GATA3 down-regulated miR-29b targets. GATA3 decreased both the expression and 3' UTR reporter activities of miR29 targets (**Fig. 3.8b–3.8c**), including *ANGPTL4*, *LOX*, *MMP9* and *VEGF*. Significantly, miR-29b loss increased metastasis despite GATA3 expression. We injected i.v. MDA231-GATA3-Zip29 and 4T1-Gata3-Zip29 cells and observed that while GATA3 suppressed metastasis, concomitant loss of miR-29b restored lung metastasis (**Fig. 3.8d, Supplemental Fig. S3.8c–S3.8d**). Similar results were obtained with orthotopic 4T1-Gata3-Zip29 primary tumors (data not shown). These results demonstrate that miR-29b is important downstream of GATA3, and that perturbing miR-29b attenuates the ability of GATA3 to suppress metastasis and promote differentiation.

Together, our data show that miR29 is an important node downstream of GATA3 that controls differentiation and the expression of pro-metastatic genes involved in modifying the tumor microenvironment, ultimately leading to metastasis suppression (**Fig. 3.8e**).

DISCUSSION

By activating and repressing gene expression, transcription factors orchestrate the establishment and maintenance of cell identity. Studies of human breast cancers have shown that GATA3 correlates with E-cadherin and ER expression and longer disease-free survival (Jacquemier et al. 2009). In this study, we show that GATA3 promotes luminal differentiation and suppresses lung metastasis through miR-29b, an anti-metastatic microRNA that regulates the tumor microenvironment (**Fig. 3.8e**). We demonstrate that GATA3 increases expression of miR-29b, which is exquisitely positioned to inhibit several steps required for metastasis: A network of mRNAs involved in angiogenesis, proteolysis, ECM signaling, and ECM remodeling have miR-29b binding sites in their 3' UTRs and are bona fide miR-29b targets. Because miRNAs act pleiotropically and modulate a range of biological processes, they represent an ideal set of targets through which transcription factors such as GATA3 can operate.

We found that miR-29b-mediated metastasis suppression depended on repression of targets including *ANGPTL4*, *LOX*, *MMP9*, and *VEGF*. Of these, *ANGPTL4* affects lung seeding by disrupting endothelial cell junctions (Padua et al. 2008); *LOX* increases tissue fibrosis and integrin-mediated survival signaling (Levental et al. 2009); *MMP9* releases VEGF from the ECM (Du et al. 2008); and *VEGF* promotes angiogenesis. In breast cancer, miR-29b expression is highest in good prognostic, luminal-type cancers and inhibits metastasis, the major cause of cancer-related deaths. miR-29b promoted luminal differentiation; conversely, loss of miR-29b resulted in the acquisition of mesenchymal traits and increased

metastasis. Concomitant loss of miR-29b in GATA3-expressing cells restored metastasis and abrogated the ability of GATA3 to promote differentiation, indicating that miR-29b is an important node that exerts its anti-metastatic effects by cell-intrinsic and extrinsic properties. Interestingly, dysregulation of miR29 has been implicated in various diseases, including tissue fibrosis (van Rooij et al. 2008; Roderburg et al. 2010; Wang et al. 2011). miR29 is decreased in many cancers, including lung cancer, liver cancer, rhabdomyosarcoma and melanoma (Fabbri et al. 2007; Wang et al. 2008a; Xiong et al. 2009; Nguyen et al. 2010).

The loss of GATA3 triggers fibroblastic transformation and cell invasion (Yan et al. 2010b), and coincides with the onset of angiogenesis, recruitment of inflammatory cells, and dissemination (Kouros-Mehr et al. 2008). We showed here that GATA3 reduces tumor vasculature, macrophage infiltrates and fibrillar collagen, components of the tumor microenvironment that promote metastasis. GATA3 also decreases macrophages in lung metastases, and endothelial cells in primary tumors (Asselin-Labat et al. 2011; Chu et al. 2012). These observations suggest that promoting differentiation in primary tumors limits metastasis by both non-cell autonomous mechanisms (e.g., angiogenesis, inflammation, collagen remodeling) and cell autonomous mechanisms (e.g., cell adhesion through E-cadherin).

TGF β signaling activates the promoters of many of the same genes that are miR-29b targets, including *LOX*, *MMPs*, and *VEGF* (Dumont and Arteaga 2000; Sanchez-Elsner et al. 2001; Sethi et al. 2011). We found that GATA3 inhibited TGF β signaling. Therefore, to repress TGF β -induced activation of its

target genes effectively, GATA3 inhibits the signaling network directly at the transcriptional level and at the post-transcriptional level through miR-29b (i.e., degradation and translational inhibition of already-made mRNAs). Furthermore, in preliminary studies, we find that miR-29b targets *TGFB* itself, which suppresses further paracrine signaling. This multi-tiered system ensures efficient GATA3-mediated inhibition of TGF β signaling. Interestingly, miR-29b also increased GATA3 expression, suggesting that GATA3 and miR-29b form a positive feedback loop and act collaboratively to reinforce fate decisions.

Whether miR29 is operative in other GATA3-regulated tissues remains to be determined. In thymocytes, miR-29b promotes T helper 2 (Th2) differentiation by repressing T-bet (Steiner et al. 2011); GATA3 also specifies Th2 fate (Pai et al. 2003), mimicking our findings in mammary cells. Although we did not find direct miR-29b binding sites in classical EMT genes, we found several candidate genes involved in differentiation that have miR-29b sites, including *DNMT3A/3B* and the *TET* family (*TET1/2/3*). Future work will be geared towards understanding if these epigenetic regulators affect EMT and differentiation in breast cancer cells. The exact mechanisms of how GATA3 and miR29 function in cell-type specific manners and their downstream targets remain to be further investigated.

Our work on miR-29b adds to the growing body of evidence implicating miRNA-mediated regulation of cancer and metastasis. Previous studies have shown that microRNAs (e.g., miR-31, miR-126, miR-335, the miR-200 family, the let-7 family) play important anti-metastatic roles by regulating diverse cellular

processes utilized during metastasis (Hurst et al. 2009; Valastyan et al. 2009; Lujambio and Lowe 2012). Interestingly, miR10b, previously shown to be pro-metastatic, was decreased with GATA3 in our screen. Future work aimed at identifying other GATA3 targets and determining their function will allow us to sort out miR29-dependent and miR29-independent effects of GATA3.

Rather than functioning as a classical tumor suppressor, GATA3 defines a distinct class of pro-differentiation factors capable of also affecting the tumor microenvironment. The identification of one GATA3 target, miR-29b a pleiotropic microRNA, illustrates how epithelial plasticity, the tumor microenvironment and metastasis are linked. Finally, microRNAs are being pursued as potential cancer therapeutics (Kota et al. 2009; Ma et al. 2010a). Whether increasing miR-29b levels in primary breast tumors will affect survival will determine if miR-29b mimics can be used therapeutically.

ACKNOWLEDGEMENTS

We thank members of the Werb laboratory for discussions and thank Elena Atamaniuc, Ying Yu, and Helen Capili for technical assistance. We also thank Charina Choi for discussions and encouragement. We thank Tara Rambaldo for flow cytometer assistance, Jayanta Debnath, Gabriele Bergers, and Dean Sheppard for discussions, and Andrei Goga, Valerie Weaver, Justin Mott and Elaine Raines for reagents. This research was supported by funds from the National Cancer Institute (R01 CA129523 to Z.W.), a Developmental Research grant from the Bay Area Breast Cancer SPORE (P50 CA058207 to Z.W.), a Department of Defense Predoctoral Fellowship (W81XWH-10-1-0168 to J.C.) and the UCSF Medical Scientist Training Program (J.C.). We dedicate this work to the memory of Libby Verber.

AUTHOR CONTRIBUTIONS

J.L. and A.B. contributed equally to this work. J.C., J.L., A.B., J.K., and S.P. designed and performed experiments. Z.W. designed experiments and supervised research. J.C. and Z.W. wrote the manuscript, and all authors discussed the results and provided comments and feedback throughout the entire preparation process.

COMPETING FINANCIAL INTERESTS

The authors declare no competing financial interests.

EXPERIMENTAL PROCEDURES

Animal Studies

All animal experiments were performed at UCSF, and reviewed and approved by the UCSF IACUC. Mice were housed under pathogen-free conditions in the UCSF barrier facility. Balb/c and nude mice were purchased from Simonsen Laboratories. FVB/n mice, originally from Jackson Laboratories, were bred in-house. For experimental metastasis experiments, age-matched female mice were injected i.v. (via tail vein) with 1×10^5 cells (4T1), 5×10^5 cells (MDA-MB-231) or $5\text{--}10 \times 10^5$ cells (PyMT) in PBS. For primary tumors and spontaneous metastasis assays, age-matched female mice were injected with the indicated number of cells in 1:1 DMEM:Matrigel (BD Biosciences) into the fourth mammary fat pad without clearing. Tumor measurements were made using a caliper at least once per week, and volumes were calculated using $V = 0.52 \times (\text{length})^2 \times \text{width}$. Bioluminescence imaging was performed using an IVIS Spectrum (Caliper Life Science). Image radiance values were normalized using the Living Image software.

Cell culture

MDA-MB-231, T47D, Hs578T, 4T1, and 4TO7 cells were obtained from the ATCC, LBNL or UCSF Cell Culture Facility, and grown in standard conditions (DMEM with 10% FBS). The immortalized human mammary epithelial cells (HMLE) were maintained as previously described (Elenbaas et al. 2001). The PyMT cell line was generated by isolating a late stage tumor from the MMTV-

PyMT/FVB mouse, dissociating the cells in collagenase, and culturing in DMEM/F12 media supplemented with 5% FBS, insulin and hydrocortisone. For TGF β stimulation, cells were serum-starved for 18–24 hr prior to adding TGF β 1, TGF β 2 or TGF β 3 (R&D Systems) at the indicated concentrations. For cells embedded in Matrigel, cells were aggregated overnight on ultra-low attachment plates (Corning) and then embedded into growth factor-reduced Matrigel (BD Biosciences) the next day. Cells were grown in serum-free conditions, supplemented with insulin-transferrin (Invitrogen) and EGF (Invitrogen) or FGF2 (Sigma).

Lentiviral and Retroviral Production

Viral production was carried out using calcium phosphate-mediated transfection of HEK 293T or GP2 cells. Virus was concentrated by ultracentrifugation. Cells were transduced with virus and then selected in puromycin, G418 or hygromycin for at least 5 days or selected by FACS.

Plasmids

Several expression plasmids were obtained from Addgene including: pcDNA-miR-29b from Joshua Mendell (plasmid 21121)(Hwang et al. 2007), pcDNA-GATA3 from Gokhan Hotamisligil (plasmid 1332)(Tong et al. 2000), p3TP-Lux from Joan Massague (plasmid 11767)(Wrana et al. 1992), SBE4-Luc from Bert Vogelstein (plasmid 16495)(Zawel et al. 1998), pWZL Blast VEGF from Robert Weinberg (plasmid 10909)(Watnick et al. 2003), and pBabe Angptl4 from Joan

Massague (plasmid 19156) (Padua et al. 2008). The pBM-IRES-puro and pBM-Mmp9 plasmids were gifts from Elaine Raines (Gough et al. 2006), the pMSCV-miR29 plasmid a gift from Andrei Goga, the pLV-LOX plasmid a gift from Valerie Weaver and the pGL3-miR29-Luc plasmid (containing the miR29a/b1 promoter) a gift from Justin Mott (Mott et al. 2010). For the 3' UTR luciferase reporters, the 3' UTRs were PCR amplified from mouse genomic DNA and cloned first into pCR2.1 by TOPO Cloning (Invitrogen) and verified by sequencing. Fragments were then digested with XhoI, SgfI and/or NotI and cloned into the siCheck2 reporter (Promega). Primer sequences used to generate 3' UTRs are detailed in Supplemental Table 2. miRZip-29b (System Biosciences) to stably knockdown miR-29b expression was used following manufacturer's instructions.

miRNA PCR arrays and Quantitative real-time PCR (qPCR)

Total RNA was isolated from cells using the miRNeasy Mini Kit (Qiagen). For miRNA PCR arrays (SABiosciences), MDA231 cells \pm GATA3 were screened according to manufacturer's instructions. cDNA was synthesized using RT² miRNA First Strand Kit (SABiosciences). Data were analyzed using SABiosciences software. For quantitative PCR (qPCR), cDNA was synthesized using the Superscript III RT First Strand Kit (Invitrogen). qPCR was performed using FastStart Universal SYBR Green master mix (Roche) in an Eppendorf Mastercycler realplex machine. Ct values were normalized to actin and GAPDH, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. For quantification of miRNA expression, Taqman probes were used according to

manufacturer's protocol (Applied Biosystems). Ct values were normalized to RNU48 (human samples) or snoRNA202 (mouse samples). Primer sequences for qPCR were found using the Harvard Primer Bank, and are detailed in Supplementary Table 2.

Serum cytokine analysis and ELISA

Sera from mice bearing 4T1 tumors were collected. Samples were allowed to clot, spun, and the serum fraction removed. Samples were analyzed in duplicate by Eve Technologies using a multiplex bead platform. n = 4–5 samples were obtained per group. For analysis of secreted VEGF from cells grown in vitro, supernatants were collected 24 or 48 hours after serum-starvation and analyzed using by ELISA according to manufacturer's instructions (R&D Systems).

Proliferation Assays

Cell proliferation was measured using the CellTiter MTT Assay according to manufacturer's instructions (Promega). Absorbance at 590 nm was read using a standard plate reader (Bio-Rad).

Luciferase Assays

For 3' UTR assays, 293T or MDA-MB-231 cells were co-transfected with miR-29b mimic or a control cel67 mimic (100 nM, Dharmacon) and the indicated siCheck2 dual luciferase reporter using Dharmafect Duo (Dharmacon). For TGF β and NF- κ B reporter assays, the SBE4-Luc, p3TP-Lux, or pGL-NF κ B-Luc reporter

was co-transfected with Renilla luciferase (pRL-TK) using Lipofectamine 2000 (Invitrogen) into MDA-MB-231 cells \pm GATA3. Cells were serum-starved overnight, and then stimulated with 5 ng/mL TGF β (R&D), 500 ng/mL sRANKL (Invitrogen), or 100 ng/mL TNF α (Peprotech) for 24 hours. For miR29-Luc reporter assays, the pGL3-miR29-Luc reporter, which contains the miR29a/b1 promoter, was co-transfected with pRL-TK into 293T or MDA-MB-231 \pm GATA3 or pcDNA-GATA3. Lysates were collected 24–48 hr post-transfection, and renilla and firefly luciferase activities were measured using the Dual-Luciferase Reporter System and a GloMax luminometer (Promega).

Immunostaining

Tissues were fixed in 4% PFA overnight, paraffin processed or embedded into OCT for frozen sections, and sectioned. Antigen retrieval was performed using citrate or proteinase K (Kouros-Mehr et al. 2006). The TSA Amplification Kit (Perkin Elmer) was used according to manufacturer instructions. Primary antibodies were incubated overnight, and secondary antibodies were incubated for 1 hour. The following antibodies were used: CD31 (BD Pharmingen), F4/80 (Invitrogen), phospho-histone H3 and cleaved caspase-3 (Cell Signaling), GATA3 (Abcam), biotinylated anti-rat (Jackson), and biotinylated anti-rabbit (Dako). Fluorescent secondary antibodies were from Molecular Probes (Invitrogen). Image analysis was performed using ImageJ software.

Western blotting

Cells were lysed in RIPA buffer plus protease inhibitors (Roche) or directly in Lamelli Buffer with DTT. Protein concentration was measured using the BCA Protein Assay Kit (Thermo Scientific). Lysates were subjected to SDS-PAGE, transferred to PVDF membranes, blocked in 5% BSA, incubated with primary antibody overnight and visualized using ECL Detection Reagents (Pierce). Exposures were acquired using a LAS-4000 Imager (Fuji). Primary antibodies to actin (Santa Cruz), LOX, (Novus), VEGF (Abcam), Angptl4 (Invitrogen), and GATA3 (Abcam) were used at manufacturer's recommended dilutions.

Flow Cytometry

To sort primary mouse mammary epithelial cells into basal and luminal fractions, mammary glands from adult virgin females were digested with collagenase. Organoids were collected by brief centrifugation and digested with trypsin to dissociate into single cells. The cells were stained with antibodies against CD49f, CD24 and lineage markers (CD45, CD31, Ter119) (eBioscience)(Stingl et al. 2006). Antibodies against CD29, GATA3, CD44, CD24, and EpCAM (all from eBioscience) were used to stain cell lines. For the intracellular GATA3 stain, cells were first permeabilized using 0.2% Triton X-100. Analysis and cell sorting was performed on a LSRII or FACS Aria II (Becton Dickinson), and analyzed using FlowJo software or FACSDiva software (BD Biosciences).

Statistical analysis

Statistical analysis was performed using Prism 4 software (GraphPad Software,

Inc.). All data are presented as mean \pm SEM, unless otherwise stated. When two groups were compared, the two-tailed Student's t-test was used, unless otherwise stated. When three or more groups were compared, the one-way analysis of variance (ANOVA) test was used, followed by Tukey's test to determine significance between groups. We considered $p < 0.05$ as significant.

SUPPLEMENTAL INFORMATION

Supplemental information includes six figures and two tables.

FIGURE LEGENDS

Figure 3.1: GATA3 suppresses spontaneous and experimental breast cancer metastases to the lungs.

(a) *GATA3* expression levels from Basal A, Basal B and Luminal breast cancer cell lines. Microarray dataset is adapted from Neve et al. (2006). ** ANOVA $p < 0.001$.

(b) Relative *Gata3* levels in 4T1 cells \pm Gata3 measured by quantitative PCR (qPCR).

(c–e) Balb/c mice were injected with 4T1 cells \pm Gata3 into the inguinal mammary fat pad. Tumors were allowed to grow for 3 weeks and measured (c), and lungs were examined for metastases (d) (n=8 per group, * $p < 0.01$). Representative H&E images depicting the lung metastases are shown (e).

(f–h) Number of phospho-histone H3 positive nuclei (f), CD31 mean intensity (g) and F4/80 mean intensity (h) in primary 4T1 \pm Gata3 tumors. Representative images are shown below graphs.

(i) MTT proliferation assay of 4T1 cells \pm Gata3 grown in vitro.

(j) Serum VEGF levels in mice bearing primary 4T1 \pm Gata3 tumors (n=4 per group, * $p < 0.05$.)

(k) Balb/c mice were injected i.v. with 4T1 cells \pm Gata3, and bioluminescence imaging was performed on day 14 prior to sacrifice (n=10 mice per group).

(l) Balb/c mice were co-injected 1:1 with control cells labeled in RFP (4T1-RFP) and Gata3-expressing cells labeled in GFP (4T1-Gata3-GFP). The percentage of RFP- and GFP-positive cells were determined by flow cytometry on day 12. (**

p<0.002, paired t test). Scale bars = 200 μ m (**e**), 100 μ m (**f–h**). All data are reported as mean \pm SEM.

Figure 3.2: GATA3 induces a more luminal phenotype and decreases cell migration.

(a) Phase-contrast images of MDA231 cells \pm GATA3 in 2D culture.

(b) Phase-contrast images of MDA231 cells \pm GATA3 embedded in 3D Matrigel. (See also Supplemental Movies 1 and 2).

(c) Relative expression of epithelial markers (*CDH1*, *KRT8*, *KRT18*), mesenchymal markers (*FN1*, *SNAIL*, *SLUG*, *ZEB1*, *ZEB2*, *VIM*, *HMGA2*, *RANK*) and major inflammatory and stem-associated pathway genes including NF- κ B (*STAT3*, *RANK*), Wnt (*FZD1*) Hedgehog (*Gli3*) and Notch (*JAG1*). (n=5 samples per group, p<0.05 for all genes.)

(d–f) Flow cytometry analysis of cell-surface markers, EpCAM and CD49f, in MDA231 cells \pm GATA3 (d) and quantification (e) of n=8 experiments, ** p<0.001. Histogram of CD49f expression is shown in (f).

(g–h) Number of tumor spheres from single MDA231 cells \pm GATA3 (* p<0.02) (g), and representative phase-contrast images (h). All scale bars = 100 μ m. All data are reported as mean \pm SEM.

Figure 3.3: miR-29b is induced by GATA3, enriched in luminal, good prognostic breast cancers, and associated with reduced metastatic potential.

(a–b) Eighty-eight miRNAs were screened using qPCR miRNA arrays in MDA231 ± GATA3 cells (a) and further validated (b).

(c) MDA231 cells were co-transfected with a miR29a/b1-Luc reporter and increasing amounts of GATA3. Firefly luciferase activity was normalized to Renilla luciferase activity, and plotted relative to the vector control.

(d–e) Relative TGFβ-Luc reporter activity (d) and NFκB-Luc reporter activity (e) in MDA231 ± GATA3 cells, with and without TGFβ1, sRANKL or TNFα stimulation.

(f–g) Relative miR29abc expression in primary human basal-like, HER2, luminal A and luminal B breast cancers (f), and in primary human estrogen-receptor (ER) positive and negative tumors (g). Microarray dataset is adapted from (Ref. 32).

(h) Relative miR29abc expression in primary mouse basal and luminal breast cancers. Microarray dataset is adapted from (Ref. 35). (* p<0.05).

(i) Relative miR-29b expression in 67NR, 168FARN, 4TO7 and 4T1 cells, with metastatic ability shown below. Microarray dataset adapted from (Ref. 37).

(j) Relative *Gata3* and miR-29b expression in normal, adenoma and carcinoma cells from MMTV-PyMT mice, inversely correlated with timing of metastasis.

Figure 3.4: miR-29b promotes luminal characteristics and loss of miR-29b induces a de-differentiated, mesenchymal phenotype.

(a) Relative expression of luminal (*Krt8*, *Esr1*, and *Gata3*) and basal (*Krt14*) epithelial markers in PyMT cells ± miR-29b by qPCR (n=5 per group, p<0.05 for all genes).

(b) Phase-contrast images of PyMT \pm miR-29b aggregates embedded in 3D Matrigel.

(c) Phase contrast images of MDA231 cells \pm Zip29 knockdown.

(d) Flow cytometer analysis of CD49f (integrin β 1) and CD29 (integrin α 6) surface protein expression in MDA231 cells \pm Zip29. Representative plot of n=5 experiments.

(e–f) Phase-contrast images (e) and flow cytometer analysis of cell-surface CD24 and CD44 expression (f) of HMLE cells \pm Zip29. Representative plot of n=5 experiments.

(g) Relative expression of epithelial and mesenchymal markers in HMLE cells \pm Zip29 by qPCR (n=5 per group, $p < 0.05$ for all genes). All scale bars = 100 μ m.

Figure 3.5: miR-29b targets pro-metastatic genes involved in remodeling the tumor microenvironment.

(a) Computationally-predicted interactions between miR-29 and the 3' UTRs of several mRNAs involved in angiogenesis, ECM cross-linking, ECM proteolysis, and ECM signaling.

(b) The 3'UTRs of the indicated miR-29b targets were cloned into a dual luciferase reporter and co-transfected with miR-29b or control mimic into 293T cells. Luciferase activity was measured 48 hours post-transfection and normalized to firefly luciferase. Data are representative of n=4 experiments performed in triplicate, $p < 0.05$ for all 3' UTR reporters.

(c) Relative miR-29b expression in stably transduced 4T1 cells \pm miR-29b.

(d) Relative mRNA expression of indicated miR-29b targets in 4T1 cells \pm miR-29b.

(e) Western blot analysis of three miR29 targets (ANGPTL4, LOX and VEGFA) in 4T1 cells \pm miR-29. All data are reported as mean \pm SEM.

Figure 3.6: miR-29b expression inhibits lung metastases and loss of miR-29b increases lung metastases.

(a) Balb/c mice were injected with 4T1 cells \pm miR-29b into the mammary fat pad. Tumors were allowed to grow for 3 weeks and measured (n=8 mice per group).

(b–c) 4T1 \pm miR-29b tumors were stained for CD31 to evaluate tumor vasculature (b) or picrosirius red to evaluate fibrillar collagen (c).

(d) Representative H&E images depicting the size of the lung metastases.

(e) Balb/c mice were injected with 4T1 cells \pm miR-29b and bioluminescence imaging was performed on day 14 prior to sacrifice (n=10 mice per group). Representative H&E images of the lung metastases are shown below.

(f) Relative expression of *Lox* and *Vegfa*, two miR29 targets, in normal MECs and MMTV-PyMT adenoma and carcinoma cells (n=4–6 samples per group).

(g) Relative miR-29b expression in stably transduced PyMT cells \pm miR-29b.

(h–i) FVB/n mice were injected i.v. with PyMT cells \pm miR-29b, and sacrificed at 6 weeks. A representative set of gross lungs (H) and H&E images of the lung metastases (I) are shown (n=8 mice per group).

(j–k) FVB/n mice were injected i.v. with PyMT cells \pm Zip29 to knockdown endogenous miR-29b, and bioluminescent imaging was performed on week 3 (j). Graph depicts number of surface lung metastases per lung lobe in each group

(k) (n=8 mice per group). Scale bars = 100 μm (b–d), 200 μm (e), 1 cm (h), and 1 mm (i). All data are reported as mean \pm SEM.

Figure 3.7: miR-29b knockdown increases target gene expression and miR-29b suppresses metastasis by repressing four microenvironmental targets.

(a) Relative expression of miR29 targets in HMLE cells \pm Zip29.

(b) The 3'UTR reporters of the indicated miR29 targets were co-transfected with anti-miR-29b or control inhibitor into 293T cells. Data are representative of n=3 experiments performed in triplicate.

(c–d) Balb/c mice were injected i.v. with 4T1 cells \pm miR-29b re-expressing ANGPTL4, LOX, Mmp9, or VegfA, and sacrificed at 2 weeks. Representative fluorescence images of lungs were taken on a dissection microscope (c) and fluorescence was quantified on ImageJ (d). Scale bar = 1 mm.

Figure 3.8: miR-29b is an important downstream target of GATA3 that mediates its ability to promote luminal differentiation and suppress metastasis.

(a) Phase contrast images of MDA231 cells \pm GATA3 \pm Zip29. Scale bar = 100 μm .

(b) Relative expression of miR29 targets *ANGPTL4*, *COL1A1*, *LOX*, *MMP9*, *PDGFC* and *VEGFA* in MDA231 cells \pm GATA3.

(c) The 3'UTR reporters of the indicated miR29 targets were transfected in MDA231 \pm GATA3 cells, and luciferase activity measured.

(d) Bioluminescent imaging of mice injected i.v. with MDA231 cells \pm GATA3 \pm Zip29.

(e) Working model of how GATA3 promotes differentiation and suppresses metastasis through regulating miR-29b.

Supplemental Figure S3.1: GATA3 suppresses primary tumor growth and experimental lung metastases in human MDA231 breast cancer cells.

(a) Immunofluorescence staining for GATA3 (green) and E-Cadherin (red) in the normal mouse mammary gland shows GATA3 staining only in the luminal epithelial cells.

(b) Relative expression of GATA3 in MDA231 cells \pm GATA3.

(c) Tumor sizes of nude mice bearing orthotopic MDA231 \pm GATA3 tumors (n=5 mice per group).

(d) MDA231 \pm GATA3 tumors stained for GATA3 to validate over-expression in vivo.

(e–f) Mean intensity of CD31 (e) and representative images (f) of primary MDA231 \pm GATA3 tumors to evaluate tumor vasculature.

(g–h) Mean intensity of F4/80 (g) and representative images (h) of primary MDA231 \pm GATA3 tumors to evaluate macrophage infiltration.

(i) Secreted VEGFA from MDA231 cells \pm GATA3 grown in vitro, analyzed by ELISA.

(j) MTT assay to measure proliferation of MDA231 cells \pm GATA3 grown in vitro.

(k) Bioluminescence imaging of nude mice injected i.v. with MDA231 cells \pm GATA3. Mice were imaged at 3 weeks post injection.

Supplemental Figure S3.2: GATA3 affects cell migration in 2D and 3D culture, and promotes a luminal differentiation program.

(a) Phase-contrast images of MDA231 \pm GATA3 cells subject to a wound scratch assay. Cells were scratched at $t = 0$, and images taken at $t = 24$ hours.

(b) Human basal breast cancer Hs578T cells \pm GATA3 were embedded into Matrigel, and grown in serum-free conditions plus FGF2 for 3 days. Scale bar = 100 μm .

(c–d) Relative GATA3 expression in Hs578T cells \pm GATA3 measured by qPCR (c) or by intracellular flow cytometry staining (d).

(e) Relative expression of epithelial and mesenchymal markers in Hs578T cells \pm GATA3 measured by qPCR. $p < 0.05$ for all genes analyzed.

Supplemental Figure S3.3: GATA3 induces miR-29b expression, inhibits TGF β -induced EMT, and transcriptionally dampens TGF β signaling components.

(a) GATA3 increases miR-29b expression in Hs578T cells \pm GATA3.

(b) The miR29a/b1 promoter contains direct binding sites for GATA3 [(A/T)GATA(A/G)].

(c) Phase-contrast images of NMuMG mammary cells \pm Gata3 stimulated with 2 ng/mL TGF β to induce EMT for 24 hours. Scale bar = 100 μm .

(d) Relative expression of epithelial and mesenchymal markers in NMuMG cells \pm Gata3 were measured by qPCR, with and without TGF β stimulation.

(e) Relative expression of TGF β and BMP genes in MDA231 cells \pm GATA3. All data are reported as mean \pm SEM, $p < 0.05$ for all genes analyzed.

(f) Relative expression of miR29a and miR-29b in MDA231 cells stimulated with TGF β 1, sRANKL or TNF α .

(g) Sequences of the mature forms of miR29a, miR-29b, and miR29c. Human miR29a-29b1 is located on chromosome 7, and miR-29b2-miR29c is located on chromosome 7.

Supplemental Figure S3.4: miR-29b is enriched in luminal epithelial cells, and its loss results in de-differentiation and an EMT-like phenotype.

(a) Normal mouse luminal and basal MECs were fractionated by FACS using previously described markers: CD49f, CD24 and Lineage markers (CD31, CD45 and Ter119).

(b) The levels of luminal (*Gata3*, *ER*, *CK8*, *CK18*) and basal markers (*CK14*, *p63*) were measured in the two cell populations by qPCR to validate successfully fractionation of the two populations. $p < 0.05$ for all genes.

(c) Relative miR29a, miR-29b and miR29c expression in normal mouse luminal and basal mammary epithelial cells (MECs).

(d) Relative miR29abc expression levels in metastatic-competent 168FARN, 4TO7, and 4T1 cells. Microarray dataset adapted from Dykxhoorn et al. (2009). All data are reported as mean \pm SEM.

(e) Relative expression of miR29a, miR-29b and miR29c in MDA231 cells \pm Zip29 to confirm knockdown. The miRZip29b virus inhibits both miR29a and miR-29b, and to a lesser extent, miR29c expression.

(f) Phase-contrast images of mouse 4TO7 cells \pm Zip29 knockdown, which shows elongated and spindle-like cells (arrowheads). Scale bar = 100 μ m.

(g) Relative expression of mesenchymal markers in 4TO7 cells \pm Zip29. Mesenchymal markers Cdh2 (N-cadherin) and vimentin are significantly increased with miR29 knockdown.

Supplemental Figure S3.5: miR29 suppresses pro-metastatic regulators of the tumor microenvironment, including *ANGPTL4*, *LOX*, *MMP9*, *VEGFA*, and inhibits lung metastases in human and mouse breast cancer cells.

(a) Relative miR-29b expression in stably transduced MDA231 cells \pm miR-29b.

(b) Relative expression of indicated miR29 target genes in MDA231 cells \pm miR-29b.

(c–d) Nude mice were injected i.v. with MDA231 cells \pm miR-29b labeled with GFP and sacrificed 7 weeks post-injection. The number of lung metastases per mouse were counted and graphed (c). Representative fluorescence images of lung metastases (left) and H&E staining (right) are shown (d).

(e) Nude mice were injected with MDA231 cells \pm miR-29b to form primary tumors in the mammary fat pad, and sacrificed 8 weeks post-injection. Representative H&E images of lung metastases are shown. Tumor sizes were not significantly different (data not shown).

(f–g) *Vegf* and *Lox* expression were measured in PyMT cells ± miR-29b by qPCR (g) and protein levels were measured by Western blot (h). All data are reported as mean ± SEM. Scale bars = 100 μm.

(h) Bioluminescent imaging of nude mice injected i.v. with MDA231 cells ± Zip29 knockdown, 6 weeks post injection.

(i) Bioluminescent imaging of Balb/c mice injected i.v. with 4T1 cells ± Zip29 knockdown, 2 weeks post injection.

Supplemental Figure S3.6: GATA3 promotes differentiation and suppresses metastasis by regulating miR-29b expression.

(a) Phase contrast images of human luminal T47D cells ± Zip29.

(b) Relative expression of mesenchymal markers (*VIM* and *TWIST*) and luminal markers (*GATA3* and *KRT8*) in T47D cells ± Zip29. All data are reported as mean ± SEM.

(c–d) Balb/c mice were injected i.v. with 4T1 cells ± GATA3 ± Zip29 knockdown, and sacrificed 2 weeks post injection. The number of lung surface metastases were counted and graphed (c), and representative H&E images of lung metastases are shown (d). ANOVA ** $p < 0.01$, $n=8$ mice per group.

Supplemental Movies 1 and 2: GATA3 suppresses cell invasion in 3D Matrigel.

MDA231-Control (**Supplemental Movie 1**) or MDA231-GATA3 (**Supplemental Movie 2**) cells were embedded into Matrigel, grown in serum-free conditions with

FGF2, and imaged using an inverted brightfield microscope every 20 minutes for 48 hours. Images were assembled and played at 8 frames/sec.

FIGURES

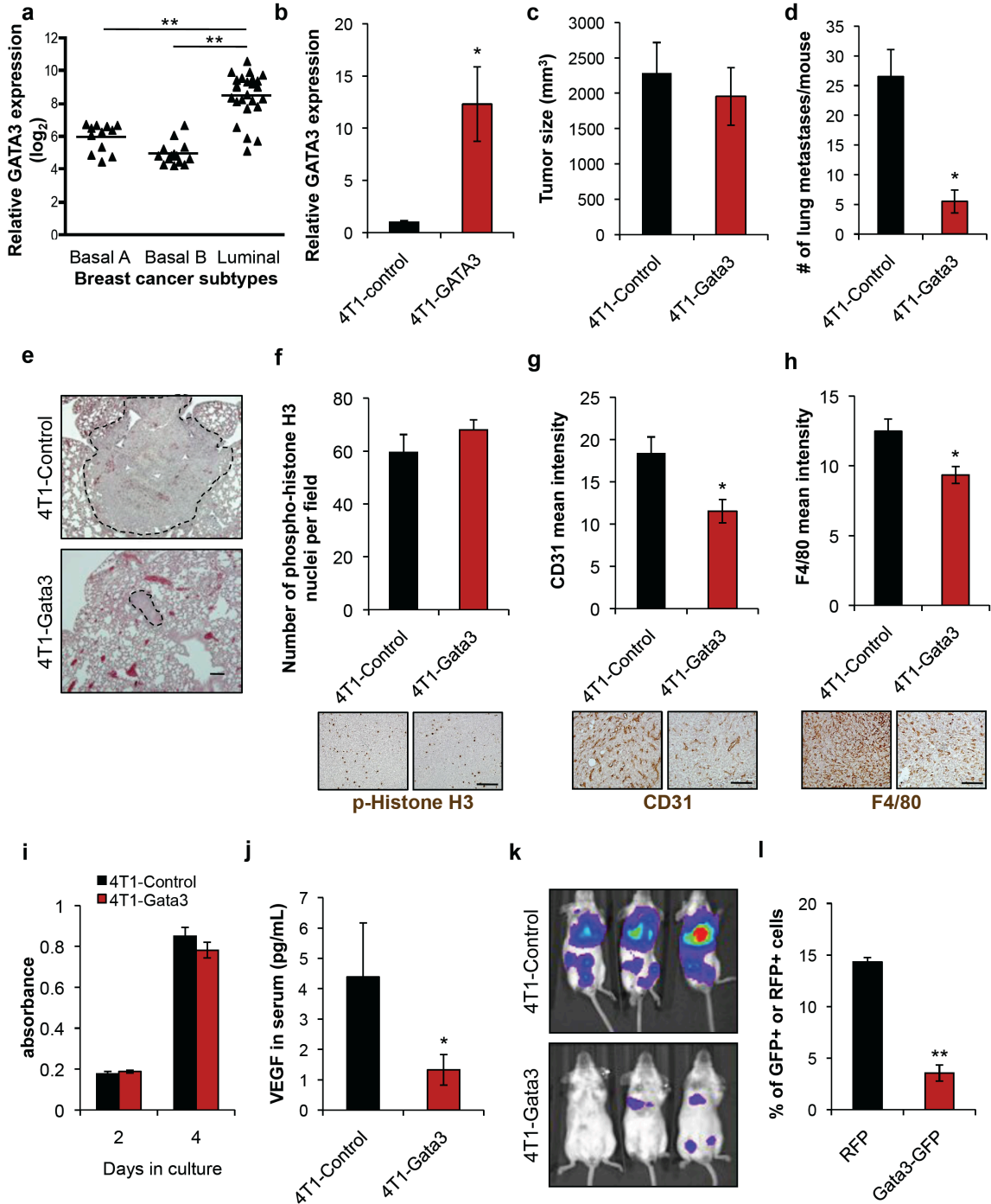


Figure 3.1

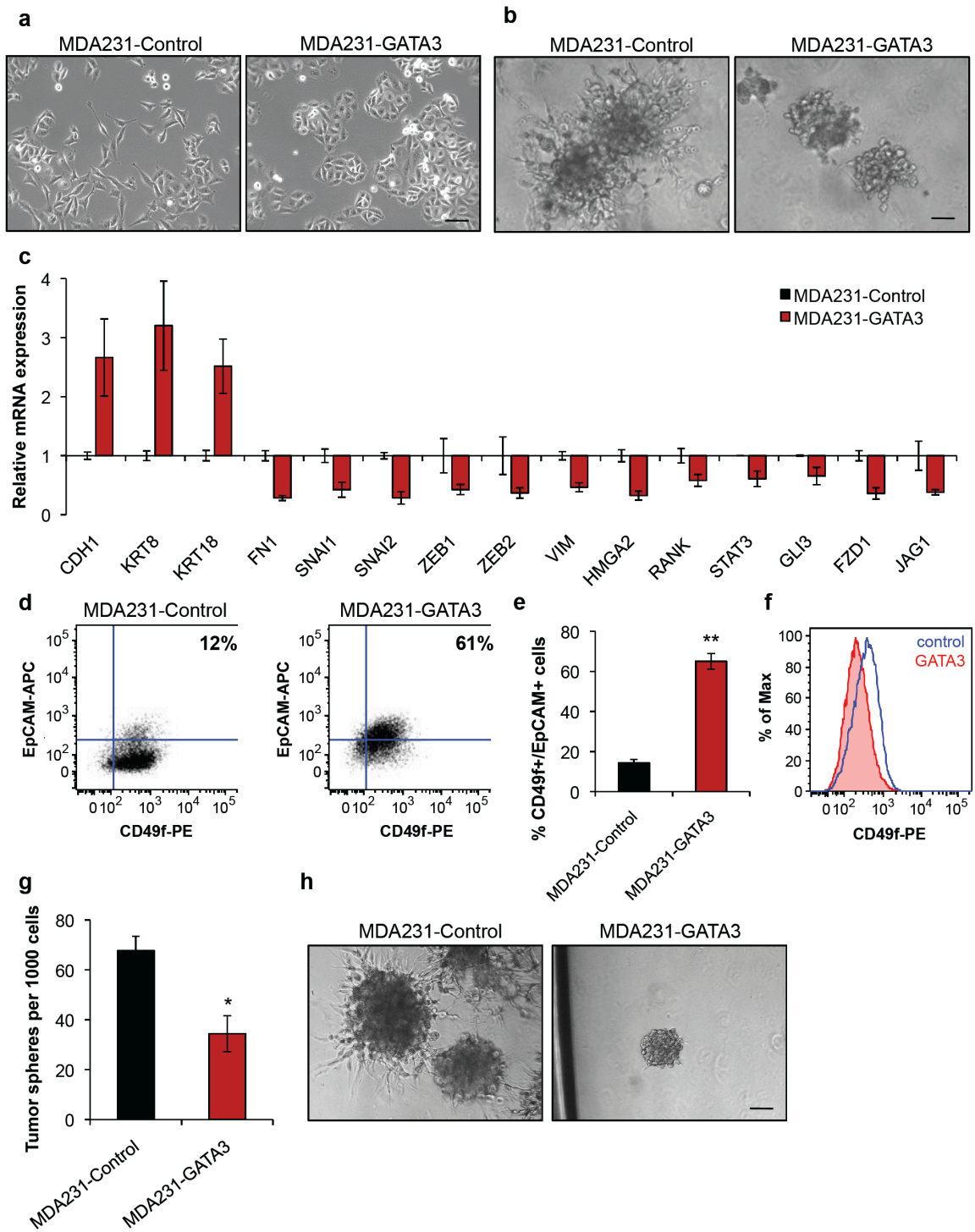


Figure 3.2

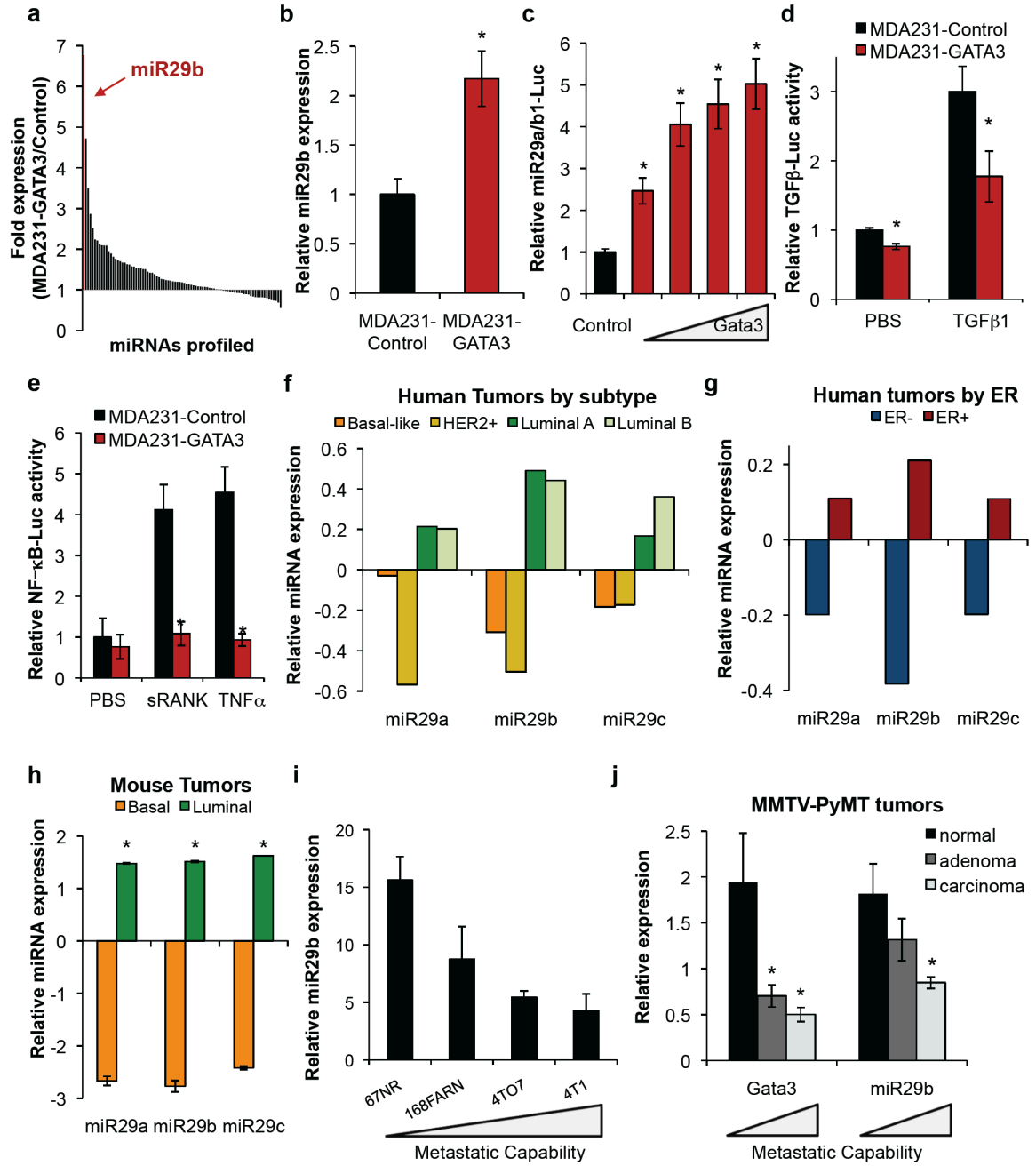


Figure 3.3

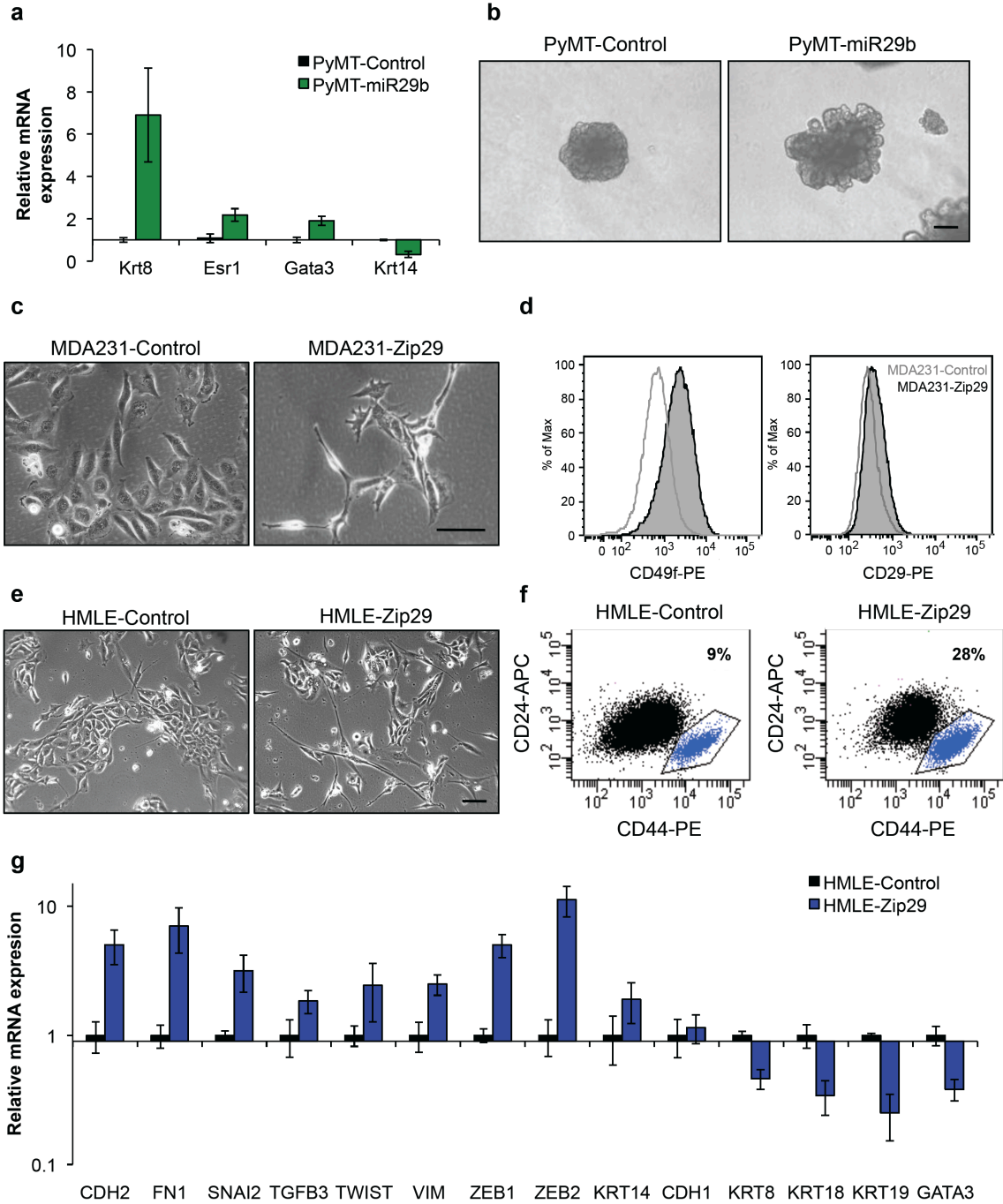


Figure 3.4

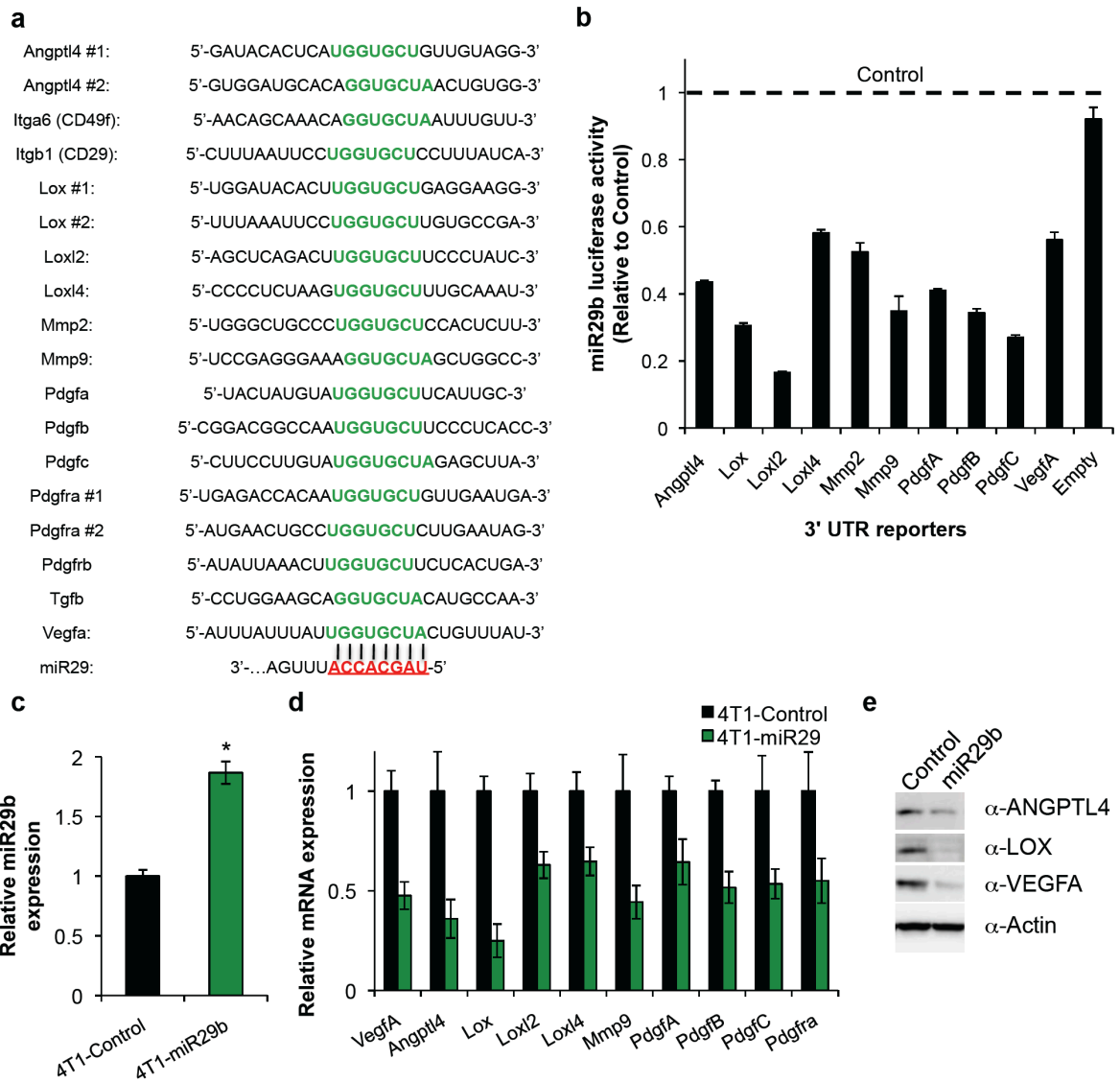


Figure 3.5

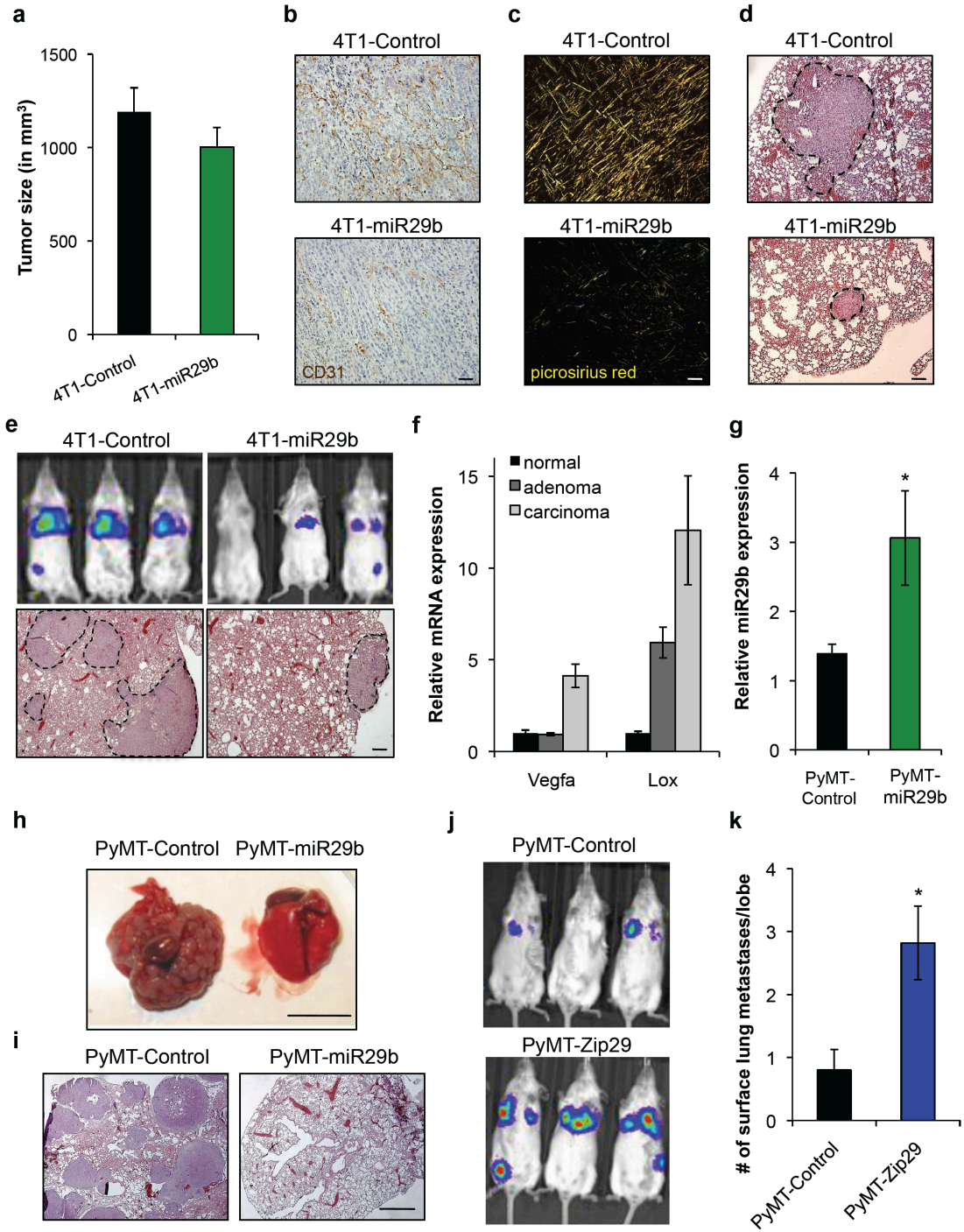


Figure 3.6

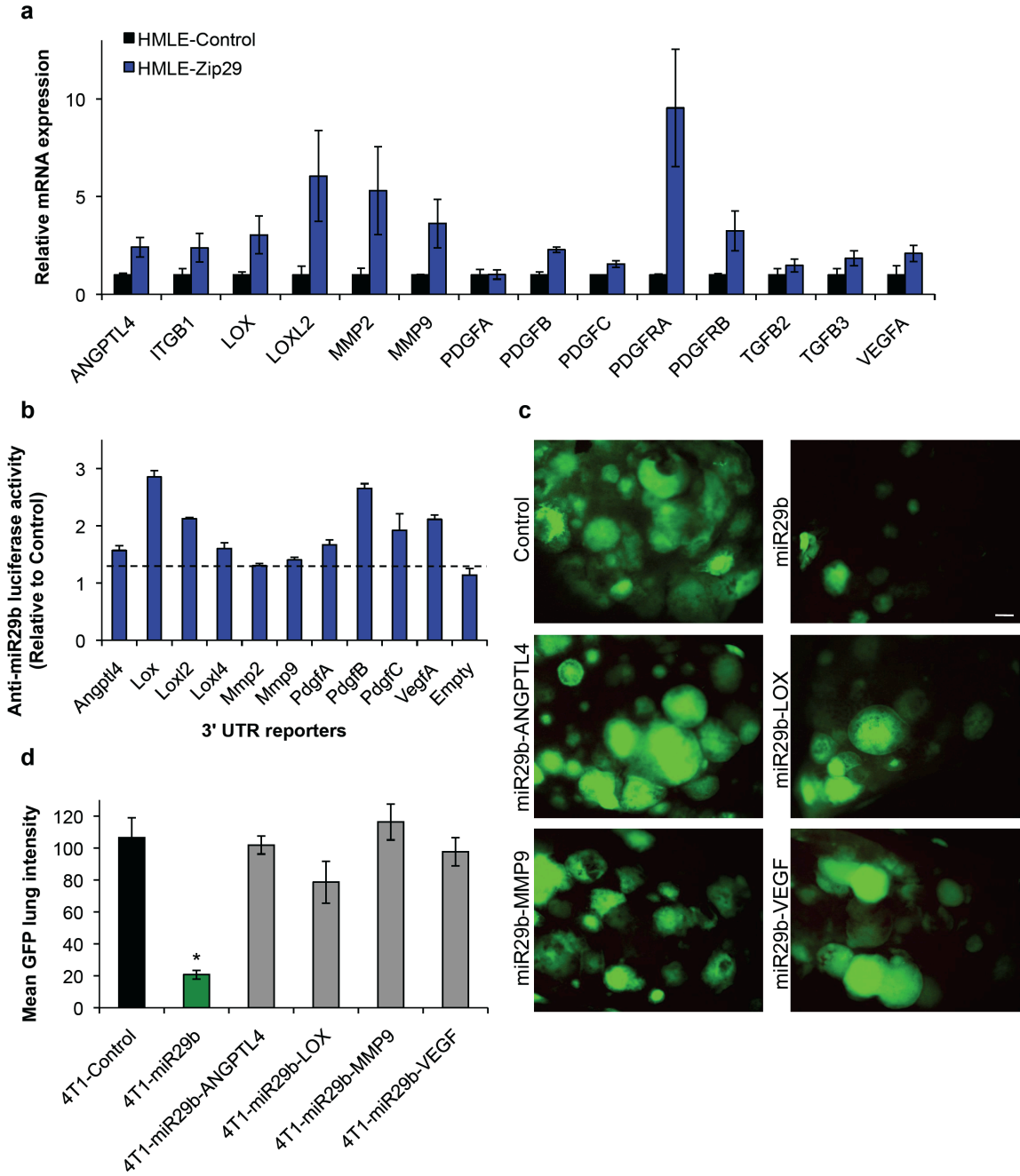


Figure 3.7

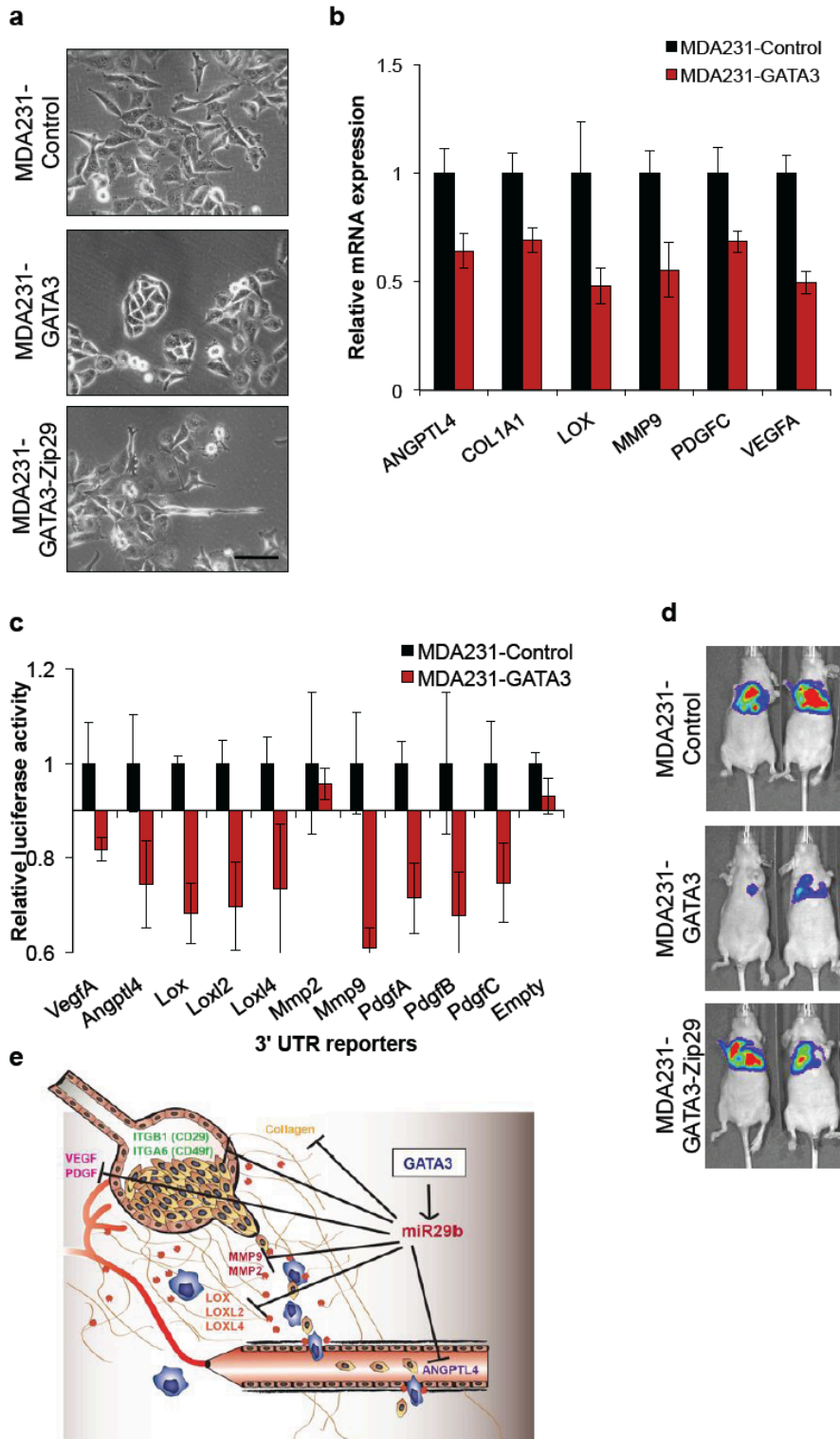
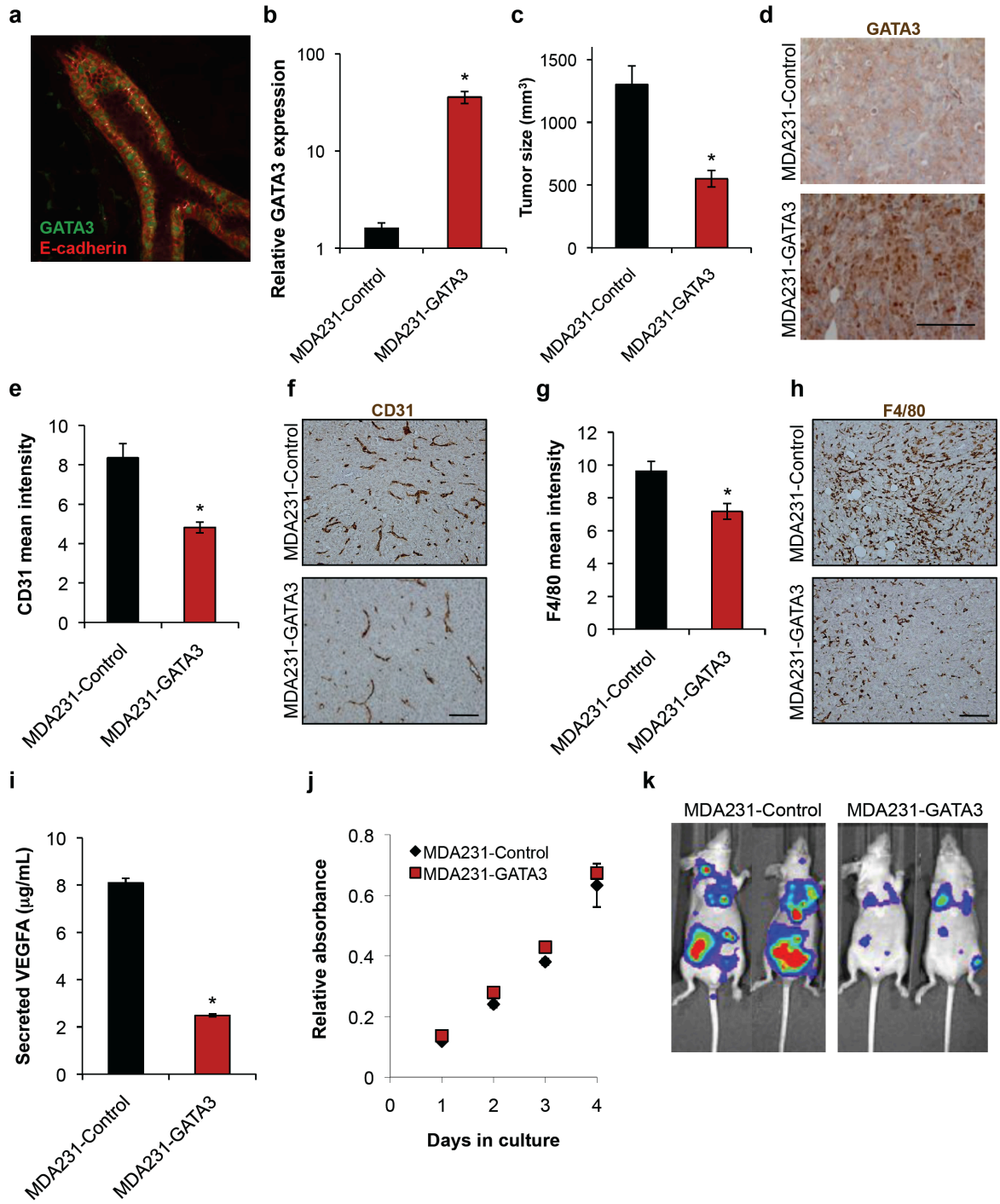
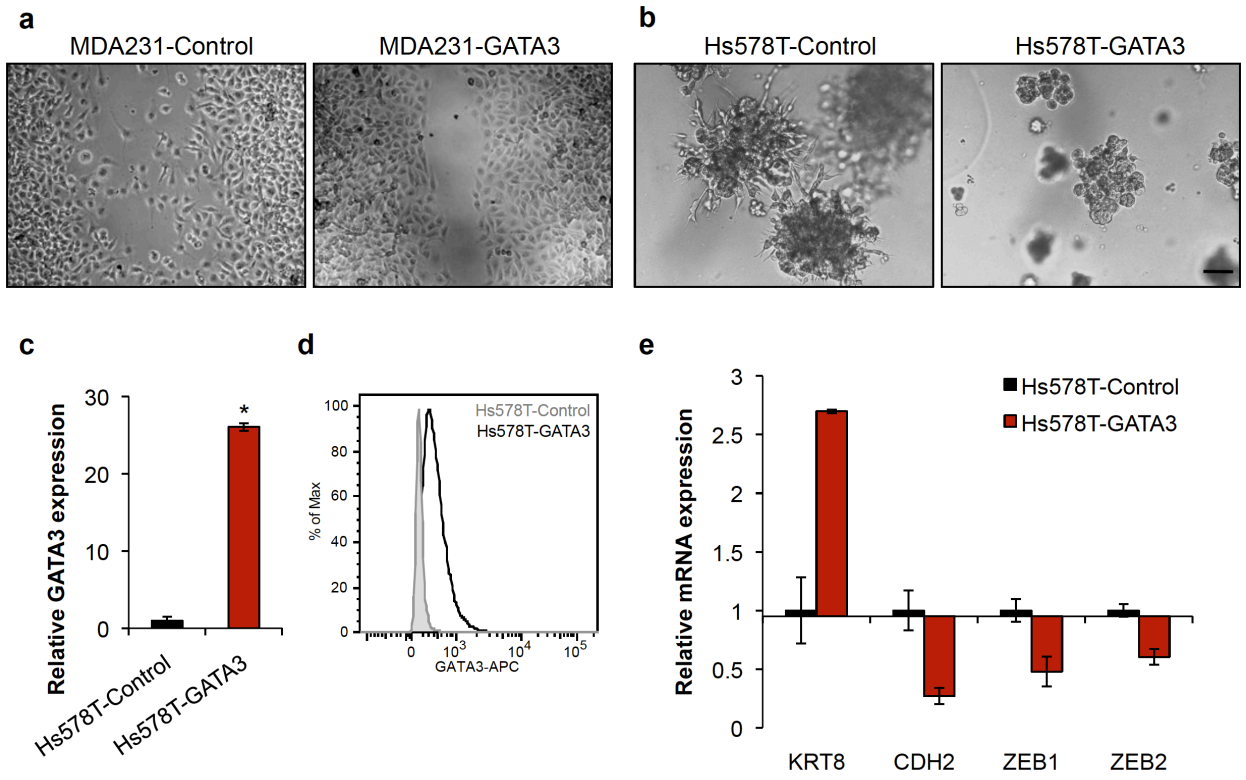


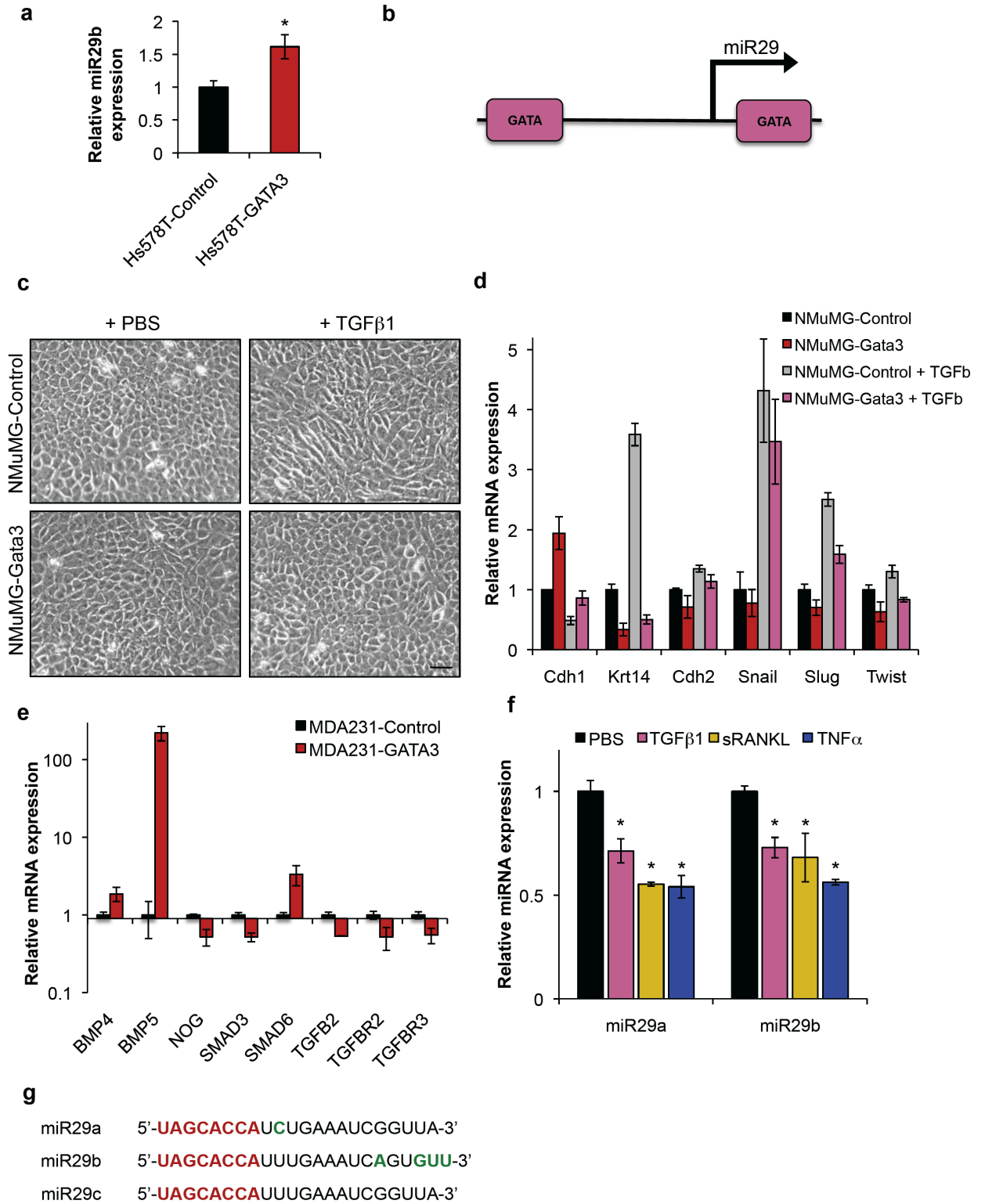
Figure 3.8



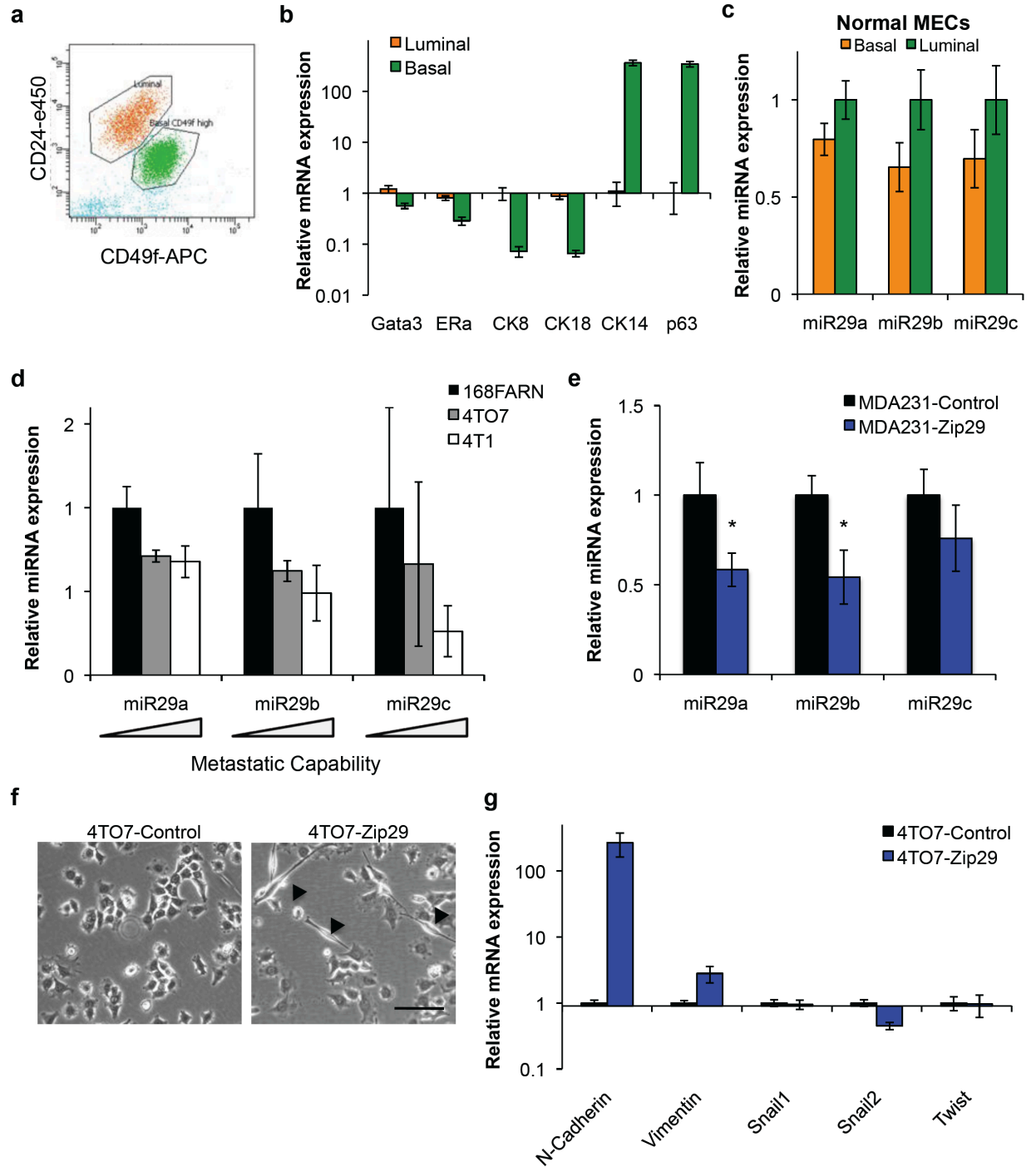
Supplemental Figure S3.1



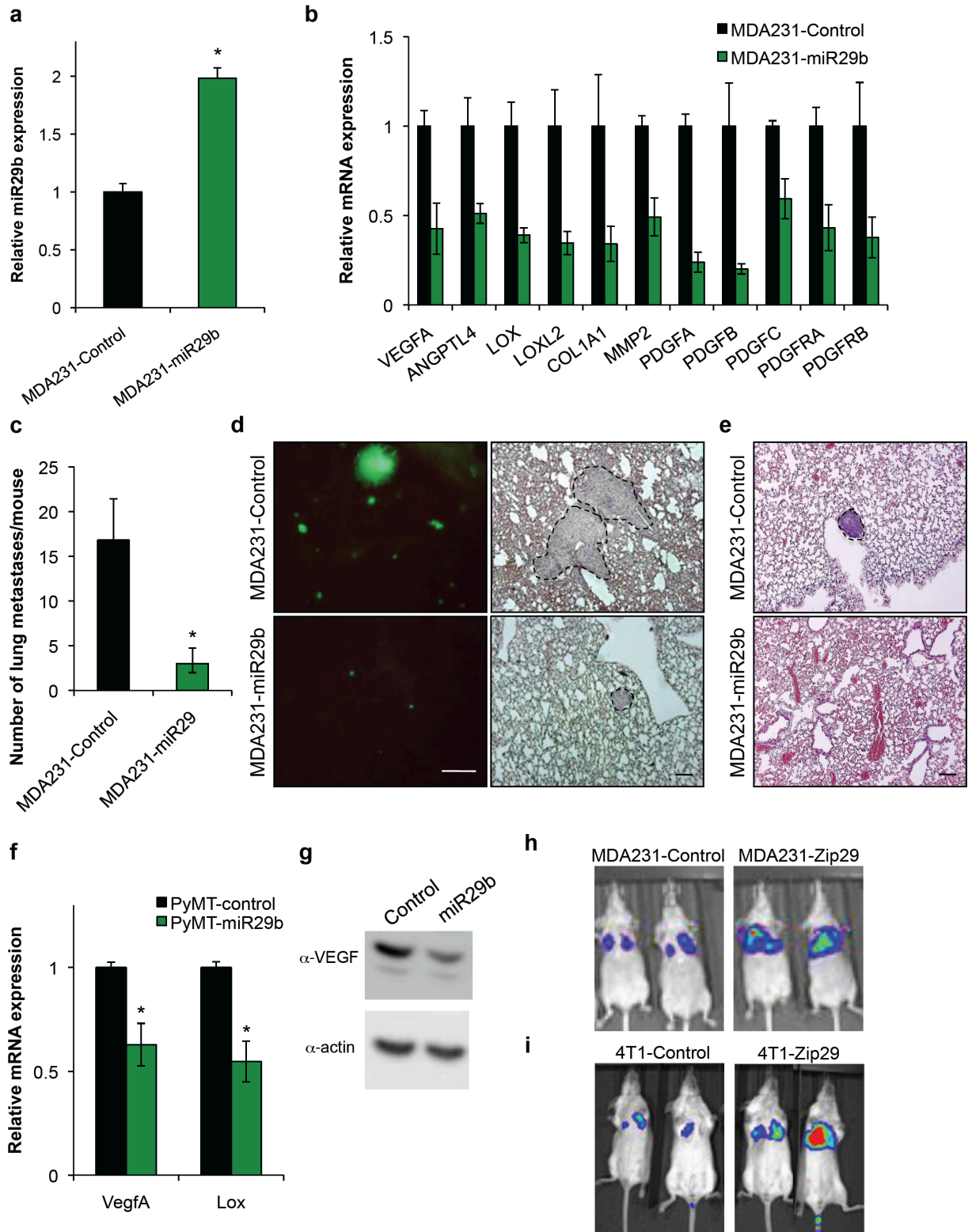
Supplemental Figure S3.2



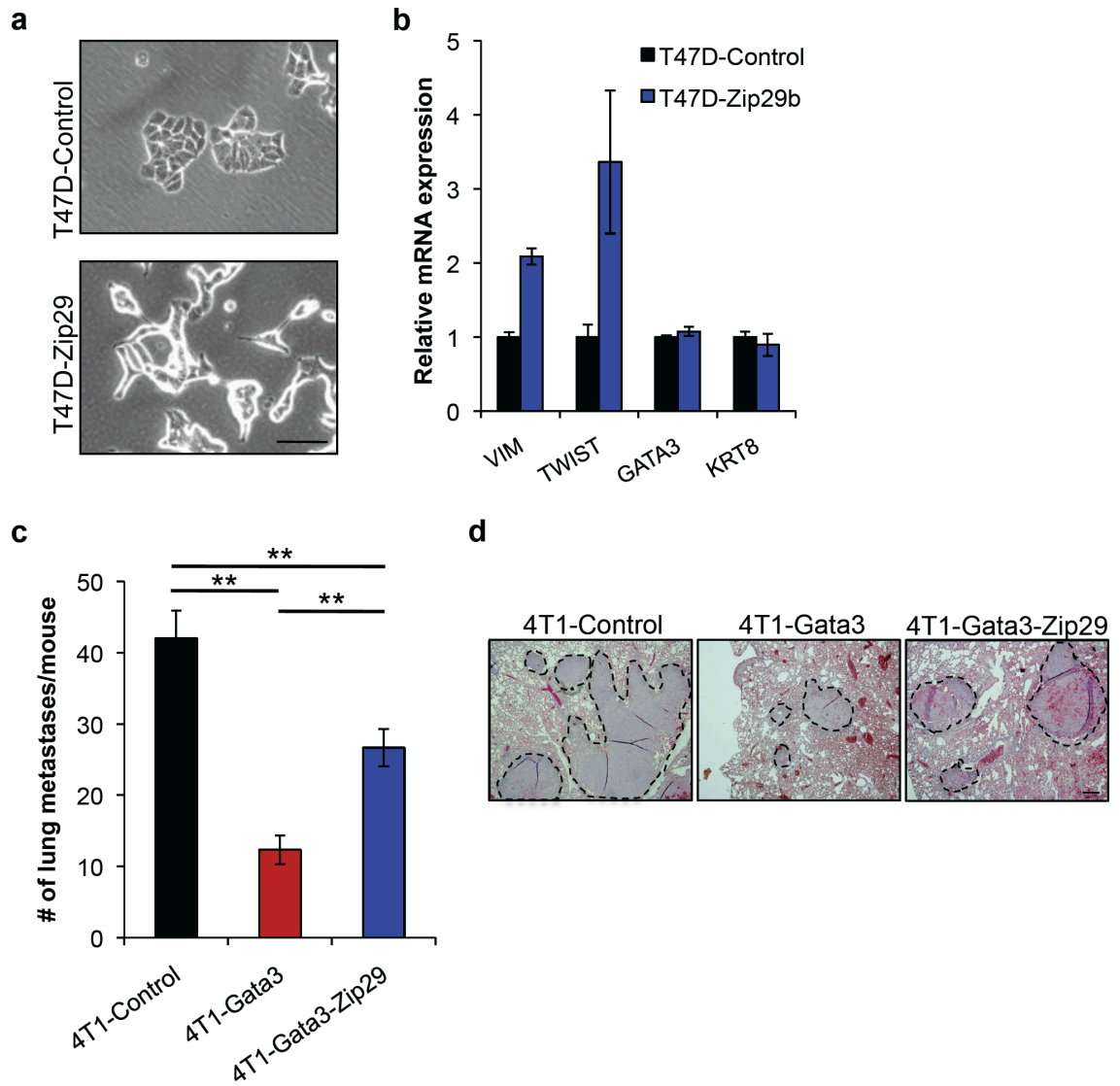
Supplemental Figure S3.3



Supplemental Figure S3.4



Supplemental Figure S3.5



Supplemental Figure S3.6

Supplemental Table 1: miRNA qPCR Screen results
(miRNAs with ΔC_t values greater than 9 were negligibly expressed, and therefore excluded from further analysis.)

miRNA ID	Well	AVG ΔC_t (Ct(GOI) - Ave Ct (HKG))		Fold Difference
		MDA231- GATA3	MDA231- Control	GATA3/Control
hsa-miR-32	H03	8.84	12.48	12.43
hsa-miR-29b	D03	5.78	8.54	6.77
hsa-miR-301a	D11	5.46	7.70	4.72
hsa-miR-19a	G05	1.70	3.50	3.49
hsa-miR-34a	F11	12.80	14.32	2.87
hsa-miR-130a	B09	4.26	5.59	2.51
hsa-miR-15a	B03	4.95	6.11	2.24
hsa-miR-34c-5p	C05	11.07	12.22	2.21
hsa-miR-214	B02	13.55	14.64	2.13
hsa-miR-96	A09	5.81	6.88	2.10
hsa-miR-133b	A02	14.82	15.89	2.09
hsa-miR-193a-5p	G06	9.75	10.82	2.09
hsa-miR-206	D07	13.57	14.53	1.95
hsa-miR-124	D08	10.40	11.32	1.89
hsa-miR-181a	B08	6.39	7.24	1.81
hsa-miR-212	D05	9.88	10.70	1.76
hsa-miR-373	C11	12.84	13.63	1.73
hsa-miR-181c	H04	7.19	7.95	1.70
hsa-miR-335	A05	10.50	11.24	1.67
hsa-miR-372	G11	10.27	11.00	1.67
hsa-miR-98	F10	7.16	7.86	1.63
hsa-miR-199a-3p	B10	11.67	12.37	1.62
hsa-let-7g	C09	2.89	3.55	1.58
hsa-miR-122	A03	13.92	14.57	1.57
hsa-miR-196a	A06	9.11	9.74	1.54
hsa-miR-210	F08	7.83	8.45	1.53
hsa-miR-27a	G10	0.71	1.32	1.52
hsa-miR-126	G09	4.75	5.34	1.51
hsa-miR-200c	D12	8.95	9.54	1.51
hsa-miR-150	E06	15.08	15.61	1.45
hsa-let-7b	B06	5.18	5.68	1.42
hsa-miR-146b-5p	C01	10.25	10.76	1.42
hsa-miR-140-5p	B11	6.40	6.87	1.39
hsa-miR-222	A10	2.08	2.50	1.33
hsa-miR-149	G12	10.17	10.56	1.31
hsa-miR-27b	E08	3.65	3.99	1.26
hsa-miR-181d	D10	7.97	8.30	1.26
hsa-miR-146a	D04	10.66	10.97	1.24
hsa-miR-215	G04	13.92	14.22	1.23
hsa-miR-181b	F05	4.23	4.52	1.23
hsa-miR-23b	H01	3.93	4.20	1.21
hsa-miR-18a	G07	2.80	3.06	1.20
hsa-miR-218	D02	7.72	7.98	1.20
hsa-miR-17	F09	1.04	1.30	1.19

miRNA ID	Well	AVG ΔC_t (Ct(GOI) - Ave Ct (HKG))		Fold Difference
		MDA231- GATA3	MDA231- Control	GATA3/Control
hsa-miR-20b	A04	3.75	3.99	1.18
hsa-miR-378	B05	4.90	5.12	1.17
hsa-let-7d	F01	5.76	5.95	1.15
hsa-miR-20a	B12	0.99	1.18	1.14
hsa-miR-7	E09	3.30	3.46	1.12
hsa-let-7c	C12	7.29	7.44	1.11
hsa-miR-30c	C06	1.82	1.96	1.10
hsa-miR-21	D09	-2.00	-1.89	1.08
hsa-miR-191	E12	2.74	2.85	1.08
hsa-miR-135b	D06	7.54	7.64	1.07
hsa-miR-183	C04	7.45	7.55	1.07
hsa-miR-29a	E11	1.09	1.18	1.07
hsa-miR-143	G03	13.79	13.85	1.05
hsa-let-7i	E07	3.97	4.02	1.04
hsa-let-7f	F03	10.04	10.08	1.03
hsa-miR-148b	A11	5.23	5.25	1.01
hsa-miR-155	E03	16.54	16.56	1.01
hsa-miR-205	B07	14.98	14.99	1.01
hsa-miR-138	C10	3.28	3.26	0.98
hsa-miR-203	H02	16.59	16.56	0.98
hsa-miR-125a-5p	A07	2.67	2.63	0.97
hsa-miR-184	B01	15.17	15.11	0.96
hsa-miR-1	E04	15.88	15.81	0.95
hsa-miR-127-5p	E10	12.84	12.75	0.93
hsa-miR-25	F12	1.93	1.82	0.93
hsa-miR-193b	C03	6.47	6.36	0.93
hsa-let-7e	D01	2.79	2.67	0.92
hsa-miR-92a	A12	1.45	1.30	0.90
hsa-miR-15b	F06	2.39	2.23	0.89
hsa-let-7a	A01	1.69	1.51	0.89
hsa-miR-9	F02	9.84	9.61	0.85
hsa-miR-132	C02	7.28	7.02	0.84
hsa-miR-148a	C07	4.81	4.52	0.82
hsa-miR-142-5p	A08	16.84	16.56	0.82
hsa-miR-363	E05	16.84	16.56	0.82
hsa-miR-144	G01	16.84	16.56	0.82
hsa-miR-16	F07	0.11	-0.19	0.81
hsa-miR-100	E01	-0.56	-0.88	0.80
hsa-miR-10b	E02	7.47	7.09	0.77
hsa-miR-10a	F04	5.77	5.35	0.75
hsa-miR-125b	G08	-0.59	-1.03	0.74
hsa-miR-128a	G02	5.15	4.70	0.73
hsa-miR-18b	B04	6.92	6.38	0.69
hsa-miR-134	C08	12.49	11.64	0.55

Supplemental Table 2: Primer sequences used for real-time qPCR and for generating 3' UTR fragments from indicated genes.

qPCR Primers			
mAngptl4 qF1	CATCCTGGGACGAGATGAACT	hKrt18 qR1	GGATGTCGTTCTCCACAGACT
mAngptl4 qR1	TGACAAGCGTTACCACAGGC	mKrt18 qF1	AAGGTGAAGCTTGAGGCAGA
hAngptl4 qF1	GGCTCAGTGGACTTCAACCG	mKrt18 qR1	CTGCACAGTTTGCATGGAGT
hAngptl4 qR1	CCGTGATGCTATGCACCTTCT	hLOXqF1	CAGGGTGCTGCTCAGATTTCC
hBMP4qF1	AAAGTCGCCGAGATTCAGGG	hLOXqR1	GGTAATGTTGATGACAACTGTGC
hBMPqR1	GACGGCACTCTTGCTAGGC	mLox qF1	TCTTCTGCTGCGTGACAACC
hBMP5qF1	AAGACTACGGAACCACGAAAGA	mLox qR1	GAGAAACCAGCTTGGAAACCAG
hBMP5qR1	TCTGTGAGGCAAACCCAAGATA	mLox qF2	GGTTACTTCCAGTACGGTCTCC
hBMP7qF1	ATCGAGAGTTCGGTTTGTACT	mLox qR2	GCAGCGCATCTCAGGTTGT
hBMP7qR1	GTCTCATTGTCGAAGCGTTCC	hLOXL2 qF1	GCCGGGTGGAGGTGTACTAT
hCol1A1qF3	GTTCGAGGGCCAAGACGAAG	hLOXL2 qR1	TGGAGATTGTCTAACCCAGATGGG
hCol1A1qR3	CAGATCACGTCATCGACAAC	mLoxl2 qF1	ATTAACCCCAACTATGAAGTGCC
hEcad qF1	CCCACCACGTACAAGGGTC	mLoxl2 qR1	CTGTCTCCTCACTGAAGGCTC
hEcad qR1	CTGGGGTATTGGGGGCATC	hLoxl4 qF1	CTGGGCACCACTAAGCTCC
mEcad qF1	CCAACAGGGACAAAGAAACAAAGG	hLoxl4 qR1	CTCCTGGATAGCAAAGTTGTCT
mEcad qR1	GATGACACGGCATGAGAATAGAGG	mLoxl4 qF4	GCTGTGGAAGTACGGTACGAC
mESR1 qF1	AATTCTGACAATCGACGCCAG	mLoxl4 qR4	CCAGACTTTTCTGTAGTAGTGGC
mESR1 qR1	GTGCTTCAACATTCTCCCTCCTC	mMMP2 qF1	TTTGCTCGGGCCTTAAAGTAT
hFzd1 qF1	ATCTTCTTGTCCGGCTGTTACA	mMMP2 qR1	CCATCAAACGGGTATCCATCTC
hFzd1 qR1	GTCTCGGCGAACTTGTCTATT	hMMP9 qF1	AGACGGGTATCCCTTCGACG
hFN1 qF1	AGGAAGCCGAGGTTTTAACTG	hMMP9 qR1	AAACCCGAGTTGGAACCACGAC
hFN1 qR1	AGGACGCTCATAAGTGTCAACC	mMMP9 qF1	GGACCCGAAGCGGACATTG
hGATA3 qF1	GCGGGCTCTATCACAATAATGA	mMMP9 qR1	CGTCGTCGAAATGGGCATCT
hGATA3 qR1	GCCTTCGCTTGGGCTTAAT	mMMP9 qF2	TGCCCATTTTCGACGACGAC
mGATA3 qF2	GAACCGGCCCTTATCAAG	mMMP9 qR2	GTGCAGGCCGAATAGGAGC
mGATA3 qR2	ACAGTTCGCGCAGGATGTC	hNoggin qF1	CCATGCCGAGCGAGATCAAA
hGAPDH qF1	CGACAGTCAGCCGCATCTT	hNoggin qR1	TCGGAAATGATGGGGTACTGG
hGAPDH qR1	CCGTTGACTCCGACCTTCA	mp63deltN qF1	CAATGCCAGACTCAATTTAGTGA
mGAPDH qF1	AGGTCGGTGTGAACGGATTTG	mp63deltN qR1	GGCCCCGGTAATCTGTGTTGG
mGAPDH qR1	TGTAGACCATGTAGTTGAGGTCA	hPDGFC qF1	ATTCACAGCCCAAGGTTTCT
hGli3 qF1	TGTTACATGGAGCCCCACTA	hPDGFC qR1	GGGTCTTCAAGCCCCAAATCTTT
hGli3 qR1	GAATCGGAGATGGATCGTAATGG	mPDGFC qF1	TTTGATGAGAGRRRRGGGCTGG
mGli3 qF1	CACAGCTCTACGGCGACTG	mPDGFC qR1	AGAATCCGGGTTCCAGATGGAA
mGli3 qR1	CTGCATAGTGATTGCGTTTTCTC	hRANK F1	GCTCCTCCATGTACCAGTGAG
hHMGA2 qF1	ACCCAGGGGAAGACCCAAA	hRANK R1	ACTGTCAGAGGTAGTAGTGCATT
hHMGA2 qR1	CCTCTTGGCCGTTTTTCTCCA	mSnail1 qF1	CACACGCTGCCTTGTGTCT
h-ITGB1 qF1	GTAACCAACCGTAGCAAAGGA	mSnail1 qR1	GGTCAGCAAAGCAGGGTT
h-ITGB1 qR1	TCCCCTGATCTTAATCGCAAAAC	hSnail1 qF1	TCGGAAGCCTAACTACAGCGA
m Itgb1 qF1	ATGCCGTATATTAGCACAAACC	hSnail1 qR1	AGATGAGCATTGGCAGCGAG
m Itgb1 qR1	GATATGCGTTGCTGACCAACA	mSnail2 qF1	TGGTCAAGAAACATTTCAACGCC
hJagged1 qF1	CAGTTCGAGTTGGAGATCCTGT	mSnail2 qR1	GGTGAGGATCTCTGGTTTTGGTA
hJagged1 qR1	CCTTGAGGCACACTTTGAAGTA	hSnail2 qF1	AAGCATTTCAACGCCTCCAAA
hKrt8 qF1	GGAGGCATCACCGCAGTTAC	hSnail2 qR1	GGATCTCTGGTTGTGGTATGACA
hKrt8 qR1	GGTTGGCAATATCCTCGTACTGT	hSmad3qF1	CCATCTCCTACTACGAGCTGAA
mKrt8 qF1	TCCATCAGGGTGACTCAGAAA	hSmad3qR1	CACTGCTGCATTCTCTGTTGAC
mKrt8 qR1	CCAGCTTCAAGGGGCTCAA	hSmad6qF1	GCTACCAACTCCCTCATCACT
hKrt14 qF1	TGAGCCGCATTCTGAACGAG	hSmad6qR1	CGTACACCGCATAGAGGCG
hKrt14 qR1	GATGACTGCGATCCAGAGGA	hSmad7qF1	TTCTCCGCTGAAACAGGG
mKrt14 qF1	CCTCTGGCTCTCAGTCACTCC	hSmad7qR1	CCTCCCAGTATGCCACCAC
mKrt14 qR1	TGAGCAGCATGTAGCAGCTT	hSOX9 qF1	AGCGAACGCACATCAAGAC
hKrt18 qF1	ACAATGCCCGCATCGTTCT	hSOX9 qR1	GCTGTAGTGTGGGAGGTTGAA

hTGFb1 qF1	CAATTCCTGGCGATACCTCAG	mTwist qF1	CGGGTCATGGCTAACGTG
hTGFb1 qR1	AGATAACCACTCTGGCGAGTC	mTwist qR1	CAGCTTGCCATCTTGGAGTC
mTGFb1 qF1	CCGCAACAACGCCATCTATG	hVegfA qF1	ACGAGGGCCTGGAGTGTGT
mTGFb1 qR1	CCCCAATGTCTGACGTATTGAAG	hVegfA qR1	CCGCATAATCTGCATGGTGTAT
hTGFb2 qF1	CAGCACACTCGATATGGACCA	mVegfA qF1	CCAGACCTCTCACCGGAAAG
hTGFb2 qR1	GTTGTAGATGGAAATCACCTCCG	mVegfA qR1	CTGTCAACGGTGACGATGATG
mTGFb2 qF1	AGAATCGTCCGCTTTGATGTC	h-vimentin qF1	AGTCCACTGAGTACCGGAGAC
mTGFb2 qR1	TCTGGTTTTACAACCTTGCT	h-vimentin qR1	GGTTCCTTTAAGGGCATCCAC
hTGFb3 qF1	CACCCAGGAAAACACCGAGTC	m-vimentin qF1	CGTCCACACGCACCTACAG
hTGFb3 qR1	GCGGAAAACCTTGGAGGTAAT	m-vimentin qR1	GGGGGATGAGGAATAGAGGCT
mTGFb3 qF1	GGACTTCGGCCACATCAAGAA	hZEB1 qF1	TTACACCTTTGCATACAGAACCC
mTGFb3 qR1	TAGGGGACGTGGGTCAACAC	hZEB1 qR1	TTTACGATTACACCCAGACTGC
hTGFBR2 qF1	CTCCAATATCCTCGTGAAGAACG	mZEB1 qF1	GCTGGCAAGACAACGTGAAAG
hTGFBR2 qR1	CCCCTGTTAGCCAGGTCATC	mZEB1 qR1	GCCTCAGGATAAATGACGGC
hTGFBR3 qF1	CTGGGGTCTCCAGACTGTTTT	hZEB2 qF1	AACAACGAGATTCTACAAGCCTC
hTGFBR3 qR1	CTGCTCCATACTCTTTTCGGG	hZEB2 qR1	TCGCGTTCTCCAGTTTTTCTT
hTwist1/2 qF1	GTCCGCAGTCTTACGAGGAG	mZEB2 qF1	ATTGCACATCAGACTTTGAGGAA
hTwist1/2 qR1	GCTTGAGGGTCTGAATCTTGCT	mZEB2 qR1	ATAATGGCCGTGTCGCTTCG

3' UTR Cloning Primers	
Angptl4 3' UTR F2 Sgfl	gcgatcgc TGCAGGCTACCA
Angptl4 3' UTR R1	CGGATCTGGGAATACGATGTCGGAG
Lox 3' UTR F1 XhoI	ctcgag AGAAGCCCAGTTCACAG
Lox 3' UTR R1	GTAGCAGGTAAGTGCATGGCTT
Loxl2 3' UTR F1 XhoI	ctcgag AGAAGCCTGCGTGG
Loxl2 3' UTR R1	TTCAGTAAAAACACAGGAGTTCAAG
Loxl4 3' UTR F1 XhoI	ctcgag GGAGGCCTTTTCGGT
Loxl4 3' UTR R1	CTGGCCTTTGTGGCAGCCCA
MMP2 3' UTR F1 XhoI	ctcgag TGAAGTACCCTCCCG
MMP2 3' UTR R1	GCGTGCCTCCAGTTCTTCC
MMP9 3' UTR F2 XhoI	ctcgag ACACTGGGCTTGGCG
MMP9 3' UTR R2	GGGCACAAGTCCCCACGTGTC
PdgfA 3' UTR F1 Sgfl	gcgatcgc GGATGTGCGTGCC
PdgfA 3' UTR R1	AGGTGGTGGCTTCCAGGCCG
PdgfB 3' UTR F1 XhoI	ctcgag CAGCATGGTGGCAG
PdgfB 3' UTR R1	AATGGCGGGCATTGAGGGCC
PdgfC 3' UTR F1 XhoI	ctcgag GCCTTCGTAGCAGCA
PdgfC 3' UTR R1	CGAGGTTTCTGGGGACGTGCTT
VegfA 3' UTR F4 XhoI	ctcgag CTGGGCGGCCTTCG
VegfA 3' UTR R4	TGTTCCCCAGAGGAAGAGGGAAAGG

Chapter 4: GATA3 inhibits bone metastasis and expression of *Osteopontin* and *Rank*

Source: The following chapter contains unpublished data that is being prepared as a manuscript.

Contributions: Sylvain Provot initiated this project to study Gata3 and bone metastasis, and most of the experimental bone metastasis experiments with Jung-whan Kim and Audrey Brenot. I generated the Opn shRNA knockdown cells, did the qPCR and western blots, and performed the Gata3 and Opn shRNA in vivo experiments together with Sylvain Provot and Audrey Brenot. Zena Werb supervised the project.

GATA3 inhibits bone metastasis and expression of *Osteopontin* and *Rank*

Jonathan Chou^{1,2,*}, Sylvain Provot^{1,4,*}, Audrey Brenot¹, Jung-whan Kim^{1,5}, and Zena Werb^{1,2,3}

¹Department of Anatomy, and ²Biomedical Sciences Program, University of California, San Francisco CA 94143-0452

*These authors contributed equally to this work.

³Corresponding Author:

Zena Werb, Ph.D.

Department of Anatomy, HSW 1323

University of California

San Francisco, CA 94143-0452 USA

Tel.: (415) 476-4622

Fax: (415) 476-4565

zena.werb@ucsf.edu

Keywords:

GATA3; bone metastasis; osteopontin; Spp1

Present Address:

⁴INSERM U606 Hôpital Lariboisière

2 Rue Ambroise Paré

75010 Paris, France

⁵ Salk Institute for Biological Studies

10010 N. Torrey Pines Road

La Jolla, CA 92037

ABSTRACT

Bone metastasis occurs commonly in patients with breast cancer, leading to severe pain and bone erosion. However, the molecular and cellular mechanisms governing metastasis to the bones remain poorly understood. We previously showed that GATA3 specifies and maintains luminal cell differentiation in mammary epithelial cells, and that its expression is lost during tumor progression. High GATA3 expression correlates with more favorable prognosis in human patients. Here, using an experimental bone metastasis model, we show that GATA3 suppresses bone metastases. Interestingly, we identified two genes down-regulated by GATA3, osteopontin (*Opn/Spp1*) and receptor activated NF- κ B (*Rank*), which have previously been implicated to play a role in bone metastasis. *Opn/Spp1* is enriched in the aggressive basal breast cancer subtype, and inhibition of *Opn/Spp1* suppresses bone metastases, without affecting primary tumor size, phenocopying the anti-metastatic effects of GATA3. This suggests that GATA3-mediated metastasis suppression may work through regulation of *Opn/Spp1*. Future studies will be aimed at investigating the role of *Rank* in metastasis and determining whether over-expression of these genes in the context GATA3-expression restores metastatic capacity to the bones. Taken together our results indicate that GATA3 has a general anti-metastatic effect to multiple organ sites, and suggest that differentiation-promoting transcription factors such as GATA3 have pleiotropic effects on tumor biology. This work has implications for preventing and treating bone metastasis in breast cancer.

INTRODUCTION

Many breast cancer patients with advanced disease develop metastases to the bones. Metastatic bone disease generally results in severe bone pain, which is likely due to the osteolytic nature of metastases (Mundy 2002). Because the bones represent a rich environment full of growth factors, a vicious cycle is established between tumor cells and the bones, such that growth factors released from the bone matrix stimulate proliferation and further osteolysis. Several genes have been shown to play a role in the development of bone metastases, including parathyroid-hormone related peptide (PTHrP), receptor activator of nuclear factor- κ B ligand (RANKL), transforming growth factor- β (TGF β) and various proteases. Elucidation of these operative pathways in bone metastasis has led to clinical trials of antibodies targeting RANK ligand (e.g., denosumab) and PTHrP.

In an effort to identify genes that mediate metastasis to the bones, Massague, Kang and colleagues have used an experimental bone metastasis model in which cancer cells are inoculated into the left ventricle of the heart. By iteratively selecting cells that form bone metastases, they identified a subpopulation with enhanced bone metastatic abilities and transcriptionally profiled their gene expression. They found a number of candidates involved in bone homing (CXCR4), angiogenesis (FGF-5, CTGF), invasion (MMP1) and osteoclast differentiation (IL-11) that were differentially expressed between the parental cells and the bone metastatic derivatives. When over-expressed in the parental cell line, CXCR4 and IL-11 increased metastasis. Interestingly, the

authors found that over-expression of osteopontin (*SPP1* or *OPN*) further enhanced metastasis and dramatically decreased tumor latency. These data suggest that many factors work collaboratively in this process, and highlighted the central role of osteopontin.

OPN is a secretory protein with multiple functions, including the ability to stimulate adhesion to bone matrix. Clinical studies have revealed a correlation between plasma OPN, tumor burden and prognosis in breast cancer patients; those with high OPN levels have higher tumor burden and worse outcomes. OPN is thought to exert its pro-metastatic effects by interacting with various integrins (e.g, $\alpha v \beta 3$, $\alpha 4 \beta 1$) and CD44 receptors and regulating the downstream cell signaling events that lead to tumor progression. OPN also regulates the bioavailability of matrix metalloproteinases (MMPs) (Rangaswami et al. 2006; Chen et al. 2009). In addition, secretion of OPN from primary tumors activates bone marrow cells and induces the outgrowth of distant otherwise indolent tumors (McAllister et al. 2008).

In addition to OPN, RANK and its ligand, RANKL, and the decoy receptor of RANKL, osteoprotegerin (OPG) play pivotal roles in bone remodeling by regulating osteoclast formation and activity. RANKL stimulates migration of RANK-expressing tumor cells, which can be inhibited by OPG. RANK is expressed by many solid primary tumors and positively correlates with bone metastasis (Santini et al. 2011a). Low RANK and high OPG mRNA levels also correlate with longer overall survival (Santini et al. 2011b). These data suggest

that RANK expression in primary tumors can serve as a predictive marker of bone metastasis.

Despite the importance both OPN and RANK to bone metastasis, their connection with tumor differentiation remains unclear. In addition, how these genes are regulated at the transcriptional and post-transcriptional levels is not well understood. We and others previously showed that GATA3 is the master transcription factor that specifies and maintains luminal cell fate in the mouse mammary gland (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007). Expression profiling of breast cancer patients has established that low GATA3 is associated with poor prognosis (Mehra et al. 2005; Jacquemier et al. 2009; Yoon et al. 2010). In this study, we have investigated the role of GATA3 in regulating bone metastasis, and identified two important downstream targets, *Opn/Spp1* and *Rank* that mediate bone metastasis.

RESULTS AND DISCUSSION

Gata3 expression in basal breast cancer cells inhibits bone metastasis

To determine whether Gata3 expression affects metastasis to the bones, we performed intra-cardiac (i.c.) injections of 4T1 cells \pm Gata3 (previously described in **Chapter 3**) into the left ventricle of the heart. We found that mice injected with 4T1-Gata3 cells formed significantly fewer bone metastases than the 4T1-Control cells by bioluminescent imaging (**Fig. 4.1A–4.1B**). We confirmed this by histological examination of the bones. Of the mice that did have bone metastases, we found that they were smaller in size (**Fig. 4.1C**). Because of the role of *Opn/Spp1* in bone metastasis, we sought to determine whether its expression was altered. Interestingly, in situ hybridization for *Opn* mRNA in the metastases showed decreased signal in 4T1-Gata3 tumors (**Fig. 4.1D**). Taken together, these results indicate that Gata3 expression inhibits experimental bone metastases in vivo, and is associated with a decrease in *Opn* expression.

Gata3 suppresses expression of two bone metastasis regulators, *Opn/Spp1* and *Rank*

We next sought to investigate the effect of Gata3 on *Opn/Spp1* expression. We utilized mouse 4T1 cells and human MDA-MB-231 cells (abbreviated MDA231 hereafter), two aggressive, triple-negative breast cancer cell lines that represent the basal subtype. By quantitative PCR (qPCR), we found that in both 4T1 and MDA231 cells, Gata3 inhibited *Opn/Spp1* expression by 50 – 70% (**Fig. 4.2A–4.2B**). Because the RANK-RANKL axis regulates multiple aspects of bone

metastasis and correlates with aggressiveness (Mundy 2002), we also asked whether *Rank* expression changed with Gata3. We found that Gata3 also inhibited *Rank* in both 4T1 and MDA231 cells (**Fig. 4.2C–4.2D**). Taken together, these results demonstrate that Gata3 suppresses two critical mediators of bone metastasis.

We wanted to determine whether Gata3 inhibits *Opn/Spp1* by direct transcriptional repression. We used a luciferase reporter containing the human *OPN* promoter (pGL3-OPN), which is ~1200 base pairs upstream of the *OPN* transcriptional start site (TSS). Transfection of pGL3-OPN with Gata3 did not decrease luminescence activity (**Fig. 4.2E**), whereas transfection of Gata3 with pGL3-Ecad (containing the *Cdh1* promoter, which is ~1000 base pairs upstream) increased luciferase activity, as expected (**Fig. 4.2F**). These results indicate that while Gata3 inhibits mRNA expression of *Opn*, this regulation does not involve direct repression at the promoter ~1200 base pairs upstream of the TSS. This suggests that Gata3 either regulates *Opn* expression indirectly through an unidentified intermediate, or that the binding site is outside this region, e.g., at an enhancer. Future work is aimed at addressing this issue.

***OPN/SPP1* expression is enriched in basal type tumors**

We wanted to ask whether *OPN* expression was enriched in particular subtypes of breast cancer. Two broad categories that have emerged from microarray analyses are luminal (which exhibit better prognoses) and basal (which are more aggressive). Using a previously published array on human breast cancer cell

lines (Neve et al. 2006), we found that *OPN* expression was enriched in basal cells (**Fig. 4.3A**). However, *RANK* expression was similar across luminal and basal subtypes (**Fig. 4.3B**). Because basal type tumors are generally more aggressive than luminal type tumors, we were intrigued by the differential *OPN* expression pattern in these two broad subtypes, which prompted us to focus more on the role of *Opn/Spp1* for subsequent studies.

Opn expression increases during PyMT tumor progression

The MMTV-PyMT model mimics progressive stages of human luminal breast cancer (Lin et al. 2003), including the development of hyperplasia to adenoma to carcinoma. We isolated primary tumor cells from these different stages of tumor progression and analyzed them for *Opn/Spp1* expression by qPCR. Adenoma cells had a significant increase in *Opn/Spp1* levels compared to normal primary mammary epithelial cells (MECs), whereas carcinoma cells expressed even higher levels of *Opn/Spp1* (**Fig. 4.3C**). We have not yet investigated whether *Rank* expression also changes during tumor progression.

Loss of Opn decreases experimental bone metastasis, recapitulating the effects of Gata3 expression

To study the consequences of down-regulating *Opn/Spp1*, we generated stable *Opn/Spp1* knockdown cell lines using shRNA lentiviruses. We used three shRNA hairpins and assessed knockdown by qPCR (**Fig. 4.4A**) and by Western blot (**Fig. 4.4B**). We found that the F10 shRNA hairpin gave the most robust

knockdown (~85% by qPCR), followed by the H5 shRNA hairpin (~65% by qPCR). Unfortunately, the D10 shRNA hairpin did not show any decrease in *Opn* mRNA or OPN protein levels (data not shown). We next measured cell viability in vitro, and found that there were no significant differences in proliferation (**Fig. 4.4C**). Because we found the most robust inhibition with the F10 hairpin, we decided to conduct our in vivo experiments using those knockdown cell lines.

To assess the effects on bone metastasis, we injected mice i.c. with 4T1-Control or 4T1-shOpn-F10 cells. In an initial cohort of mice (n = 4 mice in each group), we found that loss of *Opn* decreases bone metastases (**Fig. 4.4D**). Although these preliminary findings are encouraging, additional studies to increase the sample size will need to be conducted before we can conclude whether loss of *Opn* in aggressive breast cancer cells decreases the development of bone metastases.

Loss of *Opn* decreases experimental lung metastasis, but does not impair primary tumor growth

To assess whether loss of *Opn* also affects metastasis development at other sites such as the lung, we injected the 4T1-shOpn cells i.v. through the tail vein to allow cells to enter venous circulation. In a preliminary cohort of mice (n = 4 mice in each group), we found that mice injected with 4T1-shOpn-F10 cells had fewer lung metastases compared to the 4T1-Control cells, as detected by bioluminescence imaging (**Fig. 4.4E**).

To assess effects on primary tumor growth, we injected cells into the fourth mammary gland to form orthotopic tumors. We found that primary 4T1-Control and 4T1-shOpn-F10 tumors grew at similar rates in vivo (**Fig. 4.4F**). However, silencing *Opn* in 4T1 cells inhibited spontaneous metastases from the primary tumor (**Fig. 4.4G–4.4H**). Although these results are preliminary, the data suggest that knockdown of *Opn* reduces lung and bone metastasis, but does not affect growth of the primary tumor, similarly to the effects of Gata3. Future additional studies to increase the sample size will be required. In addition, we will need to assess whether *Opn* re-expression in 4T1-Gata3 cells restores metastasis.

CONCLUSION

We have found that Gata3 expression inhibits bone metastasis in an immunocompetent mouse model of breast cancer. This decrease in bone metastasis is accompanied by a decrease in *Opn* expression in vivo. We found that Gata3 decreases expression of two known mediators of bone metastasis: *Opn* and *Rank*. Interestingly, *OPN* expression is enriched in the poor prognostic, basal subtype of breast cancer.

OPN has previously been implicated in promoting bone metastasis formation, and its expression is associated with poor prognosis in human patients. Mechanistically, OPN enhances adhesion to the bone matrix, stimulates proliferative signaling, promotes cell migration through integrins and regulates MMPs (Rangaswami et al. 2006). The latter aspect is particularly notable, given the role of MMPs in releasing growth factors sequestered in the extracellular matrix, such as VEGF (Bergers et al. 2000). Thus, OPN regulates both cell intrinsic and extrinsic properties important for tumor cell invasion, survival and metastasis. Using a loss-of-function approach, we have found that knockdown of *Opn* inhibits metastasis to both the bones and the lungs, without affecting tumor growth. Future work will be aimed at understanding how Gata3 regulates *Opn*, and further validating that knockdown of *Opn* inhibits metastasis. We will also determine whether re-expression of *Opn* in Gata3 cells restores metastatic capacity. In addition, it will be important to assess whether regulation of *Rank* contributes to Gata3-mediated bone metastasis suppression.

Taken together, this work highlights how differentiation-promoting transcription factors such as GATA3 affect both cell intrinsic and extrinsic properties, which ultimately lead to metastasis suppression.

ACKNOWLEDGEMENTS

We thank members of the Werb laboratory for discussions and especially Elena Atamaniuc, Ying Yu, and Helen Capili for technical assistance. We thank Tara Rambaldo and Michael Kissner for flow cytometer assistance and Piia Aarnisalo for reagents. We also thank Charina Choi for discussions and encouragement. This research was supported by funds from the National Cancer Institute (R01 CA129523 to Z.W.), a Developmental Research grant from the Bay Area Breast Cancer SPORE (P50 CA058207 to Z.W.), a Department of Defense Predoctoral Fellowship (W81XWH-10-1-0168 to J.C.) and the UCSF Medical Scientist Training Program (J.C.).

EXPERIMENTAL PROCEDURES

Animal Studies

All animal experiments were performed at UCSF, and reviewed and approved by the UCSF IACUC. Mice were housed under pathogen-free conditions in the UCSF barrier facility. Balb/c mice were purchased from Simonsen Laboratories. For experimental bone metastasis experiments, age-matched female mice were injected i.c. (via left heart ventricle) with 1×10^5 cells (4T1). For experimental lung metastasis experiments, age-matched female mice were injected i.v. (via tail vein) with 1×10^5 cells (4T1). For primary tumors and spontaneous metastasis assays, age-matched female mice were injected with 1×10^5 cells (4T1) mixed 1:1 with DMEM:Matrigel (BD Biosciences) into the fourth mammary fat pad without clearing. Tumor measurements were made using a caliper at least once per week, and volumes were calculated using $V = 0.52 \times (\text{length})^2 \times \text{width}$. Bioluminescence imaging was performed using an IVIS Spectrum (Caliper Life Science). Image radiance values were normalized using the Living Image software.

Cell culture and viral transduction

MDA-MB-231 and 4T1 were grown in standard conditions (DMEM with 10% FBS). Knockdown shRNA *Opn* lentivirus and Gata3 retrovirus were generated in HEK 293T cells using standard transfection procedure with calcium phosphate. The virus was concentrated by ultracentrifugation. Cells were transduced in media containing polybrene, and then selected in puromycin for at least 5 days or

sorted by FACS using a FACS Aria II (Becton Dickinson) (UCSF Stem Cell FACS Core).

Plasmids

pGIPZ shRNA plasmids to knockdown *Opn* were obtained from Open Biosystems. The pGIPZ-scramble was used as the control. The pMIG-Gata3 plasmid was previously described (Kouros-Mehr et al. 2008). The pGL-*Opn*-Luc reporter plasmids were a kind gift from Dr. Piia Aarnisalo (Univ. of Helsinki) and from Addgene (plasmid # 11996, Steven Johnson).

Quantitative real-time PCR (qPCR)

Total RNA was isolated from cells using the miRNeasy Mini Kit (Qiagen). cDNA was synthesized using the Superscript III RT First Strand Kit (Invitrogen). qPCR was performed using FastStart Universal SYBR Green master mix (Roche) in an Eppendorf Mastercycler realplex machine. Ct values were normalized to actin and GAPDH, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. Primer sequences for qPCR were found using the Harvard Primer Bank.

Luciferase Assays

For *Opn*-Luc promoter assays, the pGL3-*Opn* reporter, which contains ~1000 bp of the *Opn* promoter upstream of the start site, was co-transfected with pRL-TK with or without GATA3 into 293T or MDA-MB-231 cells. Lysates were collected 24–48 hr post-transfection, and renilla and firefly luciferase activities were

measured using the Dual-Luciferase Reporter System and a GloMax luminometer (Promega).

Immunostaining and in situ hybridization

Tissues were fixed in 4% PFA overnight, paraffin processed or embedded into OCT for frozen sections, and sectioned. Antigen retrieval was performed using citrate buffer or proteinase K (Kouros-Mehr et al. 2006). The TSA Amplification Kit (Perkin Elmer) was used according to manufacturer instructions. Primary antibodies were incubated overnight, and secondary antibodies were incubated for 1 hour. The following antibodies were used: OPN (R&D), phospho-histone H3 and cleaved caspase-3 (Cell Signaling), biotinylated anti-rabbit (Dako) and biotinylated anti-goat (Jackson). H&E stainings were performed using standard protocols. In situ hybridization was carried out using digoxigenin-labeled probes. Image analysis was performed using ImageJ software.

Western blotting

Cells were lysed in RIPA buffer plus protease inhibitors (Roche) or directly in Lamelli Buffer with DTT. Protein concentration was measured using the BCA Protein Assay Kit (Thermo Scientific). Lysates were subjected to SDS-PAGE, transferred to PVDF membranes, blocked in 5% BSA, incubated with primary antibody overnight and visualized using ECL Detection Reagents (Pierce). Exposures were acquired using a LAS-4000 Imager (Fuji). Primary antibodies to

actin (Santa Cruz) and osteopontin (R&D) were used at manufacturer's recommended dilutions.

Cell viability assay

Cell viability was measured using the CellTiter MTT Assay according to manufacturer's instructions (Promega). Absorbance at 590 nm was read using a standard plate reader (Bio-Rad).

Statistical analysis

Statistical analysis was performed using Prism 4 software (GraphPad Software, Inc.). All data are presented as mean \pm SEM, unless otherwise stated. When two groups were compared, the two-tailed Student's t-test was used, unless otherwise stated. When three or more groups were compared, the one-way analysis of variance (ANOVA) test was used, followed by Tukey's test to determine significance between groups. We considered $p < 0.05$ as significant.

FIGURE LEGENDS

Fig. 4.1: Gata3 inhibits bone metastasis and decreases *Opn* expression vivo.

(A) Bioluminescence imaging of Balb/c mice two weeks after intracardiac (i.c.) injection of 4T1 cells \pm Gata3 to monitor bone metastasis.

(B) The percentage of hind limbs positive for bone metastasis were scored and the combined results across multiple experiments are shown. (n=20 mice in each group, * denotes $p < 0.001$)

(C) H&E images of bone metastases in mice injected with 4T1-Control or 4T1-Gata3 cells. 4T1-Gata3 metastases were consistently and significantly smaller than the 4T1-metastases.

(D) In situ hybridization for *Opn* mRNA in 4T1 \pm Gata3 bone metastases shows a significant decrease in *Opn* in vivo.

Fig. 4.2: Gata3 inhibits expression of *Opn* and *Rank*, two genes that promote bone metastasis, but not through a direct transcriptional mechanism.

(A–B) *Opn/Spp1* mRNA levels in 4T1 cells \pm Gata3 (A) and in MDA231 cells \pm GATA3 (B), measured by quantitative PCR (qPCR).

(C–D) qPCR for *Rank* expression in 4T1 cells \pm Gata3 (C) and in MDA231 cells \pm GATA3 (D).

(E) Luciferase activity in MDA231 cells using the *Opn* promoter reporter (as described in (Lammi et al. 2004)), with or without Gata3 co-transfected.

(F) As a positive control, luciferase activity in MDA231 cells using the *Cdh1* (*E-cadherin*) promoter reporter, with or without Gata3 co-transfected.

Fig. 4.3: *OPN*, but not *RANK*, is enriched in basal type cancers and increases during tumor progression.

(A–B) *OPN* (A) and *RANK* (B) mRNA levels in human luminal and basal breast cancer cell lines. Dataset adapted from Neve et al, 2006.

(C) Relative *Opn* mRNA levels in normal primary mammary epithelial cells, and MMTV-PyMT primary adenomas and carcinomas, measured by qPCR.

Fig. 4.4: *Opn* knockdown decreases experimental lung and bone metastases, without affecting primary tumor growth.

(A–B) *Opn* mRNA levels as measured by qPCR (A) and protein levels as measured by western blot (B) in 4T1 cells stably transduced with two shRNAs targeting endogenous *Opn* (F10 shRNA and H5 shRNA).

(C) MTT cell viability assay to assess proliferation in 4T1-Control, 4T1-F10 shOPN and 4T1-H5 shOPN cells in vitro.

(D) Bioluminescence imaging of Balb/c mice two weeks after i.c. injection of 4T1 cells \pm shOpn to monitor bone metastases (n = 4 mice in each group).

(E) Bioluminescence imaging of Balb/c mice two weeks after i.v. injection of 4T1 cells \pm shOpn to monitor lung metastases (n = 4 mice in each group).

(F–H) 4T1 \pm shOpn cells were injected into the fourth mammary fat pad (n = 3 mice in each group), and primary tumor size was measured (F). The number of

spontaneous lung metastases in each group (G) and representative H&E images of the spontaneous lung metastases (H) are shown.

FIGURES

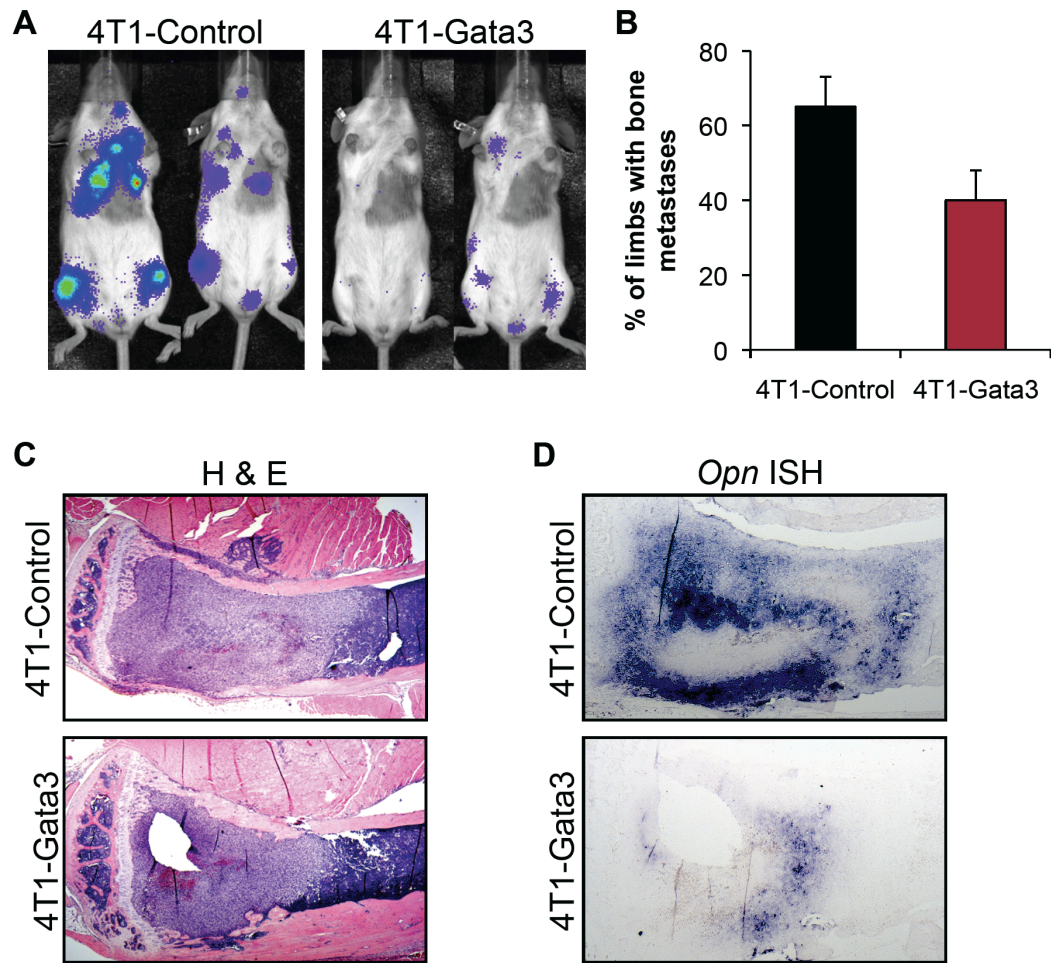


Figure 4.1

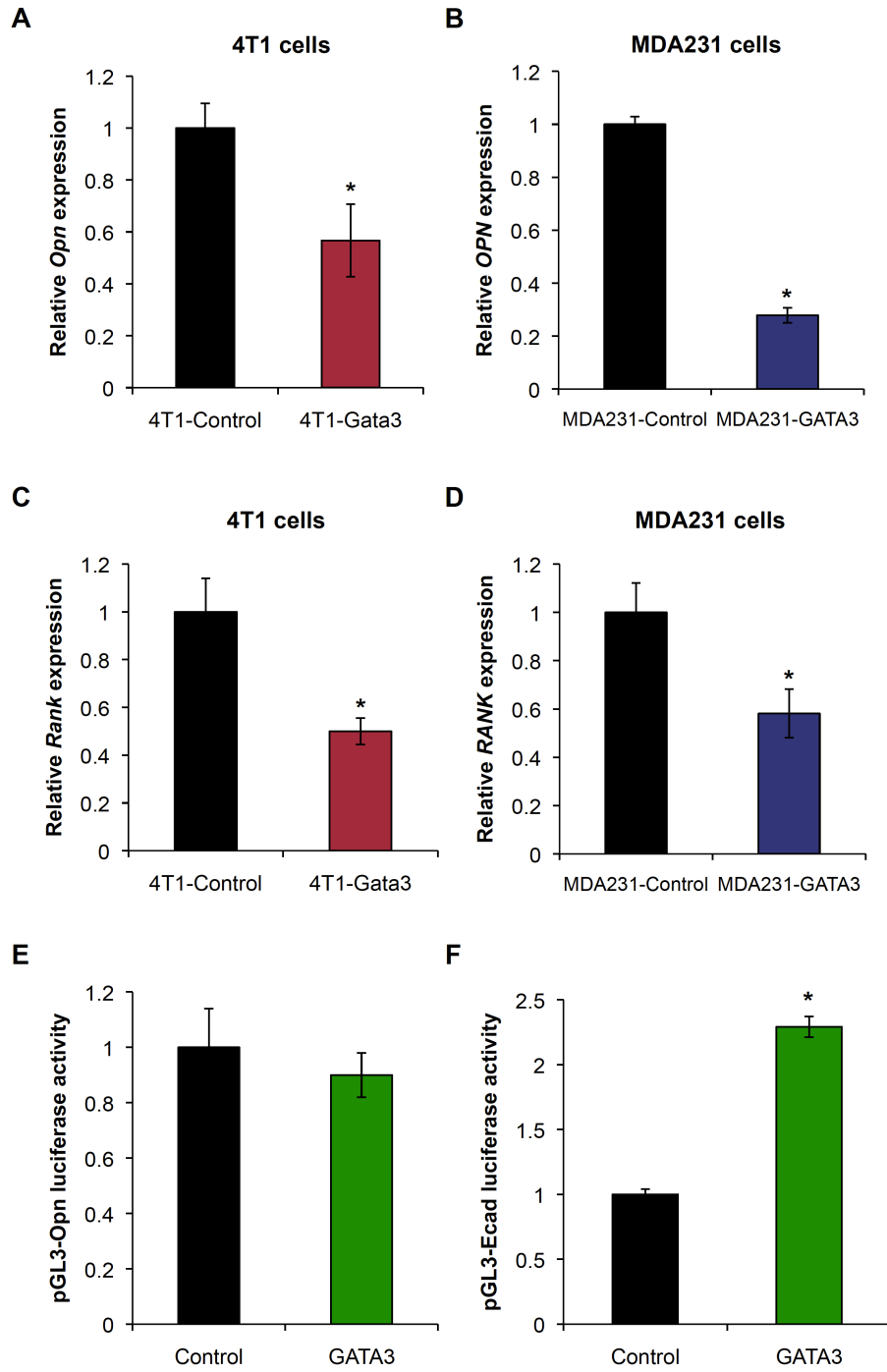


Figure 4.2

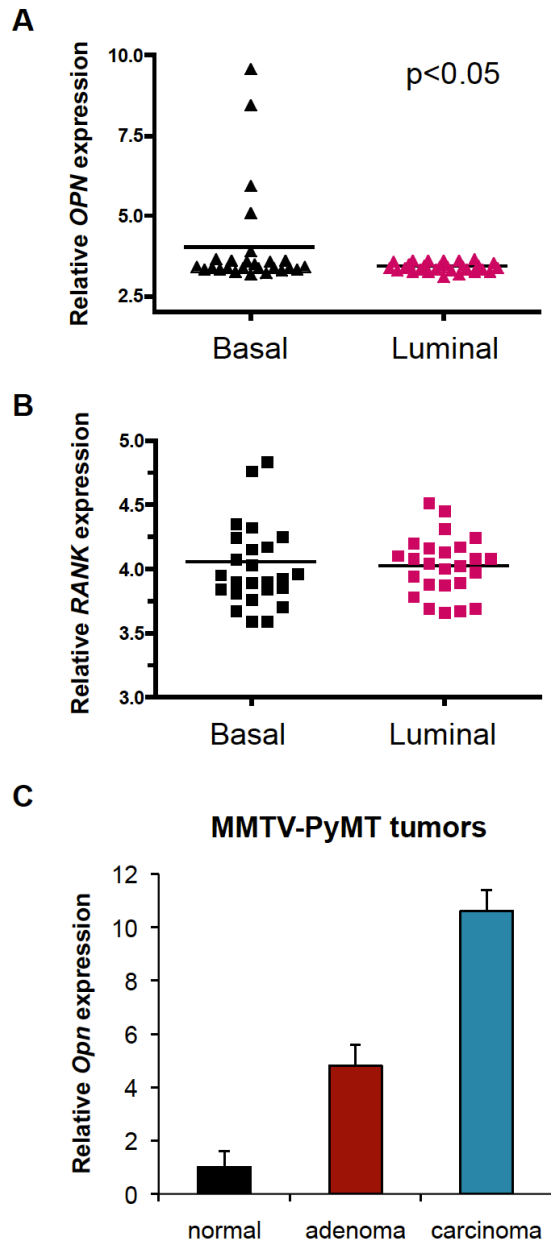


Figure 4.3

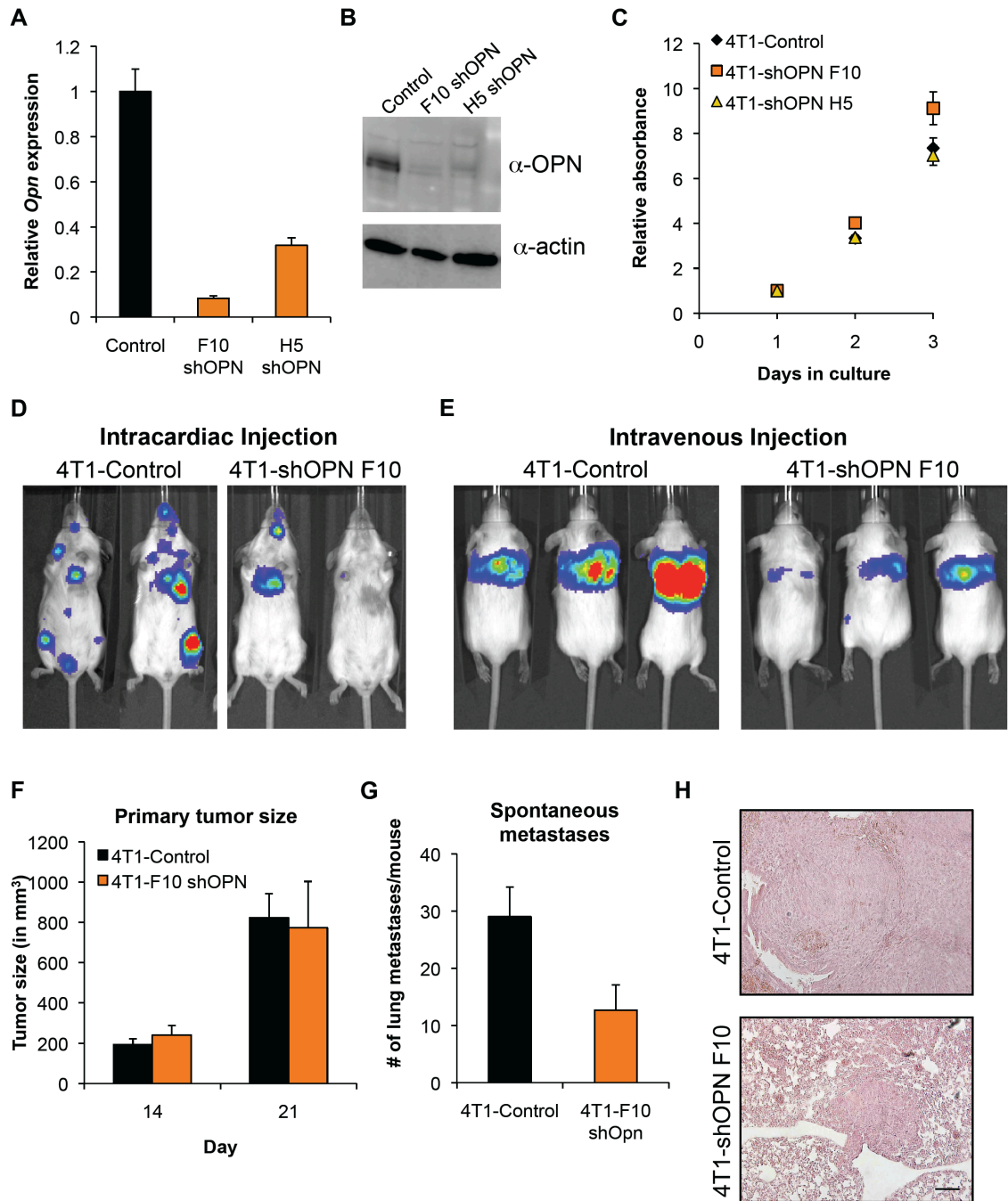


Figure 4.4

Chapter 5: Genome-wide identification of GATA3 binding sites in breast cancer cells by ChIP-seq reveals candidate genes involved in luminal differentiation and metastasis suppression

Source: The following chapter contains preliminary, unpublished data that is being prepared as a manuscript.

Contributions: This project is a collaboration with Peggy Farnham's lab at the University of Southern California (USC)/UC Davis, who have expertise in ChIP-Seq technology and analysis. I initiated this project with Henriette O'Geen when Peggy Farnham was still at UC Davis. Henriette O'Geen, Zena Werb and I conceived the idea to perform genome-wide ChIP-Seq on GATA3 in human and mouse basal breast cancer cell lines. I crosslinked and provided samples and discussed experiments with Henriette O'Geen, who performed the ChIP-seq and peak-calling analysis. I validated expression data for the targets presented in this Chapter, and assembled the expression data from the Neve et al. 2006 data set.

Since Peggy Farnham's move to USC, I have continued the project with Lijing Yao. Peggy Farnham and Zena Werb supervised the project.

ABSTRACT

Cellular differentiation is controlled by transcription, which is coordinated by transcription factors (TFs), chromatin modifying enzymes, and chromatin, composed of DNA and histone proteins. Covalent modification of DNA and histones (termed epigenetics) governs the accessibility of TFs for their target DNA sequences, which ultimately determines gene expression. Because of the advances in ChIP-seq technology, it is now possible to identify transcription factor binding sites in a genome-wide, comprehensive manner. This has revealed novel TF targets and given insight into how gene expression programs are regulated. The GATA family of TFs, of which there are six in mice and humans, are master regulators of development. These factors interact with simple DNA motifs (WGATAR, where W specifies A/T and R specifies A/G) to regulate critical processes, including mammary gland development, yet very few WGATAR motifs are actually occupied within a given cell type. We present here the initial data sets from two GATA3 ChIP-seq experiments using highly aggressive, human and murine breast cancer cells. We validate a cohort of genes identified from the ChIP-seq by expression profiling to show that these targets are indeed regulated by GATA3. Our analysis identifies the TGF β pathway as a key downstream component regulated by GATA3, and implicates several novel gene targets controlled by GATA3 that may be involved in luminal differentiation. In accordance with this, several of these genes are differentially expressed between human luminal and basal breast cancer cells, correlating with GATA3 expression. We hypothesize that many of these genes may be important for

luminal epithelial cell differentiation in the mammary gland, and may play critical roles during mammary tumorigenesis and metastasis.

INTRODUCTION

The GATA family of transcription factors (TFs), of which there are six in mice and humans, are master regulators of development. The GATA family of TFs interact with simple DNA motifs (WGATAR, where W specifies A/T and R specifies A/G) to regulate critical developmental processes, including mammary gland, skin, neuronal, hematopoietic and renal development, yet very few WGATAR motifs are actually occupied within a specific cell type. Indeed, a recent genome-wide study identifying GATA3 binding sites in distinct T cell types illustrates that GATA3-mediated gene regulation depends strongly on co-factors already expressed in different subtypes of cells (Wei et al. 2011).

We and others have previously shown that GATA3 is a master transcription factor that specifies and maintains luminal cell fate in the mouse mammary gland (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007). In addition, loss of *GATA3* is causally involved in the pathogenesis of breast cancer (Kouros-Mehr et al. 2006; Asselin-Labat et al. 2007; Kouros-Mehr et al. 2008). Expression profiling studies of human breast cancer patients has established that low GATA3 is associated with poor prognosis (Mehra et al. 2005; Jacquemier et al. 2009; Yoon et al. 2010). In several mouse models of breast cancer, including the mammary tumor virus (MMTV) promoter-driven polyoma middle T antigen (PyMT) and MMTV-Neu models, GATA3 expression inversely correlates with tumor progression and metastasis: loss of GATA3 coincides with loss of differentiation, the transition from adenoma to carcinoma and the onset of tumor dissemination. Strikingly, re-introduction of *Gata3* into MMTV-PyMT carcinomas

transplanted orthotopically induces differentiation, suppresses dissemination (Kouros-Mehr et al. 2008) and reduces tumor-initiating capacity (Asselin-Labat et al. 2011). Moreover, mutations in the zinc fingers of GATA3, which diminish or abolish DNA binding ability, have been identified in human breast cancers, emphasizing the importance of GATA3-regulated genes (Usary et al. 2004). This suggests that the transcriptional networks controlled by GATA3 are important for both normal development and cancer. However, the molecular targets downstream of GATA3 in breast cancer cells are largely unknown; these targets are likely unique and different compared to the targets identified in T cells.

Given the limited knowledge of mechanisms underlying how GATA3 establishes and coordinates complex transcriptional networks, we began a collaboration with Professor Peggy Farnham's lab at the University of Southern California using chromatin immunoprecipitation followed by massive parallel sequencing (hereafter referred to as ChIP-seq) to define GATA3 occupancy sites genome-wide in human and murine breast cancer cells.

Briefly, the ChIP-seq assay first requires crosslinking of chromatin to stabilize the interaction between the candidate TF (in this case, GATA3) and chromatin. The chromatin is then sheared into smaller-sized fragments by sonication, and GATA3 is immunoprecipitated using an antibody. In this process, the DNA bound by GATA3 is also obtained. The DNA is then sequenced to saturation by next generation sequencing techniques, which provides sufficient sequencing depth to identify all target sites, and the reads are aligned to a reference genome. After converting the sequence data to position data along

each chromosome, these sequencing reads are analyzed with various peak-calling algorithms to identify the ChIP-enriched regions as TF binding target sites, and to determine the distance between the binding site identified and the transcriptional start site of the closest gene. The Farnham Lab has great expertise in this technique and are thus excellent collaborators for this project (Farnham 2009). They have previously investigated the mechanism of GATA1 and GATA2 in hematopoietic and endothelial cell development using ChIP-seq, which provided new insights into how these GATA factors function (Fujiwara et al. 2009; Linnemann et al. 2011).

We present here two GATA3 ChIP-seq datasets performed in human and mouse breast cancer cells. In addition, we validate a cohort of genes identified from the ChIP-seq data by expression profiling to demonstrate that these targets are indeed regulated by GATA3. Several of these genes are differentially expressed between luminal and basal breast cancer cells, correlating with GATA3 expression. We believe that these findings will give new insight into how GATA3 induces luminal cell differentiation and inhibits metastatic dissemination, and will reveal several novel and unexpected GATA3 targets, thus opening up new therapeutic possibilities for breast cancer.

RESULTS AND DISCUSSION

GATA3 ChIP-seq in MDA231 and 4T1 cells

Based on expression array analysis, human breast cancers are broadly classified into two subtypes that are prognostic of clinical outcome: luminal and basal (Sorlie et al. 2001). Clinically, basal breast cancers are aggressive and poorly differentiated, while luminal type cancers are more differentiated (Foulkes et al. 2010). In human breast cancer cell lines, GATA3 is expressed higher in good prognostic, luminal cells, and lower in the basal subtype.

We therefore decided to over-express GATA3 in two basal cell lines to determine the effect of GATA3 on tumor growth and metastasis in GATA3-low cells. We chose to use MDA-MB-231 cells (referred to as MDA231), a basal human breast cancer line with low endogenous levels of GATA3 expression, as well as 4T1 cells, a basal mouse breast cancer line that is highly metastatic. These two lines are referred to as “triple-negative” cells because they lack expression of three common markers of breast cell differentiation: estrogen receptor (ER), progesterone receptor (PR), and the HER2/NEU EGF receptor.

We found that GATA3 expression in both MDA231 and 4T1 cells dramatically reduces lung metastasis, and does so in part by regulating the tumor microenvironment. Indeed, primary MDA231-GATA3 and 4T1-Gata3 tumors have decreased tumor vasculature and macrophage infiltrates, two important pro-tumorigenic components of the microenvironment (see **Chapter 3**). In addition, GATA3 also inhibits metastatic colonization in an experimental lung metastasis model.

Based on these results, we performed a GATA3 ChIP-seq on both MDA231-GATA3 and 4T1-Gata3 cells to identify novel genes and microRNAs bound by the GATA3 transcription factor (**Tables 5.1 and 5.2**). Our initial peak-calling program only identified traditional genes and not microRNAs; future work to identify microRNAs that are within proximity of these binding sites will also be performed. In addition, the Farnham Lab has previously performed a GATA3 ChIP-Seq using the luminal MCF7 breast cancer line, which expresses high levels of endogenous but mutated GATA3 (data not shown). We hypothesized that by profiling multiple tumor cell lines, we would be able to identify common targets that are shared across cell lines, and unique targets that depend on co-factors expressed in each line. The unique background and subtype of cancer cells (e.g., Ras mutation status) has recently been shown to affect how GATA2 functions in lung cancer (Kumar et al. 2012).

GATA3 regulation of the TGF β /BMP pathways

From the preliminary data, we have found that GATA3 binds to many members of the TGF β /BMP pathways. For example, regions near *BMP4*, *BMP5*, *BMP7*, *NOG*, *SMAD3*, *SMAD6*, *SMAD7*, *TGFB2*, *TGFBR2* and *TGFBR3* are all occupied by GATA3 (**Table 5.1**). This is particularly interesting because the TGF β pathway has been shown to induce and activate the epithelial-to-mesenchymal transition (EMT), which allows cancer cells to de-differentiate and acquire stem-like characteristics (Mani et al. 2008). Since GATA3 inhibits EMT and promotes differentiation (see **Chapter 3**), we hypothesized that GATA3

suppresses signaling through the TGF β pathway. We have validated by real-time quantitative PCR (qPCR) that GATA3 regulates many of the genes bound by GATA3, including *BMP4*, *BMP5*, *NOG*, *SMAD3*, *SMAD6*, *TGFB2*, *TGFBR2*, and *TGFBR3*; this has revealed a suppression of TGF β pathway components and an up-regulation of BMP pathway components, consistent with the opposing roles of these two pathways (Oshimori and Fuchs 2012) (**Fig. 5.1A**). Furthermore, we have found that GATA3 inhibits TGF β -mediated EMT, and represses the activity of a TGF β /SBE-Luciferase reporter (SBE, Smad-binding element) (**Fig. 5.1B**). Future work will be aimed at testing whether suppression of TGF β activity is critical for GATA3-mediated metastasis suppression in vivo, for example by re-expressing *NOG* or knocking down *BMP4* or *BMP5* in the context of GATA3-over-expression.

We next wanted to determine whether any of the components identified were differentially expressed in human luminal and basal breast cancer cells. We expected that genes up-regulated by GATA3 would be expressed higher in luminal breast cancer cells, while genes down-regulated by GATA3 would be expressed higher in basal cells. Indeed, we found that *BMP4* and *BMP7* were significantly higher in luminal cells, while *SMAD3*, *TGFB2* and *TGFBR2* were significantly higher in basal cells; for some genes such as *BMP5*, we found no difference between the two cancer subtypes (**Fig. 5.1C–5.1H**). Taken together, our results identify TGF β /BMP signaling as an important aspect of luminal versus basal cell differentiation in breast cancer.

Candidate genes from the ChIP-seq are differentially regulated by GATA3

We next looked at the gene list from the ChIP-seq data set and chose interesting targets to validate by qPCR. Some genes were chosen because of the high peak values, which signifies that these sequences were consistently identified in the data set. Other genes were chosen because the genes appeared multiple times on the data set, which signifies that multiple regions within the same gene were bound by GATA3. Lastly, some genes were chosen based on previous literature and interest.

We examined the effects of GATA3 expression on genes involved in differentiation and stemness pathways (*ALDH1*, *CLDN4*, *ERBB4*, *FZD1*, *GLI3*, *HMGA2*, *JAG1*, and *ZEB2*), genes involved in immune function (*S1PR1*, *STAT3*, *TNFSF18*), and miscellaneous genes involved in extracellular proteolysis, MAPK signaling, etc (*ADAMTS9*, *DUSP2*, *EDN2*, *ETS1*, *LMO4*, and *NEDD9*) (**Fig. 5.2A–5.2C**). We then correlated the expression levels of three genes (*JAG1*, *DUSP2* and *NEDD9*) in luminal and basal breast cancer cells, and found that they were differentially expressed, in a direction consistent with the effects of GATA3 (**Fig. 5.2D–5.2F**). For example, *JAG1*, which was decreased in MDA231-GATA3 cells, was significantly lower in luminal cell lines (**Fig. 5.2D**). In addition, our data suggest that GATA3 might modulate MAPK signaling by activating expression of *DUSP2*. *DUSP2* might be an interesting target to explore because it has been implicated in mediating chemoresistance (Lin et al. 2011). This family of phosphatases attenuates MAPK signaling at multiple steps along the kinase signaling cascade, making them important regulators of proliferation and cancer

(Keyse 2008). Whether GATA3 might sensitize breast cancer cells to chemotherapy, and whether this might be mediated through a gene like *DUSP2*, is an exciting possibility that awaits further investigation.

Semaphorins as potential regulators of mammary development and luminal differentiation

One interesting aspect of the ChIP-seq data is the potential to identify novel regulators of mammary development and luminal differentiation. Strikingly, a large number of semaphorin genes seem to be bound by GATA3 (**Table 5.1**). The semaphorin family are key axon guidance molecules important during neuronal development (Neufeld and Kessler 2008). Recent work indicates that various semaphorins can either promote or inhibit tumor progression by regulating processes such as tumor angiogenesis, tumor metastasis and tumor cell survival (Neufeld and Kessler 2008), thus making them interesting molecules to study. However, little is known about the role of semaphorins during breast cancer progression and metastasis (Harburg and Hinck 2011). Interestingly, the SLIT/ROBO genes, which are another family of axon guidance mediators, have been recently shown to be involved in mammary branching morphogenesis by inhibiting WNT signaling (Macias et al. 2011).

We found that two semaphorin members, *SEMA3B* and *SEMA3F* are both enriched in luminal breast cancer cells compared to basal breast cancer cells (**Fig. 5.3A–5.3B**). Future work will be aimed at understanding whether semaphorin expression is induced by GATA3, whether semaphorins are

important for luminal cell differentiation, and whether semaphorins affect metastasis. This will be an exciting area for further investigation.

Preliminary validation of a subset of GATA3 targets in 4T1 cells

We have begun to validate some of the genes identified in MDA231 cells in 4T1 cells as well to determine whether the same genes are regulated in the same manner in these two similar yet genetically distinct basal breast cancer cells. Although we have also generated a GATA3 ChIP-seq data set in 4T1-Gata3 cells, the data from that data set is less comprehensive, likely due to technical reasons during the ChIP protocol. We have found that several genes are regulated in the same direction in 4T1-Gata3 cells as in MDA231-GATA3 cells, including *Fzd1*, *Nedd9*, *Nog*, *S1pr1*, *Smad6*, *Tgfr3* and *Zeb1* (**Fig. 5.4A**). Other genes such as *Tgfr2*, *Hmga2*, and *Dusp2* show no difference in 4T1-Control versus 4T1-Gata3 cells. Future work will be aimed to more extensively validate the hits, and to further explore the 4T1-Gata3 set.

Future directions

We are currently repeating the MDA231-GATA3 ChIP-seq to establish a second independent replicate of the data, and to identify other potential targets of interest, including microRNAs, which we have not yet investigated in the present analysis. We plan to continue validating these targets by expression analysis, and conduct in vivo mouse experiments to determine their contribution to breast cancer progression and to luminal differentiation. Additionally, we plan to

determine whether dysregulation of these targets occurs in human patient samples. We believe that our analysis will generate new targets to investigate and give insight to the biological function of GATA3 in mammary epithelial cells, which will contribute to our understanding of normal mammary gland development and cancer.

ACKNOWLEDGEMENTS

We thank Peggy Farnham, Henriette O'Geen, Seth Frieze, Heather Witt and Lijing Yao for helpful discussions and project ideas, as well as performing the ChIP-seq protocol and analysis. We thank Payam Shahi for helpful discussions about gene targets to pursue.

EXPERIMENTAL PROCEDURES

Cell culture and viral transduction

MDA-MB-231 and 4T1 were grown in standard conditions (DMEM with 10% FBS) and transduced with pMIG-*hGATA3* retrovirus (MDA-MB-2331) or pMIG-*mGata3* retrovirus (4T1), which was generated in HEK 293T cells using standard transfection procedure with calcium phosphate. Cells were sorted by FACS using a FACS Aria II (Becton Dickinson) (UCSF Stem Cell FACS Core) and then expanded in culture. Over-expression of GATA3 was confirmed both by qPCR and by western blot.

ChIP-seq cell and data processing

The ChIP-seq analysis was conducted as described from Fujiwara et al. 2009. Briefly, MDA231-GATA3 and 4T1-Gata3 cells were crosslinked in PBS containing 1% formaldehyde for 10 min at room temperature. The reaction was stopped by adding glycine to a final concentration of 0.125 M, and cell pellets were flash frozen in liquid nitrogen and stored at -80°C . Chromatin was prepared as described (<http://www.genomecenter.ucdavis.edu/farnham/protocol.html>) and was fragmented with a Bioruptor (Diagenode) to ~400 bp. ChIP samples were prepared using $\sim 10^8$ cells, and samples were diluted in buffer containing 50 mM Tris pH 7.4, 150 mM NaCl, 1% (v/v) igepal, 0.25% (w/v) deoxycholic acid, 1 mM EDTA pH 8 plus protease inhibitor cocktail (Sigma). The mouse anti-GATA3 antibody to perform the ChIP was from Santa Cruz Biotech.

ChIP libraries were created using 15 cycles of amplification and were

resolved on a 2% agarose gel, and the 150–450 bp fraction was extracted and purified. ChIP-seq libraries were analyzed on an Illumina GA2 at the University of California, Davis. Sequence reads were provided as the data output format. Sequence reads obtained were aligned to the UCSC human genome assembly HG19 or the UCSC mouse genome assembly mm9 using the Eland application (Illumina). Only sequences that aligned uniquely were used to identify GATA3 occupancy peaks. Peaks were called using the Sole-search program (<http://chipseq.genomecenter.ucdavis.edu/cgi-bin/chipseq.cgi>).

Quantitative real-time PCR (qPCR)

Total RNA was isolated from cells using the miRNeasy Mini Kit (Qiagen). cDNA was synthesized using the Superscript III RT First Strand Kit (Invitrogen). qPCR was performed using FastStart Universal SYBR Green master mix (Roche) in an Eppendorf Mastercycler realplex machine. Ct values were normalized to actin and GAPDH, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. Primer sequences for qPCR were found using the Harvard Primer Bank.

Luciferase assay

The SBE4-Luc plasmid was used to assess activity of the TGF β pathway in MDA231-Control and MDA231-GATA3 cells after stimulation with 5 ng/mL of TGF β 1, as previously described (see **Chapter 3**)

FIGURE LEGENDS

Figure 5.1. GATA3 regulates components of the TGF β /BMP pathways.

(A) TGF β /BMP genes identified by ChIP-seq are differentially expressed in MDA231-GATA3 cells, as measured by qPCR.

(B) GATA3 inhibits the TGF β pathway in MDA231 cells, as measured by a TGF β -luciferase reporter (SBE-Luc).

(C–H) Expression of *BMP4* (C), *BMP5* (D), *BMP7* (E), *SMAD3* (F), *TGFB2* (G) and *TGFBR2* (H) in luminal and basal breast cancer cells. Dataset adapted from Neve et al. 2006.

Figure 5.2. GATA3 regulates genes involved in differentiation, inflammation, proteolysis and cell signaling.

(A–C) Differentiation and stemness genes (A), miscellaneous genes involved in diverse functions such as proteolysis and cell signaling (B), and immune system mediators (C) identified by ChIP-seq are differentially expressed in MDA231-GATA3 cells, as measured by qPCR.

(D–F) Expression of *JAG1* (D), *DUSP2* (E), and *NEDD9* (F) in luminal and basal breast cancer cells. Dataset adapted from Neve et al. 2006.

Figure 5.3. Semaphorins, a family of axonal guidance genes identified by GATA3 ChIP-seq, are enriched in luminal breast cancer cells.

(A–B) Expression of *SEMA3B* (A) and *SEMA3F* (B) in luminal and basal breast cancer cells. Dataset adapted from Neve et al. 2006.

Figure 5.4. Initial validation of hits in murine 4T1 breast cancer cells expressing Gata3.

(A) Genes previously identified in the MDA231-GATA3 ChIP-seq and shown to be differentially expressed in MDA231-GATA3 cells were assessed in murine 4T1-Gata3 cells. Expression was measured by qPCR.

FIGURES

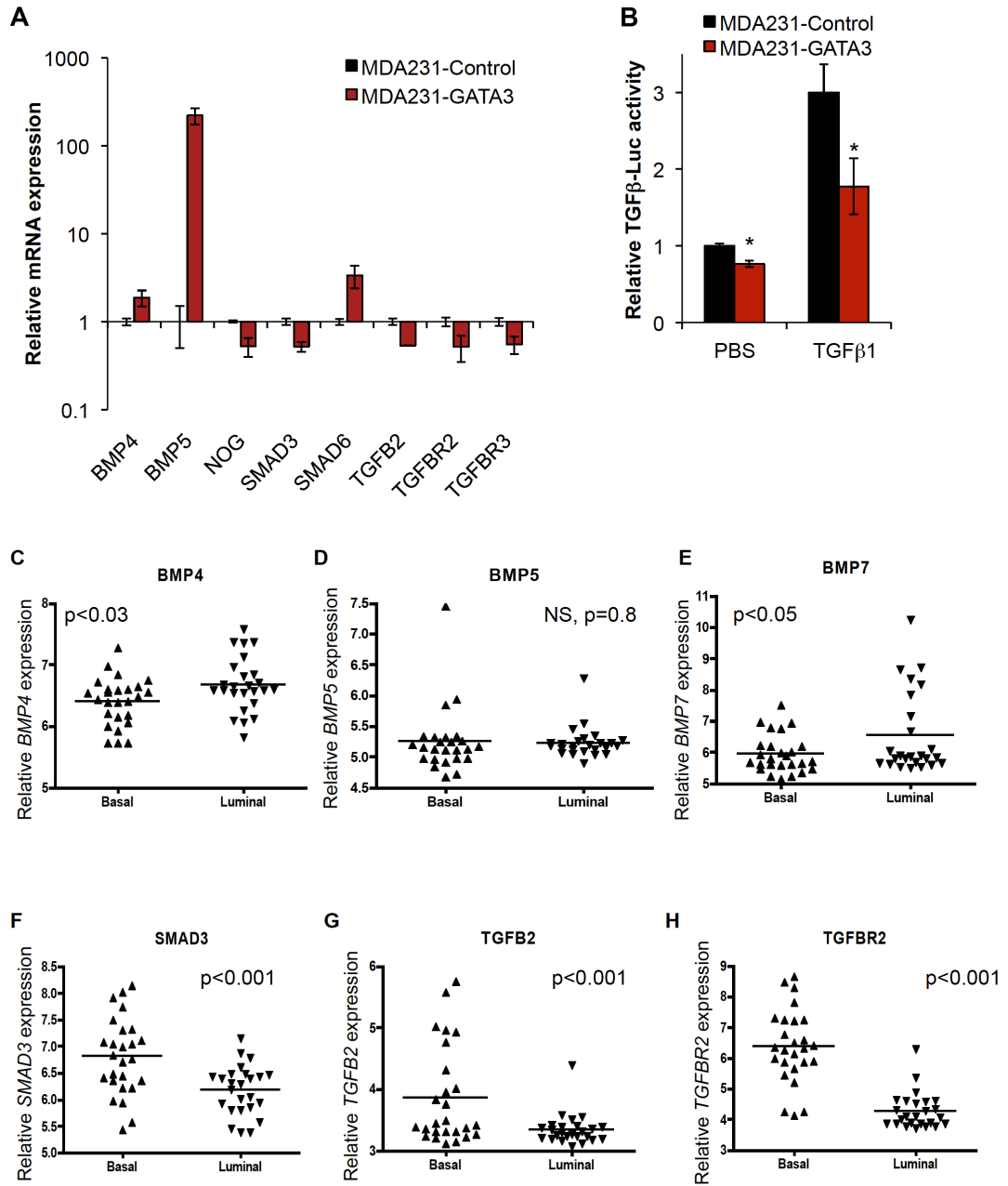


Figure 5.1

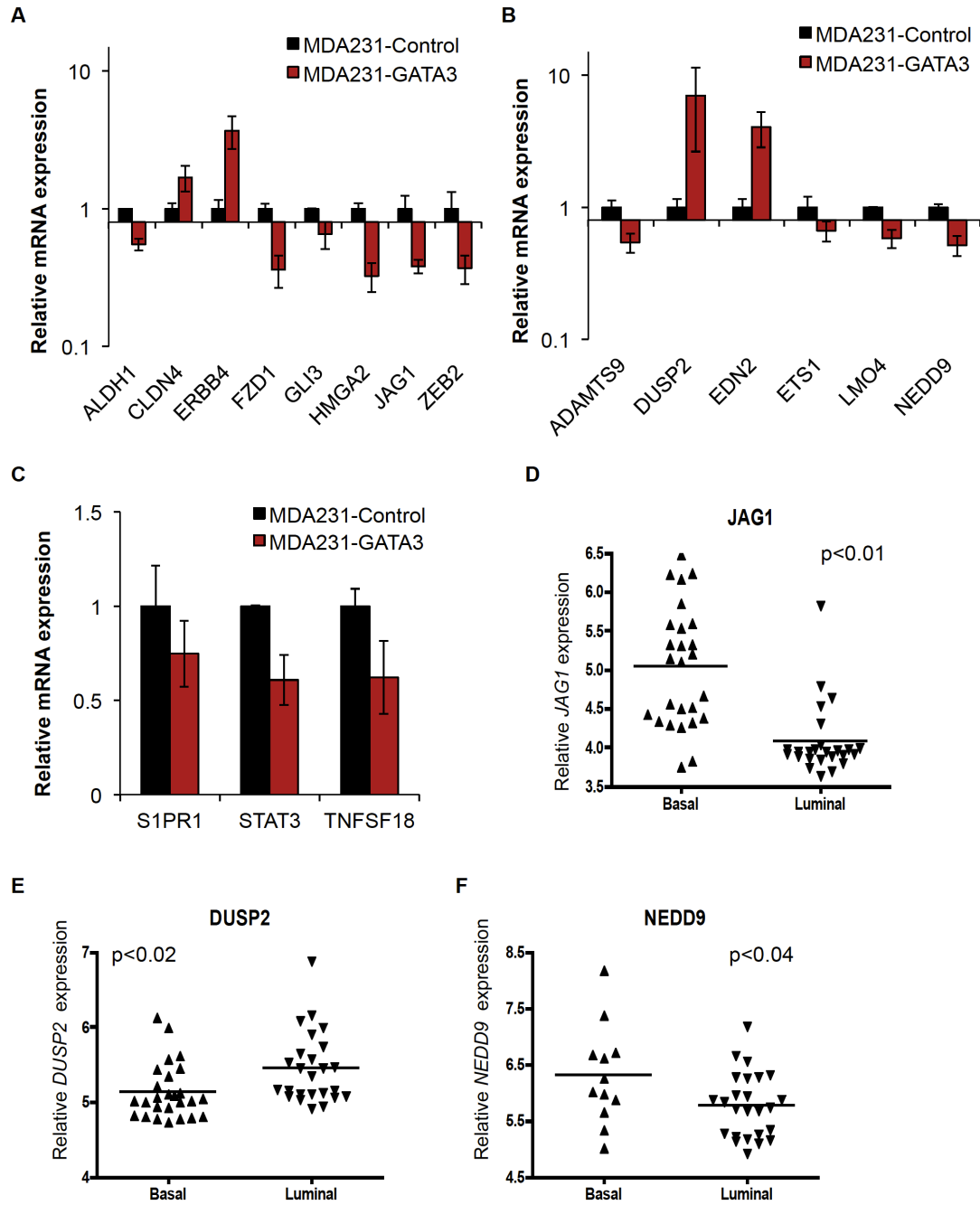


Figure 5.2

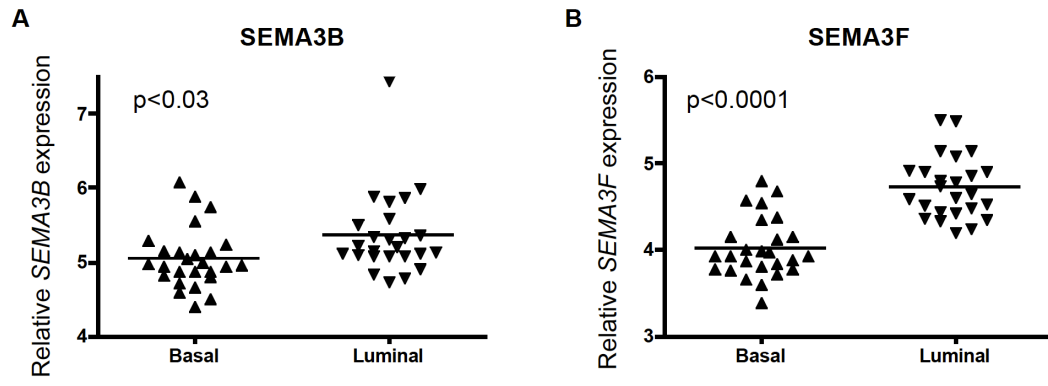


Figure 5.3

A

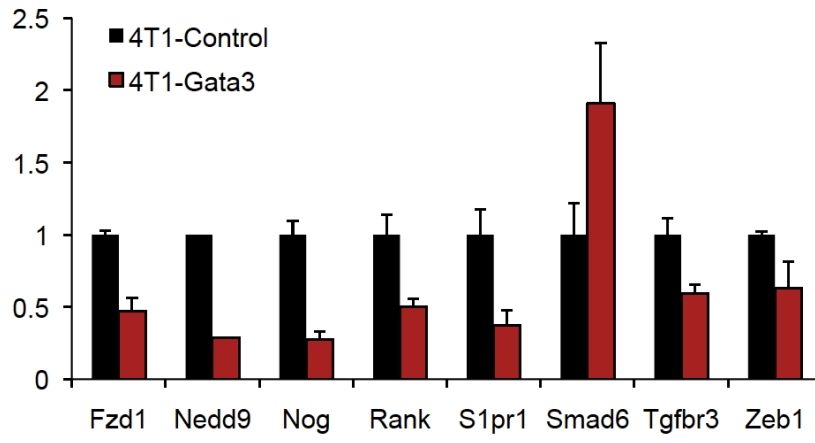


Figure 5.4

TABLES

Table 5.1. GATA3 ChIP-seq in human MDA231-GATA3 breast cancer cells (O'Geen Dataset-1)

GeneID	chromosome	strand	distance	peak start	peak end	peak value
AAGAB	chr15	-	335	67546530	67546949	31
AAK1	chr2	-	1128	69869700	69869999	22
AASS	chr7	-	-17515	121801680	121802039	28
AASS	chr7	-	-22465	121806660	121806959	16
ABCA12	chr2	-	-138424	216035100	216035369	25
ABCA4	chr1	-	127221	94459290	94459679	36
ABCA4	chr1	-	75531	94510980	94511369	36
ABCA5	chr17	-	45985	67265610	67265969	28
ABCB5	chr7	+	81925	20737020	20737319	22
ABCC12	chr16	-	19372	48161130	48161489	17
ABCC13	chr21	+	30440	15676350	15676769	27
ABCC3	chr17	+	25782	48737820	48738179	19
ABCC4	chr13	-	179168	95774340	95774699	50
ABCC4	chr13	-	200903	95752650	95752919	20
ABCD3	chr1	+	-11768	94872000	94872329	20
ABHD2	chr15	+	13524	89644770	89645039	16
ABHD5	chr3	+	80105	43812270	43812689	38
ABHD5	chr3	+	175595	43907790	43908149	24
ABL1	chr9	+	133317	133722390	133722779	24
ABLIM1	chr10	-	-128524	116415060	116415359	27
ABLIM3	chr5	+	9111	148530030	148530299	17
ABR	chr17	-	26300	985890	986159	18
ABRA	chr8	-	-346482	108128700	108129209	60
ABRA	chr8	-	-8427	107790690	107791109	42
ABRA	chr8	-	-242667	108024990	108025289	17
ABTB2	chr11	-	25248	34353390	34353719	22
ACAA2	chr18	-	89877	47250120	47250629	27
ACAA2	chr18	-	87897	47252190	47252519	21
ACACA	chr17	-	23143	35633370	35633729	23
ACBD6	chr1	-	-50022	180521880	180522209	21
ACCN1	chr17	-	5857	31614000	31614299	24
ACOX3	chr4	-	85188	8357100	8357429	24
ACPL2	chr3	+	43378	140993850	140994269	55
ACSL5	chr10	+	6094	114139860	114140159	17
ACSS3	chr12	+	65221	81536880	81537179	20
ACTBL2	chr5	-	99502	56678940	56679329	28
ACTC1	chr15	-	-14847	35102580	35102969	26
ACTL7B	chr9	-	585741	111032340	111032729	44
ACTR1B	chr2	-	952	98279460	98279759	16
ADAM19	chr5	-	3994	156998640	156998909	18
ADAM28	chr8	+	-71735	24079650	24080039	26
ADAMTS12	chr5	-	249705	33642270	33642569	22
ADAMTS18	chr16	-	134652	77334210	77334509	27
ADAMTS18	chr16	-	136437	77332290	77332859	24
ADAMTS9	chr3	-	-343324	65016510	65016869	33

ADAMTSL1	chr9	+	184861	18658740	18659189	50
ADAMTSL2	chr9	+	26139	136423260	136423589	35
ADAMTSL3	chr15	+	81817	84404490	84404819	26
ADAMTSL3	chr15	+	42007	84364710	84364979	22
ADCY5	chr3	-	116363	123050820	123051239	31
ADK	chr10	+	13545	75924300	75924719	42
ADO	chr10	+	1829	64566180	64566509	27
ADORA1	chr1	+	26144	203122770	203123189	30
ADRA1D	chr20	-	-157795	4387290	4387619	25
ADRB2	chr5	+	86469	148292490	148292759	18
AFAP1	chr4	-	15744	7925760	7926059	19
AFARP1	chr1	+	-53867	113411970	113412239	19
AFF1	chr4	+	51932	87979920	87980249	24
AGAP1	chr2	+	76299	236478870	236479199	24
AGBL4	chr1	-	495887	49993560	49993919	27
AGPAT4	chr6	-	-68147	161763060	161763449	32
AGR3	chr7	-	-108156	17029560	17029979	47
AGR3	chr7	-	34419	16886970	16887419	39
AHI1	chr6	-	150714	135668010	135668369	27
AHNAK	chr11	-	38443	62275710	62276069	31
AHR	chr7	+	149729	17487690	17488319	24
AHR	chr7	+	6359	17344500	17344769	20
AIPL1	chr17	-	53865	6284430	6284879	47
AK3L1	chr1	+	30068	65643090	65643509	34
AKAP10	chr17	-	19375	19861440	19862069	20
AKAP6	chr14	+	228356	33026700	33026969	16
AKT3	chr1	-	329109	243677220	243677669	59
ALAS1	chr3	+	794	52232730	52233089	23
ALCAM	chr3	+	35202	105120780	105121049	24
ALCAM	chr3	+	-1334448	103751070	103751459	24
ALCAM	chr3	+	147132	105232680	105233009	22
ALDH1A1	chr9	-	-27665	75595470	75595799	25
ALG10B	chr12	+	-461937	38248410	38248829	21
ALG14	chr1	-	78778	95459550	95459909	52
ALG14	chr1	-	-11747	95550060	95550449	33
ALG14	chr1	-	37183	95501100	95501549	19
ALG14	chr1	-	70288	95468070	95468369	18
ALPK2	chr18	-	17345	56278680	56279009	20
AMIGO2	chr12	-	80815	47392740	47393099	28
AMN1	chr12	-	-20266	31902150	31902599	44
ANAPC13	chr3	-	7857	134196750	134197199	44
ANGPT1	chr8	-	60315	108449790	108450089	21
ANGPT1	chr8	-	150	108509940	108510269	21
ANGPT1	chr8	-	-134235	108644340	108644639	20
ANK1	chr8	-	-77610	41600280	41600549	19
ANK3	chr10	-	-273923	62174430	62174909	75
ANK3	chr10	-	-243863	62144430	62144789	29
ANK3	chr10	-	-257003	62157570	62157929	24
ANKFN1	chr17	+	288104	54518790	54519089	55
ANKFN1	chr17	+	166349	54397050	54397319	24
ANKH	chr5	-	61893	14809830	14810159	18

ANKH	chr5	-	-149757	15021510	15021779	18
ANKRD1	chr10	-	-9997	92690790	92691269	62
ANKRD1	chr10	-	-85222	92766120	92766389	23
ANKRD2	chr10	+	2544	99334620	99334979	35
ANKRD26P1	chr16	-	-4105	46606980	46607249	17
ANKRD28	chr3	-	173494	15727320	15727799	64
ANKRD44	chr2	-	-95682	198158310	198158579	18
ANKRD50	chr4	-	432078	125199720	125199989	19
ANKS1B	chr12	-	454448	99923820	99924149	21
ANO6	chr12	+	139990	45749580	45749939	46
ANO6	chr12	+	44380	45653940	45654359	45
ANO6	chr12	+	68290	45677880	45678239	36
ANO6	chr12	+	16090	45625710	45626009	28
ANP32C	chr4	-	107029	165011700	165011969	15
ANXA11	chr10	-	41629	81923490	81923909	32
ANXA13	chr8	-	56583	124692930	124693199	24
ANXA2	chr15	-	-2154	60692100	60692579	54
ANXA2	chr15	-	35136	60654870	60655229	35
ANXA2P3	chr10	+	516150	67101300	67101569	20
ANXA2P3	chr10	+	1035960	67621110	67621379	19
AP1B1	chr22	-	44340	29740050	29740409	42
AP1B1	chr22	-	45855	29738520	29738909	36
AP1B1	chr22	-	53550	29730870	29731169	18
AP1S3	chr2	-	35515	224666670	224666939	18
AP2S1	chr19	-	9164	47344890	47345189	22
AP2S1	chr19	-	-16186	47370150	47370629	20
AP2S1	chr19	-	659	47353410	47353679	19
AP3S1	chr5	+	16036	115193460	115193849	22
APBB2	chr4	-	142076	41074380	41074739	42
APBB2	chr4	-	57956	41158500	41158859	30
APBB2	chr4	-	108431	41108070	41108339	26
APBB2	chr4	-	37481	41178990	41179319	19
APC	chr5	+	-12133	112030950	112031219	23
APCDD1L	chr20	-	-35795	57125580	57125909	15
APOB	chr2	-	-595989	21862800	21863069	23
APOLD1	chr12	+	33014	12911730	12911999	18
AQP9	chr15	+	83842	58514070	58514429	33
ARAP2	chr4	-	425665	35820120	35820509	37
ARAP2	chr4	-	232810	36013020	36013319	22
ARAP2	chr4	-	239890	36005850	36006329	21
ARAP3	chr5	-	-31814	141093390	141093839	45
ARAP3	chr5	-	-32789	141094380	141094799	20
AREG	chr4	+	98986	75579480	75579749	23
ARHGAP10	chr4	+	-17553	148635690	148636109	50
ARHGAP12	chr10	-	921	32216700	32216999	23
ARHGAP15	chr2	+	302116	144188820	144189209	36
ARHGAP18	chr6	-	42451	129988740	129989099	35
ARHGAP20	chr11	-	-267358	110850600	110851019	40
ARHGAP20	chr11	-	-181708	110764950	110765369	33
ARHGAP25	chr2	+	-13193	68948580	68948969	37
ARHGAP26	chr5	+	338083	142488120	142488629	57

ARHGAP26	chr5	+	55528	142205610	142206029	34
ARHGAP26	chr5	+	-30002	142120110	142120469	33
ARHGAP26	chr5	+	155248	142305390	142305689	31
ARHGAP26	chr5	+	-13907	142136250	142136519	21
ARHGAP26	chr5	+	452278	142602390	142602749	19
ARHGAP5	chr14	+	57655	32603970	32604329	60
ARHGDI8	chr12	-	-67	15114450	15114809	21
ARHGEF11	chr1	-	86948	156928020	156928409	21
ARHGEF12	chr11	+	130499	120338250	120338639	31
ARHGEF12	chr11	+	7274	120215070	120215369	21
ARHGEF12	chr11	+	-5416	120202350	120202709	18
ARHGEF3	chr3	-	-94989	56904570	56904899	17
ARID3A	chr19	+	9318	935160	935549	35
ARID4B	chr1	-	81518	235409760	235410269	21
ARID5B	chr10	+	104232	63765540	63765809	25
ARID5B	chr10	+	145017	63806280	63806639	24
ARID5B	chr10	+	180882	63842190	63842459	17
ARL15	chr5	-	-19556	53625630	53626289	60
ARL4A	chr7	+	264374	12991140	12991529	49
ARL4A	chr7	+	5684	12732510	12732779	21
ARMC5	chr16	+	-7247	31462920	31463219	26
ARNTL	chr11	+	-124330	13174860	13175129	28
ARPP-21	chr3	+	287006	35970720	35970989	18
ARRDC3	chr5	-	-404630	91083540	91084019	75
ART3	chr4	+	79268	77011380	77011829	39
ART4	chr12	-	7094	14989140	14989499	28
ASAP1	chr8	-	-81422	131495400	131495879	31
ASAP1	chr8	-	77788	131336250	131336609	17
ASAP2	chr2	+	28151	9374880	9375209	21
ASAP3	chr1	-	32031	23778510	23778929	46
ASB4	chr7	+	14851	95129970	95130299	24
ASCC1	chr10	-	44700	73930770	73931219	58
ASCC2	chr22	-	38895	30195210	30195509	29
ASPH	chr8	-	-19940	62646900	62647379	52
ASPH	chr8	-	-91100	62718120	62718479	20
ASPH	chr8	-	-7940	62634960	62635319	19
ATAD2B	chr2	-	532165	23617560	23617979	34
ATF3	chr1	+	30918	212769480	212769749	39
ATG16L1	chr2	+	-5987	234154080	234154379	21
ATG7	chr3	+	237385	11551200	11551589	45
ATL2	chr2	-	-18947	38623200	38623559	27
ATP1A1	chr1	+	-86255	116829390	116829689	23
ATP1B1	chr1	+	-45537	169030140	169030679	72
ATP2C2	chr16	+	11507	84413490	84413789	19
ATP6AP1L	chr5	+	91789	81692790	81693119	23
ATP6AP1L	chr5	+	7879	81608880	81609209	22
ATP6V0E1	chr5	+	707	172411290	172411649	27
ATP6V1H	chr8	-	115093	54640260	54640649	35
ATP8B1	chr18	-	-62635	55461540	55461809	22
ATXN1	chr6	-	-148783	16910310	16910699	47
ATXN1	chr6	-	-8098	16769670	16769969	35

ATXN1	chr6	-	416387	16345200	16345469	16
ATXN10	chr22	+	5473	46073010	46073309	28
ATXN7L1	chr7	-	45972	105470850	105471269	23
AZIN1	chr8	-	52848	103823280	103823819	82
B3GAT1	chr11	-	-456767	134738370	134738789	49
B4GALT4	chr3	-	5868	118953720	118954049	22
BAD	chr11	-	14252	64037730	64038119	46
BAD	chr11	-	15902	64036080	64036469	28
BAGE	chr21	-	-73332	11172090	11172449	58
BAHCC1	chr17	+	-5675	79367670	79368059	61
BAMBI	chr10	+	96661	29062950	29063219	21
BARHL2	chr1	-	107235	91075410	91075709	20
BARHL2	chr1	-	76785	91105860	91106159	19
BARHL2	chr1	-	158850	91023750	91024139	17
BARHL2	chr1	-	229290	90953370	90953639	14
BARX2	chr11	+	-128716	129117000	129117329	27
BASP1	chr5	+	143940	17361480	17361899	47
BASP1	chr5	+	141630	17359200	17359559	25
BASP1	chr5	+	477465	17695050	17695379	25
BASP1	chr5	+	35040	17252610	17252969	20
BCAR3	chr1	-	9636	94137540	94137959	38
BCAR3	chr1	-	-31134	94178340	94178699	22
BCAR3	chr1	-	52371	94094850	94095179	21
BCAR3	chr1	-	16446	94130790	94131089	15
BCAT1	chr12	-	88519	25013640	25013939	19
BCAT1	chr12	-	2344	25099830	25100099	18
BCL2L1	chr20	-	44062	30266400	30266789	33
BCL2L14	chr12	+	24222	12247860	12248339	59
BCL3	chr19	+	-13568	45238230	45238589	36
BDNF	chr11	-	-44068	27725100	27725429	23
BET1	chr7	-	-60509	93694020	93694379	25
BEYLA	chr8	+	-907043	46845300	46845629	21
BEYLA	chr8	+	-908828	46843500	46843859	20
BEYLA	chr8	+	-849338	46902960	46903379	20
BFSP2	chr3	+	53205	133171830	133172159	21
BHLHE41	chr12	-	-3271	26281110	26281439	26
BIRC6	chr2	+	62074	32644020	32644319	22
BLK	chr8	+	22019	11373330	11373749	29
BLM	chr15	+	57981	91318350	91318769	58
BMP4	chr14	-	63746	54357270	54357779	83
BMP4	chr14	-	-162394	54583230	54584099	49
BMP4	chr14	-	16106	54404970	54405359	37
BMP4	chr14	-	163316	54257700	54258209	35
BMP4	chr14	-	313616	54107490	54107819	31
BMP4	chr14	-	-154624	54575610	54576179	28
BMP4	chr14	-	102701	54318390	54318749	25
BMP4	chr14	-	71576	54349470	54349919	19
BMP5	chr6	-	-7229	55747380	55747829	30
BMP5	chr6	-	-7769	55747950	55748339	26
BMP5	chr6	-	46351	55693890	55694159	21
BMP7	chr20	-	190358	55651080	55651619	79

BMP7	chr20	-	259643	55581810	55582319	58
BMP7	chr20	-	235868	55605660	55606019	26
BMPER	chr7	+	74438	34019370	34019729	25
BMPR1B	chr4	+	-66833	95612160	95612429	17
BNC1	chr15	-	-1666	83954910	83955359	62
BOD1	chr5	-	-120038	173163570	173163839	19
BOD1	chr5	-	-248	173043780	173044049	18
BOD1L	chr4	-	-484171	14113320	14113679	40
BRAP	chr12	-	40476	112083120	112083509	57
BRD3	chr9	-	1757	136931250	136931519	25
BRD7P3	chr6	+	-113166	118709190	118709549	15
BRE	chr2	+	378793	28492080	28492469	37
BRE	chr2	+	87058	28200360	28200719	29
BSG	chr19	+	895	572070	572369	23
BTBD3	chr20	+	-347112	11524140	11524589	60
BTC	chr4	-	25323	75694410	75694709	21
BTD	chr3	+	29195	15672270	15672629	28
BTD	chr3	+	32570	15675690	15675959	22
BTG1	chr12	-	368169	92171280	92171729	57
BTG1	chr12	-	159489	92379960	92380409	57
BTG1	chr12	-	366804	92172690	92173049	43
BTG1	chr12	-	4569	92534880	92535329	38
BTG1	chr12	-	225534	92313930	92314349	36
BTG1	chr12	-	259854	92279670	92279969	20
BZRAP1	chr17	-	-7242	56413260	56413529	22
C10orf11	chr10	+	569771	78112140	78112439	15
C10orf114	chr10	-	156784	21629250	21629609	26
C10orf129	chr10	+	35563	96989280	96989759	30
C10orf26	chr10	+	21073	104524650	104524949	18
C10orf81	chr10	+	-5198	115505880	115506149	24
C10orf90	chr10	-	62821	128147040	128147339	23
C11orf70	chr11	+	-34299	101883750	101884049	41
C11orf75	chr11	-	44257	93232110	93232469	22
C12orf12	chr12	-	146404	91202280	91202819	40
C12orf12	chr12	-	58699	91290060	91290449	30
C12orf45	chr12	+	-26773	105353130	105353519	50
C12orf75	chr12	+	223901	105948120	105948509	28
C14orf101	chr14	+	-227906	56818350	56818859	32
C14orf128	chr14	-	65911	32479740	32480249	57
C14orf182	chr14	-	7924	50466030	50466599	60
C14orf182	chr14	-	-6191	50480190	50480669	40
C14orf33	chr14	-	38836	56007810	56008139	20
C14orf34	chr14	-	-51632	56314890	56315159	15
C14orf4	chr14	-	-18170	77513010	77513399	29
C14orf4	chr14	-	69295	77425560	77425919	27
C14orf43	chr14	-	-25428	74252250	74252609	41
C15orf39	chr15	+	-30196	75463830	75464219	32
C15orf39	chr15	+	-56611	75437460	75437759	24
C15orf54	chr15	+	50245	39592920	39593339	37
C15orf54	chr15	+	-148145	39394590	39394889	25
C16orf87	chr16	-	20930	46843920	46844369	32

C17orf82	chr17	+	20433	59509380	59509709	28
C18orf1	chr18	+	217314	13435830	13436369	87
C1orf106	chr1	+	23366	200887170	200887559	39
C1orf107	chr1	+	46007	210047190	210047489	18
C1orf110	chr1	-	-5249	162843690	162844019	23
C1orf130	chr1	+	3088	24885450	24885929	59
C1orf144	chr1	+	8857	16702290	16702589	27
C1orf173	chr1	-	20608	75118620	75119009	32
C1orf185	chr1	+	16824	51584580	51584879	23
C1orf203	chr1	-	-68065	117029130	117029489	37
C1orf52	chr1	-	5591	85719600	85719929	17
C1orf62	chr1	-	26452	109373040	109373489	36
C1orf9	chr1	+	-2965	172499130	172499459	32
C1orf9	chr1	+	28205	172530330	172530599	17
C1QTNF7	chr4	+	26825	15368220	15368549	18
C20orf94	chr20	+	187429	10603230	10603529	27
C21orf109	chr21	+	2654	30568320	30568589	20
C21orf63	chr21	+	2463	33787080	33787349	28
C2CD2	chr21	-	-30905	43377450	43377959	55
C2CD2	chr21	-	-39740	43386360	43386719	21
C2CD2	chr21	-	-6290	43352880	43353299	16
C2orf34	chr2	+	187322	44776140	44776589	33
C2orf34	chr2	+	191852	44780730	44781059	23
C3orf16	chr3	-	8731	149501700	149502059	23
C3orf17	chr3	-	-91729	112830090	112830479	31
C3orf26	chr3	+	7714	99544260	99544529	41
C3orf50	chr3	+	-274834	168252600	168252899	19
C3orf67	chr3	-	-614529	59650080	59650409	24
C3orf77	chr3	+	-193548	44089650	44090009	18
C4orf10	chr4	+	532	2937660	2937959	19
C4orf18	chr4	-	33209	159060360	159060659	26
C4orf45	chr4	-	-12471	159968670	159968939	21
C5orf27	chr5	+	9699	95197230	95198039	37
C5orf27	chr5	+	10599	95198340	95198729	26
C5orf36	chr5	-	399880	93554160	93554699	49
C5orf41	chr5	+	-4415	172478610	172479299	127
C5orf41	chr5	+	37870	172521030	172521449	33
C5orf56	chr5	+	-14183	131732340	131732639	26
C6orf138	chr6	-	-498564	48534840	48535139	17
C6orf15	chr6	-	25563	31054590	31054949	18
C6orf155	chr6	-	-69441	72199710	72200069	26
C6orf167	chr6	-	-39802	97770720	97770989	21
C6orf168	chr6	-	5322	99792000	99792419	23
C6orf168	chr6	-	72657	99724710	99725039	22
C6orf173	chr6	+	-164603	126496500	126496799	16
C6orf191	chr6	-	17577	130164690	130164989	22
C6orf223	chr6	+	57996	44026170	44026499	51
C6orf89	chr6	+	26890	36880350	36880709	30
C7orf10	chr7	+	85200	40259580	40259969	46
C7orf10	chr7	+	598380	40772700	40773209	36
C7orf10	chr7	+	358740	40533180	40533449	21

C7orf53	chr7	+	1122	112121820	112122239	39
C7orf58	chr7	+	58399	120686940	120687359	26
C7orf65	chr7	+	-1187	47693520	47693789	28
C8orf37	chr8	-	43673	96237570	96237959	52
C8orf4	chr8	+	21791	40032570	40032989	34
C8orf46	chr8	+	24339	67429650	67430009	26
C8orf74	chr8	+	8478	10538400	10538849	41
C8orf75	chr8	-	-4434	29609910	29610209	27
C8orf79	chr8	+	56182	12859170	12859559	34
C8orf83	chr8	-	970	93977220	93977519	22
C9orf106	chr9	+	-75	132083040	132083399	39
C9orf122	chr9	+	61050	38681940	38682329	36
C9orf3	chr9	+	144716	97633500	97633919	29
C9orf5	chr9	-	54481	111827520	111827969	32
C9orf84	chr9	-	30729	114490920	114491249	28
CA12	chr15	-	36741	63637140	63637529	23
CA14	chr1	+	-1753	150228300	150228629	22
CACNA1D	chr3	+	23759	53552580	53552999	45
CACNG4	chr17	+	-21228	64939590	64939979	53
CACNG6	chr19	+	10723	54506070	54506459	30
CALCOCO1	chr12	-	-93327	54214500	54214769	20
CALCR	chr7	-	36171	93167400	93167759	19
CALM2	chr2	-	-32454	47436000	47436389	28
CALM2	chr2	-	-74469	47478030	47478389	26
CALM2	chr2	-	-22479	47426070	47426369	24
CAPN2	chr1	+	5130	223894290	223894559	33
CAPSL	chr5	-	19198	35919360	35919719	30
CAPZA1	chr1	+	18205	113180100	113180459	26
CARD18	chr11	-	-91653	105101280	105101639	25
CASD1	chr7	+	2000	94140990	94141349	23
CASP12	chr11	-	45683	104723580	104723849	20
CASP12	chr11	-	229313	104539950	104540219	17
CASP3	chr4	-	75620	185494800	185495219	49
CASQ2	chr1	-	39197	116272080	116272379	26
CAST	chr5	+	-4871	95992920	95993219	16
CAV2	chr7	+	-32919	116106390	116106659	15
CBFB	chr16	+	23835	67086690	67087079	31
CBLB	chr3	-	62483	105525240	105525569	26
CBR4	chr4	-	-70027	170001240	170001659	27
CBS	chr21	-	11276	44484630	44484899	22
CBX7	chr22	-	-25856	39574260	39574529	19
CCDC109A	chr10	+	35171	74486790	74487329	28
CCDC112	chr5	-	-70162	114701940	114702359	47
CCDC12	chr3	-	29316	46988700	46989209	61
CCDC12	chr3	-	28431	46989630	46990049	29
CCDC12	chr3	-	13596	47004540	47004809	18
CCDC132	chr7	+	74882	92936400	92936669	14
CCDC14	chr3	-	142536	123537540	123537899	29
CCDC140	chr2	+	-223356	222939330	222939689	39
CCDC66	chr3	+	-37734	56553180	56553719	39
CCDC83	chr11	+	-27264	85538670	85539089	45

CCDC83	chr11	+	-10524	85555440	85555799	25
CCDC83	chr11	+	-44934	85521030	85521389	22
CCDC85A	chr2	+	497002	56908050	56908469	30
CCDC93	chr2	-	55788	118715730	118716089	28
CCKAR	chr4	-	-32277	26524140	26524499	20
CCL17	chr16	+	13331	57451830	57452189	33
CCNA2	chr4	-	1679	122743230	122743589	22
CCND1	chr11	+	-190328	69265380	69265709	42
CCNG2	chr4	+	-10587	78067560	78067979	23
CCR1	chr3	-	44463	46205130	46205609	32
CCR3	chr3	+	43092	46326600	46326959	32
CCR5	chr3	+	16232	46427730	46427999	23
CCRN4L	chr4	+	-117788	139818900	139819409	37
CD163L1	chr12	-	10425	7586130	7586519	19
CD180	chr5	-	-10272	66502680	66503099	47
CD36	chr7	+	9661	80241000	80241329	17
CD38	chr4	+	-23736	15756060	15756329	21
CD44	chr11	+	-27072	35133150	35133539	42
CD44	chr11	+	-106857	35053380	35053739	25
CD44	chr11	+	97098	35257380	35257649	18
CD47	chr3	-	91636	107718150	107718449	23
CD55	chr1	+	-80097	207414540	207414899	33
CD55	chr1	+	-4782	207489900	207490169	29
CD58	chr1	-	62116	117051330	117051869	23
CD59	chr11	-	-71	33744030	33744659	33
CD69	chr12	-	-32132	9945420	9945839	28
CD96	chr3	+	40184	111300840	111301379	46
CD96	chr3	+	-220201	111040560	111040889	25
CD96	chr3	+	-148801	111111990	111112259	21
CDC2	chr10	+	-114490	62423520	62423939	27
CDC2	chr10	+	-250675	62287410	62287679	19
CDC42EP1	chr22	+	3354	37959600	37960049	24
CDC42EP3	chr2	-	-40968	37940130	37940459	26
CDC42EP3	chr2	-	-123963	38023080	38023499	26
CDC42EP3	chr2	-	85722	37813440	37813769	23
CDC42EP3	chr2	-	19092	37880100	37880369	20
CDC42EP4	chr17	-	3649	71304240	71304749	48
CDC42SE2	chr5	+	7008	130606470	130606949	50
CDC5L	chr6	+	59218	44414370	44414669	21
CDGAP	chr3	+	17135	119030160	119030549	20
CDH1	chr16	+	-4700	68766300	68766689	14
CDH10	chr5	-	-144238	24788970	24789329	23
CDH10	chr5	-	-610558	25255290	25255649	22
CDH13	chr16	+	28512	82688850	82689329	24
CDH13	chr16	+	31047	82691490	82691759	22
CDH17	chr8	-	-3999	95224590	95225039	41
CDH18	chr5	-	100290	19880850	19881149	15
CDH2	chr18	-	-979754	26737020	26737379	39
CDH23	chr10	+	372561	73529070	73529459	26
CDH26	chr20	+	-3152	58530180	58530479	17
CDH6	chr5	+	-1333656	29859990	29860289	23

CDH9	chr5	-	-203410	27241920	27242279	29
CDH9	chr5	-	-1504495	28543020	28543349	26
CDK6	chr7	-	200527	92262510	92262899	35
CDK6	chr7	-	200902	92262180	92262479	16
CDKN1A	chr6	+	3776	36650040	36650429	26
CDKN3	chr14	+	-65973	54797520	54797879	30
CDKN3	chr14	+	-144858	54718440	54719189	27
CDT1	chr16	+	-32641	88837350	88837739	29
CDYL2	chr16	-	184241	80653680	80654189	44
CDYL2	chr16	-	189206	80648820	80649119	23
CEACAM16	chr19	+	23092	45225270	45225629	45
CEBPG	chr19	+	-24834	33839580	33839969	22
CENPE	chr4	-	47572	104071800	104072189	23
CEP110	chr9	+	53566	123903990	123904289	21
CEP350	chr1	+	-23673	179900010	179900459	27
CFH	chr1	+	26742	196647480	196648019	59
CFL2	chr14	-	13034	35169810	35170169	17
CFTR	chr7	+	185698	117305550	117305879	29
CFTR	chr7	+	102613	117222450	117222809	22
CFTR	chr7	+	108433	117228270	117228629	20
CGA	chr6	-	-10050	87814710	87815039	26
CHAC2	chr2	+	-363614	53631090	53631539	32
CHAC2	chr2	+	-1327469	52667250	52667669	27
CHCHD6	chr3	+	49292	126472260	126472559	28
CHD1	chr5	-	-675541	98937540	98938019	31
CHD1	chr5	-	14639	98247450	98247749	20
CHD2	chr15	+	-30811	93412590	93412889	21
CHD3	chr17	+	2832	7790730	7791179	54
CHD5	chr1	-	22099	6217920	6218249	26
CHL1	chr3	+	157665	396180	396449	22
CHRD	chr3	+	68944	184166640	184166969	28
CHRM3	chr1	+	-64773	239727450	239727749	23
CHRM3	chr1	+	-12303	239779920	239780219	19
CHST9	chr18	-	-41455	24806610	24806879	21
CHSY1	chr15	-	57012	101734920	101735309	30
CHSY3	chr5	+	610307	129850680	129850979	25
CIRH1A	chr16	+	1171	69167490	69167849	20
CISD1	chr10	+	11455	60040080	60040619	50
CIT	chr12	-	-24897	120339750	120340229	59
CITED2	chr6	-	-687729	140383110	140383919	48
CITED2	chr6	-	-334449	140030070	140030399	35
CITED2	chr6	-	-213879	139909470	139909859	33
CITED2	chr6	-	-197499	139893120	139893449	27
CITED2	chr6	-	-209019	139904640	139904969	27
CKAP2L	chr2	-	58310	113463720	113464169	45
CLASP1	chr2	-	87798	122319120	122319389	23
CLCA3P	chr1	+	-37889	87061920	87062219	18
CLCA4	chr1	+	-30534	86982000	86982449	26
CLDN14	chr21	-	-11869	37850370	37850819	36
CLDN4	chr7	+	-8233	73236720	73237199	41
CLEC12A	chr12	+	8087	10131960	10132229	17

CLEC14A	chr14	-	-400830	39126270	39126539	15
CLEC16A	chr16	+	13565	11051760	11052059	21
CLIC5	chr6	-	170133	45813330	45813599	17
CLIC6	chr21	+	-28338	36013140	36013559	37
CLIC6	chr21	+	-28743	36012780	36013109	24
CLIP1	chr12	-	1257	122905710	122906009	16
CLIP2	chr7	+	21990	73725600	73725989	54
CLIP4	chr2	+	2607	29340720	29341109	33
CLIP4	chr2	+	3837	29342010	29342279	18
CLN8	chr8	+	1955	1713630	1714019	36
CLPB	chr11	-	-141	72145560	72145859	27
CMAH	chr6	-	101597	25036320	25036589	18
CMIP	chr16	+	141400	81620040	81620309	19
CMTM8	chr3	+	10944	32290950	32291279	29
CMTM8	chr3	+	116034	32396040	32396369	25
CMTM8	chr3	+	19164	32299200	32299469	18
CNIH	chr14	-	5764	54902130	54902639	44
CNN3	chr1	-	-53149	95445720	95446049	22
COL11A1	chr1	-	-141947	103715790	103716209	42
COL11A1	chr1	-	-48737	103622580	103622999	30
COL11A1	chr1	-	-54092	103627980	103628309	29
COL11A1	chr1	-	210898	103363020	103363289	19
COL12A1	chr6	-	22689	75892800	75893069	19
COL1A2	chr7	+	-143388	93880260	93880709	25
COL24A1	chr1	-	-38278	86660190	86660609	28
COL3A1	chr2	+	27796	189866760	189867029	16
COL4A2	chr13	+	91909	111051330	111051749	21
COL6A3	chr2	-	-7634	238330350	238330619	21
COLEC12	chr18	-	1020	499530	499889	25
COMMD10	chr5	+	112378	115532970	115533239	19
COMTD1	chr10	-	-58209	77053710	77054249	29
COPB1	chr11	-	63955	14457270	14457629	30
COPB1	chr11	-	55975	14465190	14465669	29
COPG	chr3	+	-23083	128945190	128945549	37
COPS8	chr2	+	92996	238086810	238087349	87
CORO2B	chr15	+	-50193	68821170	68821589	39
COTL1	chr16	-	16165	84635370	84635639	23
COX7A2L	chr2	-	18552	42569670	42569939	24
COX7B2	chr4	-	-77	46911150	46911509	28
CP	chr3	-	-55752	148995450	148995719	19
CPA3	chr3	+	32922	148615830	148616099	15
CPB2	chr13	-	-15203	46694250	46694579	41
CPEB1	chr15	-	19959	83220420	83220809	44
CPEB3	chr10	-	52025	93998670	93998969	36
CPEB4	chr5	+	-62916	173252160	173252669	63
CPEB4	chr5	+	-65691	173249430	173249849	50
CPLX2	chr5	+	-17295	175206150	175206479	25
CPLX2	chr5	+	13350	175236810	175237109	24
CPSF4L	chr17	-	6115	71251770	71252039	18
CPSF6	chr12	+	-141392	69491760	69492089	25
CRADD	chr12	+	-52231	94018680	94019159	41

CRADD	chr12	+	107279	94178220	94178639	35
CREBBP	chr16	-	82412	3847560	3847859	27
CREBBP	chr16	-	54677	3875310	3875579	22
CRHR1	chr17	+	44089	43905600	43905869	15
CRIM1	chr2	+	47608	36630780	36631229	63
CRIM1	chr2	+	-894332	35688840	35689289	48
CRIM1	chr2	+	194023	36777270	36777569	20
CRIM1	chr2	+	-807512	35775750	35776019	20
CRIP1	chr14	+	-952	105952110	105952499	38
CRISP2	chr6	-	36670	49644450	49644809	27
CRISPLD2	chr16	+	64743	84918150	84918509	26
CRTC1	chr19	+	80660	18874830	18875339	40
CRY1	chr12	-	66254	107421210	107421479	19
CRYGC	chr2	-	-400	208994730	208995179	25
CRYL1	chr13	-	123443	20976390	20976749	23
CRYZL1	chr21	-	2856	35011140	35011469	18
CSMD3	chr8	-	1266088	113123070	113123519	56
CSNK1G3	chr5	+	188462	123036120	123036389	20
CSNK1G3	chr5	+	439472	123287040	123287489	19
CSNK1G3	chr5	+	307727	123155370	123155669	16
CSRP1	chr1	-	16323	201459870	201460259	26
CTAGE1	chr18	-	75074	19922610	19922999	44
CTNNA1	chr5	+	147308	138236190	138236639	64
CTNNA1	chr5	+	43298	138132240	138132569	36
CTNNA1	chr5	+	158198	138247110	138247499	21
CTNNA1	chr5	+	96188	138185160	138185429	18
CTNNA3	chr10	-	918107	68507130	68507489	15
CTNNB1	chr3	+	-213947	41026860	41027129	16
CTR9	chr11	+	-8256	10764360	10764749	31
CTSC	chr11	-	-14533	88085340	88085609	19
CTTNBP2	chr7	-	45427	117467940	117468329	32
CUBN	chr10	-	126582	17045070	17045399	18
CUGBP1	chr11	-	-34218	47544630	47544959	18
CUL2	chr10	-	84310	35294700	35295269	59
CUX1	chr7	+	41208	101500290	101500709	45
CXADR	chr21	+	14385	18899580	18899849	22
CXXC5	chr5	+	60624	139088760	139089089	20
CYB561	chr17	-	7408	61510590	61511009	71
CYB5B	chr16	+	52762	69511050	69511469	35
CYB5D1	chr17	+	296	7761210	7761509	45
CYFIP2	chr5	+	119159	156812070	156812429	24
CYP24A1	chr20	-	39432	52750950	52751219	15
CYP27B1	chr12	-	577	58160190	58160609	42
CYR61	chr1	+	-35889	86010270	86010839	96
CYR61	chr1	+	-26529	86019720	86020109	22
CYTH3	chr7	-	-12267	6324180	6324839	32
D2HGDH	chr2	+	6635	242680530	242680799	22
DAAM1	chr14	+	153591	59808810	59809169	40
DAAM1	chr14	+	172701	59827950	59828249	23
DAB2	chr5	-	-88204	39513300	39513779	41
DAB2IP	chr9	+	40806	124369860	124370549	38

DAB2IP	chr9	+	183726	124512990	124513259	24
DAD1	chr14	-	32694	23025240	23025659	37
DAOA	chr13	+	308220	106426650	106426919	21
DAP	chr5	-	69193	10692060	10692329	23
DCAF13	chr8	+	-17837	104408970	104409239	26
DCDC2	chr6	-	44526	24313530	24313979	39
DCP1A	chr3	-	-3167	53384580	53385029	30
DCPS	chr11	+	38288	126211770	126212099	35
DCTD	chr4	-	-58384	183896820	183897209	32
DDAH1	chr1	-	136020	85794600	85795139	69
DDAH1	chr1	-	144150	85786530	85786949	52
DDAH1	chr1	-	134640	85796040	85796459	31
DDAH1	chr1	-	-30480	85961190	85961549	26
DDAH1	chr1	-	121185	85809570	85809839	21
DDHD1	chr14	-	-304938	53924820	53925149	16
DDIT4	chr10	+	17333	74050830	74051189	26
DDIT4L	chr4	-	91684	101019780	101020079	24
DDR2	chr1	+	30757	162632850	162633119	18
DDX10	chr11	+	308474	108844110	108844469	38
DDX47	chr12	+	10115	12976230	12976559	26
DDX58	chr9	-	578	32525610	32525879	21
DDX60	chr4	-	99959	169139850	169140149	20
DEGS1	chr1	+	21737	224392470	224392859	47
DGKB	chr7	-	-357244	15238080	15238559	25
DGKB	chr7	-	-282544	15163470	15163769	16
DGKD	chr2	+	75742	234338700	234339089	34
DHRS1	chr14	-	1717	24766770	24767129	22
DHRS3	chr1	-	43966	12633720	12633989	17
DHX8	chr17	+	-14469	41546670	41547059	22
DIO2	chr14	-	208199	80469450	80469839	36
DIP2B	chr12	+	85017	50983650	50983919	25
DISC1	chr1	+	92904	231855300	231855629	28
DKK1	chr10	+	37694	54111540	54111929	39
DKK1	chr10	+	227099	54300930	54301349	30
DKK1	chr10	+	-4066	54069840	54070109	22
DLAT	chr11	+	18052	111913410	111913769	23
DLEU7	chr13	-	330406	51087330	51087629	21
DLG1	chr3	-	108733	196916520	196916909	38
DLGAP1	chr18	-	251567	3593490	3593969	50
DLGAP1	chr18	-	240542	3604620	3604889	20
DLX3	chr17	-	-6341	48078720	48079139	33
DMXL2	chr15	-	128233	51786570	51786899	15
DNAH11	chr7	+	74032	21656700	21657029	27
DNAH11	chr7	+	216547	21799200	21799559	24
DNAH14	chr1	+	324289	225441450	225441839	23
DNAH14	chr1	+	271954	225389130	225389489	17
DNAH2	chr17	+	62721	7685580	7685939	22
DNAH5	chr5	-	-94555	14038950	14039339	18
DNAH5	chr5	-	376415	13568040	13568309	17
DNAH9	chr17	+	346347	11847960	11848229	20
DNAJB6	chr7	+	68159	157197630	157198109	81

DNAJC11	chr1	-	13437	6748380	6748679	19
DNAJC5B	chr8	+	-47561	66886050	66886409	25
DNASE1L3	chr3	-	-7626	58207890	58208159	18
DNMBP	chr10	-	43457	101725890	101726549	56
DNPEP	chr2	-	19198	220233330	220233599	17
DNTTIP2	chr1	-	74603	94269900	94270379	60
DOCK10	chr2	-	112211	225794970	225795269	23
DOCK2	chr5	+	419804	169483920	169484189	18
DOCK4	chr7	-	14023	111832230	111832649	50
DOCK4	chr7	-	224188	111622080	111622469	25
DOCK5	chr8	+	134958	25177050	25177439	38
DOCK5	chr8	+	-5112	25037010	25037339	25
DPH5	chr1	-	943	101490210	101490629	18
DPYD	chr1	-	826616	97559700	97560299	87
DPYD	chr1	-	-17134	98403450	98404049	85
DPYD	chr1	-	-84469	98470800	98471369	64
DPYD	chr1	-	-68194	98454570	98455049	46
DPYD	chr1	-	-161044	98547450	98547869	38
DPYD	chr1	-	-21619	98407980	98408489	34
DPYD	chr1	-	-98854	98485260	98485679	30
DPYD	chr1	-	7376	98379060	98379419	25
DPYD	chr1	-	67301	98319150	98319479	20
DPYD	chr1	-	-106249	98492730	98492999	19
DPYD	chr1	-	-261649	98648100	98648429	19
DPYD	chr1	-	-103054	98489520	98489819	18
DPYD	chr1	-	776381	97610100	97610369	17
DPYD	chr1	-	-177424	98563890	98564189	17
DST	chr6	-	34384	56673870	56674289	21
DTWD2	chr5	-	1098716	117225390	117225659	21
DUSP10	chr1	-	-159967	222070620	222070919	23
DUSP2	chr2	-	-1760	96812730	96813149	56
DUSP4	chr8	-	-55539	29263560	29263889	25
DUSP6	chr12	-	-3473	89749590	89749949	36
DUSP6	chr12	-	-18608	89764740	89765069	26
DUSP6	chr12	-	234712	89511450	89511719	17
DYDC2	chr10	+	21597	82137960	82138349	36
DYNC111	chr7	+	78852	95480520	95480819	18
DYNLRB1	chr20	+	9166	33113190	33113549	21
DYNLRB2	chr16	+	-241484	80333160	80333579	25
DYNLRB2	chr16	+	50296	80625000	80625299	23
DYNLT1	chr6	-	-1099	159066630	159067049	22
DYX1C1	chr15	-	22813	55777470	55777769	25
E2F1	chr20	-	9346	32264430	32265299	26
E2F1	chr20	-	6601	32267430	32267789	23
E2F1	chr20	-	8131	32265930	32266229	15
E2F2	chr1	-	-8861	23866440	23866709	28
E2F7	chr12	-	-37359	77496540	77496899	30
E2F7	chr12	-	-161409	77620560	77620979	24
ECH1	chr19	-	-3692	39326010	39326369	38
EDEM1	chr3	+	7231	5236380	5236799	56
EDIL3	chr5	-	420382	83260050	83260409	24

EDN1	chr6	+	-14934	12275400	12275789	24
EDN1	chr6	+	-9789	12280590	12280889	20
EDN2	chr1	-	-1220	41951340	41951789	58
EDN2	chr1	-	48925	41901240	41901599	22
EDN2	chr1	-	41590	41908620	41908889	20
EEA1	chr12	-	-158197	93481080	93481529	55
EEA1	chr12	-	-151702	93474600	93475019	17
EED	chr11	+	-84070	85871580	85871909	33
EED	chr11	+	-71725	85883910	85884269	22
EEF1E1	chr6	-	19224	8083470	8083739	21
EEPD1	chr7	+	-115721	36076920	36077309	42
EEPD1	chr7	+	-104981	36087720	36087989	15
EFCAB1	chr8	-	314761	49332960	49333259	24
EFCAB4B	chr12	-	111407	3750810	3751109	29
EFEMP1	chr2	-	81017	56069160	56069519	19
EFHB	chr3	-	45567	19929900	19930379	59
EFHB	chr3	-	2142	19973400	19973729	30
EFHD2	chr1	+	2374	15738600	15738929	28
EFHD2	chr1	+	7774	15744000	15744329	23
EFNA5	chr5	-	246902	106759500	106759889	45
EFNA5	chr5	-	244952	106761180	106762109	20
EFNA5	chr5	-	113852	106892550	106892939	18
EFNB2	chr13	-	250448	106936710	106937069	26
EGFR	chr7	+	54910	55141440	55141829	34
EGFR	chr7	+	-22805	55063710	55064129	32
EGFR	chr7	+	-49265	55037220	55037699	29
EGLN3	chr14	-	-187370	34607400	34607909	44
EGLN3	chr14	-	-8960	34429110	34429379	23
EGLN3	chr14	-	-248585	34668690	34669049	20
EHBP1	chr2	+	13952	62914710	62915219	57
EHD4	chr15	-	44271	42220320	42220649	21
EIF2B3	chr1	-	16128	45435960	45436349	39
EIF2B3	chr1	-	91878	45360240	45360569	22
EIF2C2	chr8	-	-779	141646230	141646619	37
EIF2C2	chr8	-	-18239	141663750	141664019	23
EIF2C3	chr1	+	93863	36490500	36490769	19
EIF3H	chr8	-	401238	117366690	117366959	18
EIF3IP1	chr7	-	-311089	109911210	109911509	21
EIF4E	chr4	-	100169	99749940	99750209	19
ELK3	chr12	+	25383	96613440	96613739	19
ELL2	chr5	-	-106079	95403630	95404079	35
ELOVL2	chr6	-	10445	11034030	11034329	24
ELTD1	chr1	-	173046	79299270	79299629	25
EMID2	chr7	+	39428	101045370	101045729	36
EMID2	chr7	+	79793	101085780	101086049	19
EML4	chr2	+	-36160	42360090	42360569	87
EML4	chr2	+	-45145	42351120	42351569	46
EML4	chr2	+	70190	42466500	42466859	28
EML6	chr2	+	228396	55180410	55180679	21
ENC1	chr5	-	324945	73612110	73612499	56
ENC1	chr5	-	439785	73497270	73497659	31

ENC1	chr5	-	65565	73871520	73871849	26
ENDOD1	chr11	+	60698	94883520	94883909	24
ENOX1	chr13	-	383844	43819590	43819949	20
ENOX1	chr13	-	92289	44111190	44111459	15
ENPP1	chr6	+	15384	132144300	132144779	47
ENPP2	chr8	-	103427	120547500	120547859	35
ENPP2	chr8	-	64532	120586440	120586709	29
ENPP2	chr8	-	52217	120598650	120599129	18
ENPP2	chr8	-	-2938	120653910	120654179	15
ENPP6	chr4	-	20845	185118090	185118449	36
ENPP6	chr4	-	52360	185086560	185086949	31
EPB41L4A	chr5	-	89491	111665280	111665759	24
EPB41L4A	chr5	-	208696	111546180	111546449	18
EPC1	chr10	-	48944	32586990	32587349	43
EPHA2	chr1	-	4465	16477920	16478279	34
EPHA4	chr2	-	952516	221484330	221484659	21
EPM2A	chr6	-	4647	146052180	146052509	23
EPM2A	chr6	-	238632	145818210	145818509	20
EPS8	chr12	-	106471	15835800	15836279	49
EPS8	chr12	-	153301	15789000	15789419	33
EPS8L3	chr1	-	-32925	110339340	110339639	20
ERAP1	chr5	-	-15857	96159570	96159929	36
ERBB4	chr2	-	-304317	213707520	213707819	22
ERBB4	chr2	-	-305472	213708660	213708989	19
ERC2	chr3	-	677162	55824900	55825559	114
ERC2	chr3	-	570602	55931580	55931999	47
ERC2	chr3	-	483632	56018580	56018939	33
ERC2	chr3	-	556937	55945290	55945619	23
EREG	chr4	+	29915	75260610	75260939	27
ERGIC1	chr5	+	41767	172302810	172303169	27
ERI1	chr8	+	72441	8932620	8932889	17
ERMP1	chr9	-	14942	5817930	5818349	32
ERN1	chr17	-	71968	62135370	62135699	25
ERRFI1	chr1	-	11189	8075040	8075369	30
ESRRG	chr1	-	96576	216800040	216800399	29
ETNK1	chr12	+	459669	23237580	23237909	25
ETS1	chr11	-	185811	128206170	128206619	59
ETS1	chr11	-	114066	128277990	128278289	21
ETS2	chr21	+	33926	40211580	40211969	38
ETV1	chr7	-	244200	13781730	13782149	40
ETV6	chr12	+	-15683	11786940	11787269	35
ETV6	chr12	+	-93953	11708640	11709029	33
EVI5	chr1	-	253417	93004380	93004709	23
EXOC5	chr14	-	11463	57723990	57724319	24
EXOC6B	chr2	-	68698	72984330	72984629	27
EXOC6B	chr2	-	136348	72916680	72916979	19
EXOC6B	chr2	-	263383	72789660	72789929	14
EXT1	chr8	-	275444	118848420	118848809	31
EXT1	chr8	-	107054	119016780	119017229	26
EXT1	chr8	-	153944	118969980	118970249	21
EXT1	chr8	-	264644	118859280	118859549	20

EXT1	chr8	-	200639	118923270	118923569	20
EXT1	chr8	-	-8926	119132850	119133119	19
EYA2	chr20	+	60087	45583140	45583559	45
EYA2	chr20	+	61107	45584190	45584549	25
EYA2	chr20	+	76497	45599580	45599939	25
EYA4	chr6	+	-70610	133491690	133492079	36
F2RL1	chr5	+	-4623	76110000	76110419	39
F3	chr1	-	-104763	95111820	95112449	93
F3	chr1	-	-74823	95081880	95082509	43
F3	chr1	-	-39078	95046270	95046629	35
F3	chr1	-	-1608	95008800	95009159	30
F3	chr1	-	-90183	95097360	95097749	22
F3	chr1	-	-127473	95134650	95135039	14
F5	chr1	-	56985	169498620	169498949	23
FADD	chr11	+	30416	70079490	70079879	35
FAF2	chr5	+	26359	175901520	175901909	30
FAF2	chr5	+	-9221	175865940	175866329	30
FAIM2	chr12	-	34626	50262870	50263319	62
FAM100B	chr17	+	1509	74262540	74263049	16
FAM102A	chr9	-	-51156	130764000	130764299	22
FAM10A4	chr13	+	147736	50893680	50894099	41
FAM113B	chr12	+	5483	47615370	47615699	31
FAM114A2	chr5	-	23443	153394920	153395189	21
FAM114A2	chr5	-	222193	153196170	153196439	17
FAM135B	chr8	-	1195431	138313410	138313859	51
FAM13C	chr10	-	-151162	61273380	61273649	19
FAM169A	chr5	-	-50259	74212740	74213009	23
FAM170A	chr5	+	249391	119214510	119214779	27
FAM172A	chr5	-	322410	93124800	93125189	24
FAM172A	chr5	-	291360	93155910	93156179	18
FAM174A	chr5	+	156301	100027230	100027619	22
FAM176A	chr2	-	57788	75730170	75730439	28
FAM177B	chr1	+	47792	222958170	222958529	60
FAM179B	chr14	+	106334	45537570	45537929	17
FAM188A	chr10	-	-18735	15921090	15921419	27
FAM188B	chr7	+	33017	30843840	30844259	33
FAM190A	chr4	+	528041	91576590	91576859	18
FAM190B	chr10	+	-19565	86068650	86069039	31
FAM32A	chr19	+	-145	16295910	16296269	43
FAM38B	chr18	-	-343790	11041410	11041799	31
FAM38B	chr18	-	-417110	11114760	11115089	24
FAM38B	chr18	-	-225560	10923240	10923509	15
FAM3C	chr7	-	-148472	121184700	121185089	32
FAM40B	chr7	+	61686	129135720	129136199	60
FAM46C	chr1	+	12531	118160970	118161299	23
FAM49B	chr8	-	-93144	131045010	131045279	31
FAM49B	chr8	-	-57024	131008860	131009189	26
FAM65B	chr6	-	-69519	24980520	24980909	33
FAM65B	chr6	-	-15414	24926460	24926759	19
FAM69A	chr1	-	-54765	93481650	93482039	34
FAM71A	chr1	+	12494	212810160	212810459	29

FAM71D	chr14	+	32254	67688190	67688609	48
FAM92B	chr16	-	-31925	85177890	85178189	26
FANCC	chr9	-	-65878	98145600	98146139	55
FANCL	chr2	-	-392774	58861140	58861439	29
FANCL	chr2	-	-490289	58958640	58958969	21
FAR2	chr12	+	19637	29396100	29396369	23
FAR2	chr12	+	-2473	29373960	29374289	21
FAR2	chr12	+	-289243	29087190	29087519	18
FARS2	chr6	+	236966	5498280	5498819	29
FASLG	chr1	+	87890	172715940	172716209	16
FAT1	chr4	-	-448547	188093340	188093729	27
FAT1	chr4	-	-512942	188157780	188158079	25
FBLN1	chr22	+	46371	45944910	45945269	24
FBLN1	chr22	+	94911	45993480	45993779	23
FBN1	chr15	-	230474	48707250	48707639	27
FBXL17	chr5	-	42655	107674950	107675339	35
FBXL2	chr3	+	41051	33359760	33360209	38
FBXO32	chr8	-	11519	124532730	124532999	24
FBXO43	chr8	-	1617	101156280	101156639	34
FCHO2	chr5	+	75162	72326820	72327119	27
FER	chr5	+	-129708	107953650	107953979	24
FEZ2	chr2	-	38688	36786510	36786779	18
FGD5	chr3	+	107446	14967720	14968109	36
FGD6	chr12	-	16211	95594820	95595239	50
FGD6	chr12	-	133391	95477580	95478119	27
FIGN	chr2	-	387649	164204670	164205059	41
FILIP1	chr6	-	43937	76159410	76159709	23
FLG	chr1	-	-3800	152301270	152301689	38
FLI1	chr11	+	39992	128603610	128603999	24
FLJ10661	chr8	+	60918	8146860	8147159	21
FLJ25363	chr3	+	297783	109426470	109426769	28
FLJ32065	chr17	-	-9741	62981310	62981579	23
FLJ32810	chr11	+	-15987	100542210	100542629	44
FLJ32810	chr11	+	32418	100590660	100590989	23
FLJ32810	chr11	+	4773	100563000	100563359	21
FLJ36031	chr7	-	-57228	106358610	106359029	35
FLJ37543	chr5	+	167454	61100880	61101299	45
FLJ42393	chr3	+	-117151	187778970	187779389	29
FLJ42709	chr5	-	444100	92462220	92462579	23
FLJ45079	chr17	-	8805	75871200	75871529	26
FLNB	chr3	+	-8837	57985080	57985499	42
FLOT1	chr6	-	4959	30705330	30705659	26
FLRT2	chr14	+	-249798	85746540	85746839	22
FMN1	chr15	-	265421	33094500	33094829	18
FMNL2	chr2	+	67914	153259440	153259889	33
FMOD	chr1	-	-59520	203379570	203380049	40
FBNP1	chr9	-	51154	132754110	132754529	36
FBNP1L	chr1	+	107121	94020720	94021199	34
FBNP1L	chr1	+	23811	93937440	93937859	25
FNDC3A	chr13	+	86512	49636410	49636709	22
FNDC3B	chr3	+	90697	171847950	171848279	27

FNIP2	chr4	+	105238	159795240	159795599	28
FNIP2	chr4	+	118873	159808890	159809219	24
FOLR4	chr11	+	39277	94077900	94078259	20
FOS	chr14	+	-1556	75743760	75744089	19
FOSL2	chr2	+	-26019	28589610	28589909	23
FOXB1	chr15	+	-62406	60233880	60234149	19
FOXG1	chr14	+	-383967	28852140	28852499	27
FOXJ3	chr1	-	108369	42692400	42692669	23
FOXJ3	chr1	-	112734	42688020	42688319	22
FOXL1	chr16	+	243555	86855370	86855969	116
FOXL1	chr16	+	5835	86617800	86618099	16
FOXL2	chr3	-	25288	138640500	138640889	34
FOXN2	chr2	+	29630	48571200	48571649	42
FOXP1	chr3	-	139256	71493660	71494109	38
FOXP1	chr3	-	780341	70852650	70852949	23
FOXP1	chr3	-	3986	71629020	71629289	18
FOXP4	chr6	+	-86489	41427540	41427809	25
FOXP4	chr6	+	-47099	41466930	41467199	18
FOXQ1	chr6	+	13340	1325850	1326179	32
FREM2	chr13	+	-30143	39230850	39231209	33
FRG1	chr4	+	-319479	190542360	190542629	16
FRG1	chr4	+	-297324	190564500	190564799	15
FRMD3	chr9	-	37779	86115390	86115749	31
FRMD3	chr9	-	106644	86046570	86046839	18
FRMD4A	chr10	-	448547	13924110	13924529	33
FRMD4A	chr10	-	436697	13936020	13936319	28
FRMD4A	chr10	-	345887	14026830	14027129	27
FRMD4A	chr10	-	406382	13966350	13966619	17
FRMD4B	chr3	-	-47794	69483060	69483389	29
FRMD4B	chr3	-	31166	69404040	69404489	22
FRMD5	chr15	-	37315	44449950	44450279	22
FRMD5	chr15	-	110185	44377080	44377409	17
FRMD6	chr14	+	45755	52001460	52001759	20
FSIP1	chr15	-	61610	40013250	40013609	19
FSTL4	chr5	-	154424	132793620	132793979	30
FSTL4	chr5	-	397799	132550260	132550589	23
FSTL5	chr4	-	227172	162857790	162858239	63
FSTL5	chr4	-	364032	162720960	162721349	32
FTO	chr16	+	64725	53802450	53802749	19
FURIN	chr15	+	-7915	91403760	91404179	42
FYB	chr5	-	80468	39139020	39139379	32
FYN	chr6	-	-21122	112215600	112215899	26
FZD1	chr7	+	248692	91142250	91142699	47
FZD6	chr8	+	9910	104320680	104321339	88
G2E3	chr14	+	-274984	30753180	30753509	19
GAD1	chr2	+	27090	171700110	171700469	20
GADL1	chr3	-	-182726	31118730	31119029	27
GADL1	chr3	-	-271601	31207620	31207889	19
GALNT10	chr5	+	95075	153665100	153665639	37
GALNT10	chr5	+	118130	153688140	153688709	17
GALNT3	chr2	-	2319	166648230	166648739	37

GALNT6	chr12	-	7766	51777210	51777659	23
GALNTL2	chr3	+	-140339	16075710	16075979	23
GAPVD1	chr9	+	82429	128106390	128106689	21
GARNL4	chr17	+	156223	2855820	2856089	27
GAS2L3	chr12	+	18421	100985700	100986119	40
GATA3	chr10	+	776388	8872920	8873189	19
GBAP	chr1	-	2691	155194440	155194829	40
GBAS	chr7	+	27924	56060040	56060399	17
GBP2	chr1	-	-65	89591640	89592089	41
GBP4	chr1	-	-2171	89666580	89667029	20
GCLM	chr1	-	-67132	94441950	94442339	28
GCNT2	chr6	+	-2233	10519200	10519469	24
GCNT4	chr5	-	-26865	74353440	74353739	30
GCNT4	chr5	-	76470	74250090	74250419	29
GDF9	chr5	-	10353	132189990	132190259	18
GDPD4	chr11	-	42899	76955370	76955759	38
GDPD4	chr11	-	-12166	77010420	77010839	33
GDPD4	chr11	-	43769	76954560	76954829	21
GEM	chr8	-	-702	95275080	95275439	27
GFI1	chr1	-	-21678	92970840	92971229	32
GFOD1	chr6	-	114403	13373190	13373579	30
GJA1	chr6	+	377080	122133690	122133959	19
GLCCI1	chr7	+	-57163	7950990	7951529	66
GLI3	chr7	-	246004	42030420	42030809	48
GLI3	chr7	-	166789	42109620	42110039	23
GLI3	chr7	-	139054	42137400	42137729	21
GLI3	chr7	-	-48701	42325140	42325499	20
GLI3	chr7	-	131614	42144870	42145139	17
GLRA1	chr5	-	-218237	151522500	151522769	20
GLTSCR1	chr19	+	40257	48151560	48151859	24
GMDS	chr6	-	604067	1641600	1641959	30
GMDS	chr6	-	83687	2162010	2162309	21
GMFB	chr14	-	-1555	54957120	54957479	25
GNA14	chr9	-	-64692	80327760	80328089	25
GNA15	chr19	+	19074	3155040	3155489	67
GNAI1	chr7	+	-46880	79717110	79717409	27
GNAI2	chr3	+	10823	50284290	50284649	23
GNAL	chr18	+	-380321	11308650	11308979	21
GNAT3	chr7	-	-30667	80171700	80172119	37
GNAT3	chr7	-	30443	80110620	80110979	23
GNG12	chr1	-	-74172	68373150	68373479	19
GNG7	chr19	-	74657	2627910	2628269	21
GOLGA7	chr8	+	17884	41365830	41366099	22
GOLGA7	chr8	+	-49706	41298240	41298509	17
GPATCH1	chr19	+	29549	33601200	33601469	21
GPM6A	chr4	-	342613	176391270	176391869	115
GPR110	chr6	-	-64807	47074710	47075069	31
GPR116	chr6	-	33795	46855770	46856069	17
GPR116	chr6	-	-10695	46900260	46900559	17
GPR125	chr4	-	7503	22510020	22510319	16
GPR126	chr6	+	111664	142734540	142734899	22

GPR128	chr3	+	-26928	100301370	100301639	26
GPR151	chr5	-	-72413	145967880	145968299	26
GPR81	chr12	-	4410	123210510	123210929	56
GPR98	chr5	+	604458	90458880	90459269	31
GPSM2	chr1	+	612	109420050	109420379	23
GRAMD3	chr5	+	-8883	125686680	125687129	32
GRAMD3	chr5	+	6372	125701980	125702339	28
GRAMD3	chr5	+	5877	125701530	125701799	16
GRB14	chr2	-	23606	165454560	165454949	29
GREM2	chr1	-	-153652	240928980	240929249	22
GREM2	chr1	-	35828	240739470	240739799	14
GRHL2	chr8	+	-31538	102472950	102473309	28
GRHL2	chr8	+	-46223	102458250	102458639	24
GRHL3	chr1	+	-35396	24610290	24610679	40
GRHL3	chr1	+	-52571	24593160	24593459	20
GRIA2	chr4	+	50694	158192220	158192639	29
GRIA4	chr11	+	91270	105571920	105572219	21
GRID1	chr10	-	724821	87401280	87401579	17
GRIK2	chr6	+	1334120	103180830	103181219	33
GRINL1A	chr15	+	26349	58025070	58025429	26
GRLF1	chr19	+	37017	47458710	47459189	46
GRM1	chr6	+	383378	146731920	146732399	41
GSDMC	chr8	-	107165	130691790	130692149	32
GSDMC	chr8	-	79100	130719870	130720199	29
GSTM1L	chr3	-	35578	12264210	12264539	30
GTDC1	chr2	-	182811	144869070	144869429	30
GTF2A2	chr15	-	7773	59941740	59942189	47
GTF2IRD1	chr7	+	-1235	73866690	73867079	35
GUCY2C	chr12	-	53115	14796210	14796599	26
GULP1	chr2	+	-309835	188847420	188847689	22
HACE1	chr6	-	230020	105077580	105077969	32
HACE1	chr6	-	220945	105086700	105086999	22
HACE1	chr6	-	154690	105152940	105153269	19
HAND1	chr5	-	-105790	153963390	153963839	44
HAS2	chr8	-	-100489	122753970	122754269	48
HAS2AS	chr8	+	-126096	122525340	122525639	23
HDAC4	chr2	-	132564	240189900	240190259	27
HEATR7B2	chr5	-	58730	41012580	41012849	24
HECW2	chr2	-	332691	197124480	197124809	26
HECW2	chr2	-	91761	197365440	197365709	19
HECW2	chr2	-	5661	197451540	197451809	17
HEG1	chr3	-	-23892	124798470	124798919	54
HEG1	chr3	-	-7197	124781850	124782149	30
HELLS	chr10	+	3696	96309030	96309509	29
HERC1	chr15	-	-23302	64149270	64149629	27
HERC3	chr4	+	7898	89521410	89521679	17
HERC4	chr10	-	-11811	69846750	69847079	23
HES1	chr3	+	4671	193858470	193858739	27
HEY1	chr8	-	-40166	80720100	80720429	25
HGF	chr7	-	386133	81013110	81013529	25
HINT1	chr5	-	787	130499940	130500539	40

HIPK2	chr7	-	3202	139474230	139474559	30
HIPK2	chr7	-	120697	139356690	139357109	27
HIST1H3C	chr6	+	1636	26047140	26047409	22
HIST1H4H	chr6	-	-10607	26296170	26296499	23
HIVEP1	chr6	+	136136	12148560	12149159	41
HIVEP1	chr6	+	58976	12071550	12071849	17
HIVEP2	chr6	-	188674	143077410	143077919	91
HIVEP2	chr6	-	102304	143163810	143164259	34
HIVEP3	chr1	-	57092	42327180	42327629	24
HK2	chr2	+	18728	75078300	75078719	48
HMCN1	chr1	+	-172703	185530830	185531129	28
HMCN1	chr1	+	418567	186122040	186122459	24
HMCN1	chr1	+	-653	185702850	185703209	22
HMCN1	chr1	+	-89108	185614440	185614709	19
HMGA2	chr12	+	54355	66272370	66272819	56
HMGA2	chr12	+	29710	66247770	66248129	34
HMGA2	chr12	+	55765	66273810	66274199	33
HMGA2	chr12	+	-220370	65997690	65998049	30
HMGA2	chr12	+	123925	66342000	66342329	25
HMGXB3	chr5	+	17266	149397240	149397629	27
HMP19	chr5	+	326021	173798610	173798879	56
HNF1B	chr17	-	30052	36074910	36075179	20
HNRNPC	chr14	-	51179	21686280	21686639	27
HNRNPH1	chr5	-	1543	179049000	179049359	27
HNRNPL	chr19	-	12118	39328200	39328799	38
HOPX	chr4	-	-43201	57565740	57566039	25
HPGD	chr4	-	84748	175358880	175359209	20
HPSE2	chr10	-	571250	100424220	100424519	20
HRH1	chr3	+	101866	11280510	11280779	21
HS2ST1	chr1	+	178835	87558870	87559469	43
HS2ST1	chr1	+	78875	87459030	87459389	40
HS3ST1	chr4	-	-76727	11507100	11507429	32
HS3ST1	chr4	-	-184937	11615310	11615639	28
HS3ST1	chr4	-	-312617	11743020	11743289	23
HS3ST5	chr6	-	-588228	114972090	114972449	24
HSD17B12	chr11	+	109752	43811730	43812059	17
HSD3B1	chr1	+	-71	120049560	120049949	35
HSF2BP	chr21	-	4480	45074700	45075089	32
HSP90AB2P	chr4	+	-244282	13090620	13090889	21
HSPB1	chr7	+	-11220	75920520	75920789	19
HTR1B	chr6	-	170256	78002610	78003119	41
HTR1B	chr6	-	436881	77736060	77736419	24
HTR1B	chr6	-	447771	77725140	77725559	21
HTR1E	chr6	+	-581834	87064950	87065429	36
HTR1E	chr6	+	-213299	87433590	87433859	19
HTR2A	chr13	-	-223719	47693640	47694149	77
HTR2A	chr13	-	-279579	47749470	47750039	39
HTR2A	chr13	-	-315099	47785110	47785439	28
HTR2A	chr13	-	-395229	47865270	47865539	24
HTR7	chr10	-	10222	92607300	92607599	17
HTRA3	chr4	+	84468	8355810	8356109	22

HTT	chr4	+	119297	3195480	3195929	41
HULC	chr6	+	829433	9481680	9482069	41
HULC	chr6	+	844838	9497130	9497429	18
HYDIN	chr16	-	379115	70885290	70885619	33
ICA1	chr7	-	-62699	8364720	8365049	27
ID4	chr6	+	-569142	19268340	19268609	17
IER2	chr19	+	22103	13283130	13283639	62
IER3	chr6	-	-24802	30736890	30737369	45
IER5	chr1	+	1377	181058880	181059149	21
IFI35	chr17	+	4066	41162700	41162969	21
IGF1	chr12	-	-86731	102960900	102961319	37
IGF1R	chr15	+	227689	99420240	99420659	39
IGF2	chr11	-	61555	2098440	2098859	32
IGF2BP3	chr7	-	112816	23397000	23397359	36
IGF2BP3	chr7	-	135676	23374170	23374469	20
IGFBP3	chr7	-	-913	45961620	45961949	29
IGFBP3	chr7	-	-225898	46186590	46186949	24
IGFBP3	chr7	-	-637423	46598160	46598429	19
IGFBP7	chr4	-	-6080	57982440	57982799	21
IKBKB	chr8	+	26961	42155580	42155999	33
IL12B	chr5	-	-11593	158768820	158769329	77
IL12B	chr5	-	-7393	158764740	158765009	19
IL16	chr15	+	69801	81558870	81559169	19
IL2	chr4	-	-51754	123429240	123429569	26
IL22	chr12	-	13237	68633910	68634179	16
IL2RB	chr22	-	-102	37545930	37546199	19
ILK	chr11	+	3671	6628440	6628829	21
INADL	chr1	+	1906	62209920	62210189	19
INHBA	chr7	-	-213823	41956380	41956679	28
INO80	chr15	-	-569	41408730	41409089	25
INPP5F	chr10	+	72891	121558320	121558679	22
INSIG2	chr2	+	24630	118870470	118870889	45
IPCEF1	chr6	-	36736	154614300	154614659	25
IPCEF1	chr6	-	-58904	154709970	154710269	24
IPMK	chr10	-	49940	59977440	59978069	40
IPO9	chr1	+	20907	201819030	201819359	23
IQCJ	chr3	+	122358	158909160	158909789	83
IQCJ	chr3	+	13443	158800380	158800739	19
IQGAP1	chr15	+	91047	91022340	91022699	26
IQGAP1	chr15	+	91542	91022880	91023149	25
IQSEC1	chr3	-	-139304	13148130	13148399	21
IQSEC1	chr3	-	-149534	13158360	13158629	16
IRAK1BP1	chr6	+	-78839	79498140	79498559	35
IRAK1BP1	chr6	+	-251579	79325460	79325759	18
IRAK2	chr3	+	8572	10214970	10215299	31
IREB2	chr15	+	19107	78749490	78749759	21
IRF2BP2	chr1	-	-46248	234791370	234791669	18
IRGC	chr19	+	-14614	44205420	44205779	34
IRS1	chr2	-	683552	226979760	226980149	34
IRS1	chr2	-	375167	227288130	227288549	30
IRS1	chr2	-	373532	227289780	227290169	30

IRS1	chr2	-	322892	227340480	227340749	25
IRS1	chr2	-	217592	227445720	227446109	24
IRS1	chr2	-	125372	227538000	227538269	18
IRS2	chr13	-	199655	110239080	110239439	35
IRS2	chr13	-	-17485	110456190	110456609	31
IRS2	chr13	-	449750	109989030	109989299	18
IRX3	chr16	-	-136296	54456510	54456839	36
IRX3	chr16	-	150264	54169920	54170309	28
IRX3	chr16	-	-139386	54459540	54459989	19
IRX3	chr16	-	-145026	54465270	54465539	19
ITCH	chr20	+	27848	32978730	32979089	31
ITGA1	chr5	+	134284	52218240	52218599	27
ITGA11	chr15	-	47533	68676750	68677169	34
ITGA2	chr5	+	42554	52327530	52327889	32
ITGA2	chr5	+	-1516	52283490	52283789	20
ITGA6	chr2	+	12031	173304180	173304509	25
ITGA9	chr3	+	330397	37823970	37824449	58
ITGAV	chr2	+	-45690	187408950	187409249	21
ITGB3	chr17	+	47452	45378450	45378869	29
ITGB3	chr17	+	51712	45382740	45383099	24
ITGB5	chr3	-	49340	124556550	124557059	68
ITGB5	chr3	-	100310	124505640	124506029	25
ITGB5	chr3	-	62900	124543110	124543379	19
ITGB6	chr2	-	-22559	161079000	161079299	25
ITGB6	chr2	-	-29699	161086140	161086439	22
ITGBL1	chr13	+	37159	102141990	102142259	21
ITPR1	chr3	+	228261	4763100	4763489	27
ITPR2	chr12	-	202327	26783670	26783939	18
ITPRIPL2	chr16	+	-2804	19122300	19122599	18
IVNS1ABP	chr1	-	-124398	185410620	185411099	24
JAG1	chr20	-	-34545	10689060	10689419	22
JAG1	chr20	-	-127695	10782210	10782569	20
JAK1	chr1	-	45358	65386650	65387009	25
JARID2	chr6	+	1468	15247800	15248189	38
JRKL	chr11	+	148837	96271830	96272159	30
JUB	chr14	-	-382	23446680	23446949	17
JUN	chr1	-	-31099	59280720	59281049	20
JUP	chr17	-	-9445	39952260	39952559	26
KATNAL1	chr13	-	-31586	30912570	30912929	27
KBTBD3	chr11	-	-311669	106259970	106260299	21
KBTBD3	chr11	-	-210524	106158780	106159199	20
KCNA10	chr1	-	-16982	111078570	111078989	30
KCNA5	chr12	+	129660	5282520	5282969	36
KCNA5	chr12	+	-45870	5107080	5107349	17
KCNC4	chr1	+	-8150	110745780	110746049	24
KCND3	chr1	-	-198547	112730160	112730489	17
KCNF1	chr2	+	-11613	11040300	11040599	22
KCNH1	chr1	-	116548	211190700	211191119	57
KCNH1	chr1	-	250198	211057110	211057409	21
KCNH7	chr2	-	17356	163677630	163678139	38
KCNH7	chr2	-	256021	163439040	163439399	28

KCNIP4	chr4	-	179180	21770970	21771419	51
KCNIP4	chr4	-	935570	21014610	21014999	35
KCNIP4	chr4	-	497630	21452580	21452909	23
KCNIP4	chr4	-	906920	21043320	21043589	21
KCNJ2	chr17	+	124194	68289720	68290019	22
KCNJ3	chr2	+	278212	155833170	155833439	22
KCNK2	chr1	+	199525	215455860	215456309	37
KCNK2	chr1	+	108705	215287440	215287739	16
KCNMA1	chr10	-	449998	78947430	78947729	29
KCNMA1	chr10	-	442513	78954900	78955229	22
KCNMA1	chr10	-	221278	79176150	79176449	20
KCNMB2	chr3	+	-153854	178100160	178100579	18
KCNQ1	chr11	+	401389	2867430	2867789	29
KCNT2	chr1	-	868975	195708390	195708659	20
KCNT2	chr1	-	839305	195738060	195738329	20
KCNV1	chr8	-	-1112225	112098990	112099379	31
KCNV1	chr8	-	-186665	111173460	111173789	21
KCTD1	chr18	-	34601	24093690	24094109	62
KCTD16	chr5	+	313423	143863680	143864039	31
KCTD16	chr5	+	251338	143801580	143801969	22
KCTD8	chr4	-	290765	44159910	44160209	19
KDM5A	chr12	-	56466	442020	442289	19
KIAA0087	chr7	-	-10840	26589150	26589419	17
KIAA0101	chr15	-	3033	64670490	64670849	39
KIAA0513	chr16	+	27695	85088910	85089299	38
KIAA0564	chr13	-	264337	42270690	42271079	28
KIAA0802	chr18	+	173851	8891010	8891429	36
KIAA0831	chr14	-	29212	55849110	55849619	45
KIAA1026	chr1	+	-112098	14812950	14813279	18
KIAA1161	chr9	-	-120	34376820	34377209	29
KIAA1199	chr15	+	25863	81097440	81097709	49
KIAA1370	chr15	-	63211	52907430	52907789	33
KIAA1462	chr10	-	-50859	30387420	30387899	37
KIAA1462	chr10	-	45516	30291090	30291479	27
KIAA1486	chr2	+	112238	226377660	226378019	33
KIAA1486	chr2	+	-59647	226205790	226206119	17
KIAA1618	chr17	+	-5932	78228540	78228929	50
KIF13A	chr6	-	123025	17864550	17864999	27
KIF13A	chr6	-	111265	17876400	17876669	18
KIF14	chr1	-	-19387	200609070	200609429	35
KIF20B	chr10	+	525833	91986900	91987499	70
KIF20B	chr10	+	448193	91909380	91909739	28
KIF20B	chr10	+	520913	91982100	91982459	23
KIF20B	chr10	+	479858	91940970	91941479	22
KIF20B	chr10	+	492188	91953420	91953689	17
KIF2B	chr17	+	-360194	51539850	51540239	28
KIF3B	chr20	+	8208	30873480	30873869	25
KIF5B	chr10	-	55262	32289930	32290289	32
KIFAP3	chr1	-	148390	169895310	169895669	26
KIFAP3	chr1	-	140200	169903530	169903829	20
KILLIN	chr10	-	-331025	89954040	89954399	36

KILLIN	chr10	-	-206450	89829510	89829779	22
KILLIN	chr10	-	-178775	89801820	89802119	21
KIT	chr4	+	-132050	55391820	55392269	58
KITLG	chr12	-	-345646	89319720	89320049	51
KLC2	chr11	+	-15	66024540	66024959	43
KLF10	chr8	-	-102406	103768350	103768799	39
KLF14	chr7	-	-153229	130571910	130572269	29
KLF6	chr10	-	-346501	4173750	4174199	34
KLF6	chr10	-	-20746	3848040	3848399	31
KLF6	chr10	-	-317866	4145160	4145519	22
KLHL3	chr5	-	116575	136954980	136955429	26
KLRAQ1	chr2	+	-19908	48647760	48648239	55
KRR1	chr12	-	-111071	76016310	76016669	16
KRT18	chr12	+	27787	53370480	53370779	23
KRT19	chr17	-	-1998	39686400	39686879	67
KRTAP2-1	chr17	-	-1841	39205260	39205559	20
KTN1	chr14	+	25835	56072520	56072999	59
KTN1	chr14	+	19070	56065830	56066159	25
LAMA2	chr6	+	615644	129819780	129820079	20
LAMB1	chr7	-	7135	107636520	107636819	22
LAMB4	chr7	-	38837	107731710	107732219	47
LANCL2	chr7	+	-16796	55416120	55416569	58
LASP1	chr17	+	31463	37057350	37057799	45
LBR	chr1	-	2075	225613560	225613859	24
LCNL1	chr9	+	-810	139876500	139876769	19
LCORL	chr4	-	-545759	18569010	18569279	17
LDB2	chr4	-	40260	16860030	16860299	19
LDB2	chr4	-	-243765	17144040	17144339	18
LDHB	chr12	-	6852	21803730	21804119	32
LDLRAD3	chr11	+	90998	36056340	36056879	61
LEKR1	chr3	+	66614	156610560	156610859	28
LEO1	chr15	-	-10471	52274220	52274639	42
LEPREL1	chr3	-	108784	189729960	189730289	18
LEPREL1	chr3	-	4369	189834360	189834719	17
LGALS8	chr1	+	-1200	236680200	236680529	20
LHX6	chr9	-	10915	124979910	124980299	31
LIMA1	chr12	-	7419	50608890	50609249	48
LIMA1	chr12	-	-31896	50648220	50648549	31
LIMA1	chr12	-	20634	50595720	50595989	26
LIMCH1	chr4	+	-40999	41321670	41321939	31
LIMCH1	chr4	+	84941	41447610	41447879	22
LIMCH1	chr4	+	-36919	41325750	41326019	19
LIMK2	chr22	+	31685	31639740	31640129	30
LIPA	chr10	-	-10784	91022160	91022729	51
LIPA	chr10	-	-8489	91019910	91020389	50
LIPA	chr10	-	931	91010580	91010879	24
LIPC	chr15	+	119805	58843800	58844159	24
LIPG	chr18	+	123533	47211750	47212169	30
LLGL2	chr17	+	31572	73553190	73553519	24
LLPH	chr12	-	132089	66392310	66392579	22
LMCD1	chr3	+	-45276	8498100	8498369	22

LMCD1	chr3	+	-8691	8534670	8534969	19
LMO4	chr1	+	145029	87938910	87939449	102
LMO4	chr1	+	76014	87869970	87870359	21
LMO4	chr1	+	467229	88261230	88261529	17
LMOD1	chr1	-	29772	201885780	201886109	24
LNX1	chr4	-	-205758	54629940	54630449	29
LOC100128023	chr3	-	-9057	193720920	193721249	29
LOC100128292	chr10	+	44550	79730910	79731329	41
LOC100129716	chr5	+	-161364	90514620	90514979	29
LOC100129716	chr5	+	-213969	90462000	90462389	26
LOC100130987	chr11	+	61950	67147050	67147469	25
LOC100132288	chr21	-	134736	9833700	9833999	15
LOC100133893	chr12	-	-1590	27925650	27925949	22
LOC100188947	chr10	-	233233	93137790	93138179	34
LOC100189589	chr2	+	5555	74618190	74618609	40
LOC100216001	chr10	-	-26367	4746420	4746839	38
LOC100233209	chr12	-	-144388	47754480	47754749	44
LOC100268168	chr5	-	3004	172382610	172382939	23
LOC100287227	chr3	-	-43372	156436740	156437009	19
LOC144776	chr13	-	509937	91068720	91069109	42
LOC145820	chr15	+	-185797	95790390	95790659	22
LOC145820	chr15	+	-151567	95824620	95824889	17
LOC152118	chr3	+	107241	153309390	153309659	17
LOC153910	chr6	-	68704	142890090	142890449	31
LOC196993	chr15	-	18640	71388930	71389469	60
LOC254312	chr10	-	36747	10957200	10957559	25
LOC283404	chr12	+	4951	52609530	52609799	16
LOC284632	chr1	+	36700	24563280	24563579	21
LOC285593	chr5	+	-122541	172883910	172884299	43
LOC285593	chr5	+	-124821	172881630	172882019	28
LOC285627	chr5	-	2005	158891100	158891459	27
LOC285627	chr5	-	55675	158837430	158837789	22
LOC285696	chr5	-	29212	17188140	17188499	21
LOC285696	chr5	-	36247	17181150	17181419	18
LOC285954	chr7	+	-113539	41619720	41620229	59
LOC285954	chr7	+	-504874	41228460	41228819	24
LOC285954	chr7	+	-446254	41287020	41287499	22
LOC338588	chr10	+	-413493	4284690	4285019	19
LOC338758	chr12	+	223833	90326400	90326729	34
LOC338758	chr12	+	120783	90223380	90223649	19
LOC388965	chr2	+	-1174896	83342760	83343059	26
LOC388965	chr2	+	-978441	83539230	83539499	18
LOC399959	chr11	-	13956	122059620	122060009	26
LOC399959	chr11	-	56331	122017260	122017619	19
LOC399959	chr11	-	65661	122007960	122008259	17
LOC400804	chr1	-	649	221508750	221509229	56
LOC400804	chr1	-	-101816	221611260	221611649	24
LOC643955	chr7	-	795130	61969140	61969469	15.31531532
LOC646982	chr13	-	-22911	41077890	41078219	22
LOC646999	chr7	+	-1026	39647820	39648299	29
LOC647979	chr20	-	-11987	34650720	34651019	16

LOC723972	chr15	+	28948	35558340	35558609	21
LOC729020	chr10	+	-13504	104991900	104992379	48
LOC729176	chr6	-	-101734	147226440	147226949	51
LOC729176	chr6	-	-111394	147236160	147236549	20
LOC729467	chr1	-	45560	59566770	59567069	24
LOC729920	chr7	-	4253	16456560	16456829	17
LOC732275	chr16	-	-51524	86430600	86431019	41
LOC732275	chr16	-	-47219	86426280	86426729	34
LOC91450	chr15	-	27218	78259200	78259499	21
LOC91948	chr15	-	-28075	98445540	98445929	24
LOC91948	chr15	-	377525	98040000	98040269	21
LOXL1	chr15	+	8126	74226720	74227109	46
LOXL1	chr15	+	13976	74232630	74232899	16
LOXL2	chr8	-	48098	23213460	23213789	25
LPAR1	chr9	-	67906	113732220	113732699	36
LPHN2	chr1	+	310193	82576020	82576529	63
LPHN2	chr1	+	7433	82273290	82273739	39
LPHN2	chr1	+	764498	83030370	83030789	20
LPHN2	chr1	+	139118	82404990	82405409	19
LPHN2	chr1	+	149183	82415100	82415429	18
LPHN3	chr4	+	-1781139	60581520	60581879	20
LPIN1	chr2	+	430555	12317100	12317489	40
LPP	chr3	+	-16121	187914450	187914749	27
LRAT	chr4	+	-23123	155641890	155642189	22
LRBA	chr4	-	705110	151231620	151231919	33
LRFN2	chr6	-	-146113	40700970	40701509	67
LRFN5	chr14	+	1319826	43396440	43396739	23
LRP1B	chr2	-	-59349	142948440	142948799	28
LRP6	chr12	-	-16043	12435630	12436079	70
LRRC16A	chr6	+	318137	25597620	25597949	21
LRRC20	chr10	-	5545	72135660	72136079	37
LRRC28	chr15	+	2373	99793800	99794249	49
LRRC29	chr16	-	2657	67257870	67258259	20
LRRC3B	chr3	+	-185010	26479110	26479469	34
LRRC4C	chr11	-	19920	40295550	40295939	43
LRRC59	chr17	-	-952	48475650	48475919	19
LRRC8D	chr1	+	69392	90355770	90356159	36
LRRFIP2	chr3	-	1382	37216320	37216619	18
LRRIQ3	chr1	-	287587	74376150	74376419	22
LRRIQ3	chr1	-	407737	74255970	74256299	17
LRRTM1	chr2	-	-1930247	82461540	82461929	18
LSAMP	chr3	-	247274	115916880	115917329	44
LSAMP	chr3	-	532304	115631850	115632299	43
LTA4H	chr12	-	1416	96427770	96428129	26
LTA4H	chr12	-	-5364	96434580	96434879	25
LTBP2	chr14	-	-8490	75087330	75087719	27
LYPD6B	chr2	+	55909	149950710	149951069	23
LYPD6B	chr2	+	31009	149925840	149926139	22
LYPLAL1	chr1	+	130528	219477540	219477899	26
LYPLAL1	chr1	+	-286412	219060600	219060959	22
LYPLAL1	chr1	+	344953	219691980	219692309	19

LYZL1	chr10	+	-147090	29430690	29431109	48
LYZL1	chr10	+	29370	29607150	29607569	22
LYZL2	chr10	-	-58047	30976470	30976919	52
LYZL2	chr10	-	-52857	30971280	30971729	38
MACF1	chr1	+	61672	39608580	39608999	29
MACROD2	chr20	+	1932554	15908550	15908849	20
MAF	chr16	-	-31832	79666290	79666619	21
MAF	chr16	-	-440837	80075310	80075609	21
MAGI2	chr7	-	66326	79016370	79016759	31
MAGI2	chr7	-	211796	78870960	78871229	20
MAGI2	chr7	-	18476	79064280	79064549	17
MAGI3	chr1	+	4200	113937510	113937839	21
MAL2	chr8	+	-28895	120191430	120191999	37
MALAT1	chr11	+	4897	65269950	65270309	30
MAML2	chr11	-	197755	95878320	95878859	91
MAML2	chr11	-	352045	95724120	95724479	22
MAML2	chr11	-	31330	96044880	96045149	19
MAML2	chr11	-	129520	95946690	95946959	17
MAN2A1	chr5	+	-32381	108992550	108992999	44
MANEA	chr6	+	23892	96049080	96049529	36
MAP1LC3C	chr1	-	-7154	242169360	242169719	21
MAP2	chr2	+	-440286	209848350	209848619	20
MAP2K4	chr17	+	27430	11951400	11951729	27
MAP2K6	chr17	+	258377	67668960	67669469	43
MAP3K3	chr17	+	53699	61753200	61753799	80
MAP3K5	chr6	-	185197	136928250	136928669	42
MAP3K71P2	chr6	+	16372	149655240	149655629	33
MAP4	chr3	-	-466	47952180	47952509	25
MAP4K5	chr14	-	2857	50996280	50996759	61
MAPK10	chr4	-	-23153	87051780	87052139	29
MAPK6	chr15	+	12534	52323780	52324109	24
MAPKAP1	chr9	-	31924	128437320	128437859	35
MAPKAP1	chr9	-	51079	128418300	128418569	15
MARK1	chr1	+	103217	220804560	220805009	48
MAST4	chr5	+	332326	66456750	66457109	17
MAT2B	chr5	+	928045	163860450	163860809	28
MAT2B	chr5	+	66490	162998850	162999299	27
MAT2B	chr5	+	1012735	163945170	163945469	26
MAT2B	chr5	+	1143925	164076360	164076659	21
MAT2B	chr5	+	1816510	164748960	164749229	19
MATN1	chr1	-	710988	30485280	30485609	22
MBD2	chr18	-	66199	51684780	51685139	28
MBIP	chr14	-	2803	36786870	36787289	37
MBIP	chr14	-	81958	36707730	36708119	27
MBIP	chr14	-	102868	36686850	36687179	25
MBL2	chr10	-	-993889	55525170	55525529	24
MBL2	chr10	-	-112444	54643770	54644039	23
MBNL2	chr13	+	-16209	97858200	97858529	22
MBNL2	chr13	+	-100194	97774230	97774529	17
MBOAT7	chr19	-	1394	54692070	54692609	50
MCF2L	chr13	+	54383	113676930	113677349	38

MCF2L2	chr3	-	46556	183099090	183099509	29
MDGA1	chr6	-	30377	37635210	37635569	42
ME3	chr11	-	152891	86230080	86230619	34
ME3	chr11	-	120641	86262420	86262779	30
MECOM	chr3	-	-285086	169149030	169149329	24
MED12L	chr3	+	218459	151022940	151023329	43
MED13L	chr12	-	146587	116568210	116568599	30
MED24	chr17	-	28315	38182350	38182799	35
MED27	chr9	-	71339	134883720	134884109	35
MED9	chr17	+	-30505	17349570	17350019	48
MEGF9	chr9	-	12048	123464430	123464699	25
MEGF9	chr9	-	105828	123370650	123370919	16
MELK	chr9	+	94115	36666810	36667229	41
MEOX2	chr7	-	-161736	15887910	15888179	22
MEST	chr7	+	2959	130128810	130129199	30
METT5D1	chr11	+	511682	28641330	28641629	22
METT5D1	chr11	+	292142	28421790	28422089	21
METTL4	chr18	-	159970	2411310	2411729	53
METTL4	chr18	-	268480	2302800	2303219	38
METTL4	chr18	-	42010	2529300	2529659	27
MFHAS1	chr8	-	21027	8729970	8730239	27
MFSD1	chr3	+	-12302	158507460	158507759	20
MGAT4A	chr2	-	-73118	99352800	99353309	51
MGAT4C	chr12	-	-278758	87511290	87511589	27
MGC34034	chr6	+	-52350	134089800	134090069	20
MGC45800	chr4	-	187184	182878290	182878679	38
MGC45800	chr4	-	178349	182887170	182887469	20
MIAT	chr22	+	46131	27099450	27099779	35
MICAL2	chr11	+	90012	12222000	12222299	28
MIPOL1	chr14	+	349182	38016120	38016479	27
MITF	chr3	+	67552	69856020	69856349	37
MITF	chr3	+	23152	69811530	69812039	28
MITF	chr3	+	108757	69897210	69897569	28
MKKS	chr20	-	16746	10395660	10395989	24
MLANA	chr9	+	-11194	5879580	5879849	33
MLLT3	chr9	-	68795	20553540	20553899	30
MLPH	chr2	+	13927	238409580	238410029	58
MMADHC	chr2	-	-403964	150848100	150848489	47
MMADHC	chr2	-	-24449	150468570	150468989	36
MMP10	chr11	-	20100	102631080	102631409	26
MMP16	chr8	-	-310002	89649570	89649869	23
MMP2	chr16	+	20939	55533810	55534229	45
MOCS2	chr5	-	-144024	52549080	52549619	74
MOV10	chr1	+	24702	113241570	113241929	17
MPDZ	chr9	-	226751	13023450	13023779	24
MPHOSPH8	chr13	+	-46560	20161110	20161529	29
MPHOSPH8	chr13	+	-25665	20182050	20182379	20
MPP4	chr2	-	703	202562520	202562909	45
MPP4	chr2	-	28	202563240	202563539	27
MPPE1	chr18	-	20872	11887590	11887949	38
MPZL2	chr11	-	-11005	118145790	118146239	59

MREG	chr2	-	217297	216660870	216661229	24
MRI1	chr19	+	1193	13876380	13876679	22
MRPS22	chr3	+	-30936	139031700	139032149	53
MRPS22	chr3	+	-41541	139021170	139021469	22
MRPS22	chr3	+	-4521	139058190	139058489	17
MRPS6	chr21	+	40592	35486280	35486549	19
MRVI1	chr11	-	-7546	10681200	10681589	38
MSI2	chr17	+	224419	55558110	55558589	85
MSI2	chr17	+	39904	55373610	55374059	47
MSI2	chr17	+	-41036	55292730	55293059	22
MSI2	chr17	+	202894	55536660	55536989	21
MSMP	chr9	-	-17457	35771520	35771939	34
MTA3	chr2	+	50719	42846150	42846629	46
MTDH	chr8	+	-4182	98652060	98652389	23
MTHFD2L	chr4	+	126156	75149790	75150179	18
MTIF3	chr13	-	-16108	28040640	28040999	23
MTPAP	chr10	-	141213	30496890	30497219	24
MTR	chr1	+	59314	237017580	237018209	27
MTUS1	chr8	-	-68733	17623800	17624159	30
MTUS1	chr8	-	49272	17505840	17506109	20
MUSK	chr9	+	70403	113501310	113501669	27
MUSK	chr9	+	22433	113453340	113453699	26
MYB	chr6	+	62792	135565110	135565379	19
MYBPHL	chr1	-	7934	109841460	109841999	37
MYEOV	chr11	+	6043	69067470	69067859	36
MYH9	chr22	-	56969	36726930	36727259	30
MYH9	chr22	-	26744	36757140	36757499	23
MYH9	chr22	-	-11311	36795210	36795539	21
MYLIP	chr6	+	-19977	16109190	16109489	27
MYLK	chr3	-	-169284	123508530	123509099	71
MYLK	chr3	-	-130929	123470280	123470639	29
MYO10	chr5	-	41631	16894500	16895009	50
MYO10	chr5	-	-56244	16992480	16992779	23
MYO18B	chr22	+	5575	26143560	26143829	19
MYO5A	chr15	-	197288	52623720	52624199	34
MYOF	chr10	-	6235	95235660	95236019	39
MYPN	chr10	+	-15390	69853710	69854009	21
MYT1L	chr2	-	539321	1795530	1795919	24
N4BP3	chr5	+	-32746	177507600	177508019	33
N4BP3	chr5	+	-34996	177505380	177505739	18
N6AMT1	chr21	-	62394	30195150	30195449	19
N6AMT1	chr21	-	53094	30204420	30204779	19
NAALADL2	chr3	+	476084	175053000	175053389	63
NACAP1	chr8	+	-12151	102368760	102369179	34
NACAP1	chr8	+	-80671	102300210	102300689	31
NACAP1	chr8	+	-26596	102354360	102354689	24
NACAP1	chr8	+	-4366	102376590	102376919	23
NAF1	chr4	-	328019	163759860	163760249	25
NAMPT	chr7	-	68614	105856890	105857159	21
NAMPT	chr7	-	-59816	105985320	105985589	20
NAP1L1	chr12	-	-18696	76497240	76497629	32

NAP1L4	chr11	-	-8427	3021900	3022169	22
NARS2	chr11	-	-32000	78317730	78318089	28
NAV2	chr11	+	691504	20063520	20064029	45
NAV2	chr11	+	296464	19668510	19668959	33
NAV3	chr12	+	14796	78239730	78239999	20
NAV3	chr12	+	93666	78318600	78318869	18
NBAS	chr2	-	363130	15338190	15338459	20
NBEA	chr13	+	-529946	34986360	34986659	30
NBEA	chr13	+	-371081	35145240	35145509	22
NBN	chr8	-	930	90995820	90996119	15
NBPF4	chr1	-	-84091	108870600	108870989	19
NCOR1	chr17	-	64781	16053840	16054289	38
NCRNA00052	chr15	+	-227915	87892110	87892379	28
NCRNA00052	chr15	+	42520	88162530	88162829	23
NCRNA00052	chr15	+	-207440	87912570	87912869	20
NCRNA00081	chr10	-	6315	112672230	112672529	13
NCRNA00158	chr21	-	1226134	25577610	25578149	28
NCRNA00161	chr21	+	97945	30009450	30009719	23
NCRNA00161	chr21	+	102580	30014040	30014399	22
NCRNA00161	chr21	+	149215	30060720	30060989	19
NCRNA00164	chr2	-	-97489	133016700	133017329	16.01362862
NCRNA00164	chr2	-	-110239	133029630	133029899	15.16183986
NDC80	chr18	+	16515	2587860	2588189	15
NDFIP1	chr5	+	60766	141548940	141549239	30
NDST1	chr5	+	9681	149897160	149897549	39
NDUFB5	chr3	+	9160	179331600	179331869	19
NEAT1	chr11	+	-1784	65188350	65188619	19
NEBL	chr10	-	6097	21180210	21180659	44
NEBL	chr10	-	171892	21291090	21291359	23
NEBL	chr10	-	-17558	21480540	21480809	21
NEDD1	chr12	+	-44691	97256100	97256519	38
NEDD4	chr15	-	-125615	56334720	56335169	45
NEDD4L	chr18	+	117421	55828860	55829219	23
NEDD9	chr6	-	-45423	11427870	11428139	23
NEDD9	chr6	-	-46218	11428620	11428979	22
NEDD9	chr6	-	-29838	11412240	11412599	17
NEK1	chr4	-	7780	170525670	170525999	26
NEK10	chr3	-	57723	27353010	27353369	29
NEK11	chr3	+	82581	130828050	130828499	31
NF1	chr17	+	154225	29575980	29576459	73
NF1	chr17	+	174895	29596680	29597099	40
NF1	chr17	+	-42260	29379570	29379899	22
NFASC	chr1	+	95948	204893490	204893969	27
NFE2L3	chr7	+	-304442	25887120	25887689	81
NFIB	chr9	-	30511	14283300	14283569	21
NFIL3	chr9	-	-136555	94322520	94322879	24
NFKBIA	chr14	-	48126	35825640	35826029	29
NFRKB	chr11	-	30890	129731880	129732149	19
NFX1	chr9	+	38500	33328860	33329159	12
NFYB	chr12	-	28946	104502870	104503319	58
NFYB	chr12	-	-33649	104565510	104565869	27

NID2	chr14	-	-10508	52546260	52546649	32
NKAIN2	chr6	+	490896	124615770	124616159	42
NKX2-4	chr20	-	-35667	21413520	21413909	45
NLGN1	chr3	+	513811	173629920	173630189	25
NLRP1	chr17	-	-242017	5729610	5730089	73
NME7	chr1	-	51192	169285830	169286159	17
NMT2	chr10	-	-32109	15242580	15243029	53
NOG	chr17	+	86245	54757020	54757589	96
NOG	chr17	+	42085	54713010	54713279	22
NOS1AP	chr1	+	72469	162111870	162112229	32
NOTCH1	chr9	-	-6706	139446780	139447109	25
NOX5	chr15	+	114561	69337290	69337559	17
NPAS2	chr2	+	82847	101519220	101519699	54
NPAS2	chr2	+	-18598	101417850	101418179	25
NPAS3	chr14	+	630786	34038900	34039589	89
NPAS3	chr14	+	626151	34034460	34034759	18
NPAS3	chr14	+	447096	33855390	33855719	16
NPC2	chr14	-	1260	74958690	74958959	19
NPM1	chr5	+	21537	170836110	170836559	38
NPTX2	chr7	+	2173	98248590	98248949	28
NPY	chr7	+	-231064	24092580	24092909	21
NPY6R	chr5	+	25853	137169120	137169509	32
NR2F2	chr15	+	-60762	96808170	96808619	40
NR2F2	chr15	+	-437847	96431100	96431519	37
NR2F2	chr15	+	-51027	96817950	96818309	31
NR2F2	chr15	+	-255117	96613860	96614219	23
NR3C1	chr5	-	86635	142696440	142696799	33
NR3C1	chr5	-	-9425	142792530	142792829	21
NR3C2	chr4	-	-192206	149555670	149556029	23
NR4A2	chr2	-	388363	156800790	156801059	18
NR6A1	chr9	-	62407	127470930	127471409	58
NR6A1	chr9	-	18052	127515360	127515689	22
NRAP	chr10	-	-2229	115425900	115426169	15
NRG1	chr8	+	691862	32188920	32189339	43
NRG1	chr8	+	45559	32624760	32625059	22
NRG1	chr8	+	235232	31732320	31732679	20
NRG2	chr5	-	31930	139390800	139391099	27
NRG3	chr10	+	991300	84626190	84626549	15
NRP1	chr10	-	71174	33552480	33552839	26
NRP1	chr10	-	-31171	33654840	33655169	26
NRP1	chr10	-	141149	33482550	33482819	22
NRP2	chr2	+	103531	206650560	206650949	26
NRXN1	chr2	-	-1573010	52147650	52148159	34
NRXN3	chr14	+	921507	79791420	79791779	30
NSMAF	chr8	-	-142238	59713980	59714429	56
NSMCE1	chr16	-	33274	27246630	27247049	33
NSUN4	chr1	+	1710	46807950	46808249	19
NSUN6	chr10	-	6	18940410	18940679	19
NT5C1B	chr2	-	-240266	19010940	19011269	18
NT5C1B	chr2	-	-173156	18943830	18944159	13
NT5C2	chr10	-	24322	104928600	104928869	15

NTNG1	chr1	+	335596	108018090	108018359	14
NTSR1	chr20	+	-5204	61334730	61335239	59
NUAK1	chr12	-	77222	106456410	106456769	18
NUAK1	chr12	-	7277	106526400	106526669	17
NUCB2	chr11	+	-841	17297310	17297579	26
NUDT12	chr5	-	-627354	103525650	103526039	32
NUDT12	chr5	-	93366	102804960	102805289	22
NUDT12	chr5	-	-305139	103203450	103203809	18
NUDT16	chr3	+	21118	131121630	131122019	40
NUDT4	chr12	+	-148361	93623190	93623489	21
NUFIP1	chr13	-	172039	45391410	45391739	22
NUFIP1	chr13	-	147199	45416280	45416549	20
NUP210L	chr1	-	153343	153974100	153974399	20
NUP93	chr16	+	77358	56841120	56841629	48
NXN	chr17	-	46176	836670	836999	23
NXN	chr17	-	51486	831360	831689	16
NXPH2	chr2	-	-169308	139706940	139707299	27
NXPH2	chr2	-	-168153	139705800	139706129	26
NXPH2	chr2	-	-97368	139635000	139635359	24
OASL	chr12	-	12616	121464030	121464299	17
OBFC2A	chr2	+	-31778	192510870	192511169	29
OBFC2A	chr2	+	-40448	192502200	192502499	20
ODZ2	chr5	+	10757	166722420	166722779	20
ODZ2	chr5	+	54572	166766280	166766549	20
ODZ3	chr4	+	-83452	183161520	183161849	25
ODZ4	chr11	-	724766	78426720	78427139	45
OGFRL1	chr6	+	-532	71997810	71998079	19
OLFM3	chr1	-	163661	102298950	102299309	26
OR10V1	chr11	-	6799	59474370	59474669	20
OR13F1	chr9	+	-118379	107148000	107148329	22
OR111	chr19	+	-13992	15183750	15184019	18
OR2B11	chr1	-	1375	247613760	247614059	23
OR2J3	chr6	+	30472	29109960	29110319	25
OR2V2	chr5	+	13122	180594900	180595229	18
OR51B6	chr11	+	12892	5385480	5385779	22
OR5B3	chr11	-	9013	58161690	58162049	30
ORC5L	chr7	-	83324	103764990	103765289	26
OSBPL10	chr3	-	150054	31873020	31873349	29
OSGIN2	chr8	+	-18956	90894960	90895319	41
OSR1	chr2	-	232433	19325610	19326269	20
OSTC	chr4	+	-6071	109565460	109565879	32
OSTCL	chr6	-	17620	159260880	159261209	38
OSTCL	chr6	-	3040	159275490	159275759	26
OTOL1	chr3	+	393709	161608110	161608499	52
OTOL1	chr3	+	1575124	162789510	162789929	32
OTUD1	chr10	+	75272	23803290	23803649	28
OXA1L	chr14	+	-66161	23169300	23169839	70
OXA1L	chr14	+	-13841	23221740	23222039	25
OXGR1	chr13	-	-8480	97654920	97655249	22
OXSM	chr3	+	309717	26141130	26141429	16
P4HA1	chr10	-	52243	74804250	74804729	58

PABPCP2	chr2	+	-993270	146351070	146351639	42
PACSN2	chr22	-	20843	43322070	43322339	21
PAIP2B	chr2	-	45549	71408490	71408879	25
PAK1	chr11	-	91679	77093280	77093579	24
PAK6	chr15	+	32221	40541610	40542089	37
PALLD	chr4	+	125598	169543590	169544039	27
PALLD	chr4	+	139788	169557810	169558199	23
PALM2	chr9	+	-84527	112318350	112318739	26
PALM2	chr9	+	-79502	112323420	112323719	23
PALM2-AKAP2	chr9	+	206618	112749000	112749389	38
PALM2-AKAP2	chr9	+	351173	112893570	112893929	26
PALMD	chr1	+	-188151	99923070	99923489	23
PALMD	chr1	+	-46011	100065240	100065599	22
PALMD	chr1	+	-46461	100064730	100065209	20
PAP2D	chr1	-	97955	99372300	99372689	23
PAPSS2	chr10	+	45849	89465160	89465489	34
PAPSS2	chr10	+	-24051	89395260	89395589	27
PARD3	chr10	-	115824	34987890	34988309	52
PARD3	chr10	-	408759	34694940	34695389	33
PARD3	chr10	-	148764	34955010	34955309	23
PARD3B	chr2	+	405024	205815330	205815749	43
PARD3B	chr2	+	79674	205489980	205490399	24
PARD6B	chr20	+	30044	49377930	49378319	40
PARD6B	chr20	+	-41521	49306350	49306769	35
PARVA	chr11	+	40489	12439470	12439799	25
PARVB	chr22	+	43767	44438790	44439089	19
PAWR	chr12	-	141346	79943280	79943609	20
PAWR	chr12	-	144316	79940310	79940639	18
PBX1	chr1	+	176193	164704830	164705159	22
PBX3	chr9	+	107718	128617080	128617589	54
PCDH15	chr10	-	-343173	56904090	56904359	14
PCDH21	chr10	+	1498	85955820	85956209	30
PCDH7	chr4	+	42993	30764880	30765179	26
PCDH8	chr13	-	39080	53383500	53383889	40
PCDH9	chr13	-	-1209706	69014010	69014339	30
PCNXL2	chr1	-	285425	233145870	233146199	35
PCSK1	chr5	-	180463	95588340	95588639	20
PCSK6	chr15	-	43613	101986350	101986799	74
PCTK2	chr12	-	-45431	96839490	96839819	35
PCTK2	chr12	-	2269	96791760	96792149	33
PCTK2	chr12	-	-87401	96881490	96881759	20
PDC	chr1	-	-43993	186461640	186462059	42
PDC	chr1	-	5822	186411810	186412259	28
PDCD1LG2	chr9	+	-805	5509560	5509919	40
PDCD6IP	chr3	+	83375	33953370	33953639	25
PDE10A	chr6	-	164425	165910950	165911369	36
PDE11A	chr2	-	-2242	178939530	178939919	45
PDE1C	chr7	-	120807	31989960	31990409	30
PDE3A	chr12	+	-179782	20342250	20342579	20
PDE3B	chr11	+	121626	14786670	14787119	50
PDE7B	chr6	+	239431	136412070	136412459	22

PDE7B	chr6	+	199666	136372350	136372649	19
PDE8A	chr15	+	49195	85574250	85574549	24
PDE8B	chr5	+	14069	76520610	76520939	26
PDGFC	chr4	-	20262	157872060	157872509	51
PDGFC	chr4	-	-4998	157897350	157897739	37
PDGFC	chr4	-	200007	157692330	157692749	30
PDGFC	chr4	-	387132	157505250	157505579	26
PDGFC	chr4	-	332277	157560120	157560419	21
PDGFC	chr4	-	197472	157694880	157695269	16
PDGFD	chr11	-	29393	104005470	104005799	19
PDHX	chr11	+	-106417	34831080	34831439	33
PDHX	chr11	+	61178	34998570	34999139	29
PDIK1L	chr1	+	-13401	26424120	26424389	19
PDLIM1	chr10	-	30437	97020210	97020479	17
PDLIM5	chr4	+	93652	95466480	95466899	40
PDLIM5	chr4	+	96712	95469600	95469899	22
PDP2	chr16	+	6024	66920310	66920609	17
PECI	chr6	-	-184768	4320420	4320779	40
PEX3	chr6	+	967	143772750	143773019	16
PFDN1	chr5	-	-8125	139690680	139690949	15
PFN2	chr3	-	-111158	149799750	149800049	39
PFTK1	chr7	+	107598	90446070	90446549	40
PFTK2	chr2	+	-8148	202662870	202663229	19
PGC	chr6	-	-313	41715300	41715569	21
PGCP	chr8	+	444976	98102220	98102729	77
PGPEP1	chr19	+	24437	18475710	18475979	26
PGR	chr11	-	-167270	101167680	101167949	23
PGRMC2	chr4	-	-35461	129244230	129244589	30
PGRMC2	chr4	-	-51121	129259860	129260279	26
PGRMC2	chr4	-	-173701	129382500	129382799	23
PHACTR1	chr6	+	22	12717720	12717989	18
PHF11	chr13	+	4384	50073990	50074379	29
PHF20	chr20	+	50047	34409790	34410149	34
PHLDA1	chr12	-	247067	76178250	76178729	52
PHLDA1	chr12	-	49412	76375980	76376309	34
PHLDB1	chr11	+	12962	118490010	118490339	36
PHLDB2	chr3	+	119348	111570480	111570869	43
PHLDB2	chr3	+	191108	111642210	111642659	27
PHTF2	chr7	+	35956	77463870	77464259	50
PHTF2	chr7	+	13246	77441100	77441609	30
PHTF2	chr7	+	143043	77612340	77612639	24
PID1	chr2	-	2488	230133420	230133719	24
PIK3AP1	chr10	-	37185	98442930	98443259	19
PIK3C2A	chr11	-	-39640	17230830	17231159	34
PIK3C2G	chr12	+	14586	18428880	18429239	27
PIM1	chr6	+	-35407	37102350	37102679	26
PINX1	chr8	-	37955	10659150	10659539	28
PINX1	chr8	-	45920	10651230	10651529	21
PION	chr7	-	56643	76988940	76989209	16
PISD	chr22	-	-15874	32042490	32042879	55
PITPNC1	chr17	+	13461	65387220	65387549	24

PITRM1	chr10	-	-24996	3239820	3240179	37
PJA2	chr5	-	-124054	108869400	108870059	22
PKD1L1	chr7	-	31748	47956080	47956499	28
PKHD1	chr6	-	408104	51544110	51544529	25
PKIG	chr20	+	1299	43161570	43161899	26
PKN2	chr1	+	-507302	88642410	88642829	31
PKNOX2	chr11	+	136316	125170740	125171009	20
PKP1	chr1	+	26135	201278430	201278999	100
PKP4	chr2	+	46284	159359580	159359939	34
PLA2G4A	chr1	+	210193	187008060	187008389	37
PLA2G4A	chr1	+	580033	187377900	187378229	19
PLAC8	chr4	-	1712	84033960	84034439	64
PLAGL1	chr6	-	46631	144338970	144339239	19
PLB1	chr2	+	48828	28767630	28767989	25
PLB1	chr2	+	-45957	28672890	28673159	16
PLCE1	chr10	+	305609	96059160	96059549	31
PLCL1	chr2	+	23124	198692370	198692729	27
PLCL2	chr3	+	110398	17036670	17037029	25
PLCXD2	chr3	+	60828	111454200	111454469	21
PLCXD3	chr5	-	2266	41508300	41508629	18
PLD1	chr3	-	-1286	171529380	171529739	28
PLD1	chr3	-	98044	171430080	171430379	27
PLD1	chr3	-	-34706	171562680	171563279	23
PLEKHA6	chr1	-	72755	204255960	204256619	35
PLEKHA6	chr1	-	55490	204273420	204273689	19
PLEKHA9	chr12	-	23455	45586170	45586499	29
PLEKHF2	chr8	+	-32638	96113190	96113609	62
PLEKHG4B	chr5	+	36402	176640	176909	17
PLEKHG7	chr12	+	10825	93140940	93141239	19
PLEKHM3	chr2	-	-9030	208899150	208899479	37
PLEKHM3	chr2	-	11520	208878540	208878989	28
PLK2	chr5	-	271654	57484020	57484499	59
PLOD2	chr3	-	603968	145275090	145275539	25
PLSCR4	chr3	-	-57473	146025960	146026289	31
PLXDC2	chr10	+	-86177	20019030	20019359	37
PLXNA2	chr1	-	54961	208362570	208362839	22
PLXNA4	chr7	-	221334	132039810	132040169	25
PLXNC1	chr12	+	-48889	94493370	94493849	60
PLXNC1	chr12	+	21281	94563570	94563989	24
PLXNC1	chr12	+	-49639	94492680	94493039	22
PLXNC1	chr12	+	-165784	94376580	94376849	13
PMFBP1	chr16	-	-119255	72325080	72325559	62
PNPLA1	chr6	+	47550	36258300	36258689	32
PNPLA8	chr7	-	105649	108060780	108061199	27
PNPT1	chr2	-	-86003	56006760	56007269	70
PODXL	chr7	-	-48373	131289570	131289929	53
POLR3G	chr5	+	14099	89784570	89784989	27
POR	chr7	+	43880	75588120	75588479	37
POTEA	chr8	+	676820	43824210	43824599	24
POU2AF1	chr11	-	-48162	111298140	111298499	31
POU2F1	chr1	+	28252	167218230	167218559	24

POU3F2	chr6	+	-86590	99195840	99196139	24
PPA2	chr4	-	74073	106321020	106321289	18
PPAP2A	chr5	-	-65586	54896250	54896669	44
PPAPDC1A	chr10	+	-40866	122175390	122175809	49
PPAPDC3	chr9	+	15529	134180430	134180789	25
PPARA	chr22	+	28291	46574550	46575029	67
PPARG	chr3	+	118146	12447270	12447719	30
PPFIA2	chr12	-	-183855	82336830	82337099	22
PPFIA2	chr12	-	201210	81951720	81952079	19
PPFIA2	chr12	-	-61170	82214130	82214429	19
PPFIBP1	chr12	+	99745	27776640	27776939	32
PPIA	chr7	+	14599	44850660	44851019	19
PPIF	chr10	+	20490	81127530	81127889	23
PPM1B	chr2	+	33685	44429520	44429849	23
PPP1R15B	chr1	-	-5695	204386220	204387059	44
PPP1R3C	chr10	-	-39366	93432030	93432419	34
PPP2CA	chr5	-	-53609	133615350	133615769	24
PPP2R2B	chr5	-	89550	146168460	146168759	26
PPP4R2	chr3	+	40181	73086150	73086449	18
PPPDE1	chr1	+	66883	244882980	244883489	79
PRCC	chr1	+	29931	156767010	156767399	33
PRDM10	chr11	-	-18867	129836160	129836609	34
PRDM4	chr12	-	1630	108153150	108153419	17
PRELID2	chr5	-	283218	144931500	144931799	18
PREP	chr6	-	-29770	105880530	105880949	41
PRICKLE1	chr12	-	-160418	43037640	43038029	25
PRICKLE1	chr12	-	-438608	43315890	43316159	20
PRKAA1	chr5	-	6353	40791750	40792139	20
PRKACB	chr1	+	-4645	84538920	84539279	28
PRKACB	chr1	+	103145	84646740	84647039	18
PRKAR2A	chr3	-	59116	48825960	48826349	25
PRKAR2B	chr7	+	6717	106691670	106692119	40
PRKCA	chr17	+	-29476	64269240	64269659	44
PRKCA	chr17	+	265004	64563780	64564079	29
PRKCA	chr17	+	371324	64670100	64670399	20
PRKCE	chr2	+	416162	46294950	46295459	56
PRKCE	chr2	+	421802	46300680	46301009	27
PRKCH	chr14	+	150640	61939020	61939289	24
PRKD1	chr14	-	-296100	30692820	30693179	22
PRKG2	chr4	-	38461	82087560	82087949	25
PRKRIP1	chr7	+	-2709	102033960	102034229	24
PRL	chr6	-	-65829	22363380	22363739	25
PRMT6	chr1	+	-1571362	106027770	106028039	18
PRMT7	chr16	+	21515	68366310	68366609	22
PRNP	chr20	+	-19932	4646610	4647119	39
PRNP	chr20	+	-52287	4614360	4614659	24
PRNT	chr20	-	-21055	4742220	4742519	24
PRPF39	chr14	+	28998	45582090	45582509	26
PRPS1L1	chr7	-	-153073	18220380	18220739	18
PRR5L	chr11	+	156935	36474450	36474869	29
PRRC1	chr5	+	-364	126852810	126853079	18

PRSS2	chr7	+	26017	142505730	142506119	28
PRSS2	chr7	+	26707	142506450	142506779	28
PRSS23	chr11	+	-10121	86501220	86501519	23
PRTG	chr15	-	5498	56029440	56029919	57
PSAP	chr10	-	38	73610820	73611269	39
PSAP	chr10	-	1118	73609770	73610159	21
PSAT1	chr9	+	242996	81154920	81155189	21
PSD3	chr8	-	-146054	18812280	18812639	33
PSMB2	chr1	-	35459	36071550	36071819	20
PSMD1	chr2	+	11766	231933210	231933539	23
PSMD1	chr2	+	75381	231996810	231997169	23
PSMG1	chr21	-	88186	40467060	40467449	33
PSMG2	chr18	+	38491	12741360	12741749	63
PSPH	chr7	-	17469	56101590	56102009	27
PTBP1	chr19	+	3248	800430	800849	34
PTBP2	chr1	+	-332400	96854490	96855059	84
PTGDR	chr14	+	11944	52746150	52746599	34
PTGER4	chr5	+	-182162	40497660	40498079	31
PTGER4	chr5	+	-297827	40382040	40382369	25
PTGR2	chr14	+	916	74319300	74319599	29
PTHLH	chr12	-	-10118	28134780	28135289	20
PTK2	chr8	-	273753	141737370	141737789	44
PTPLAD1	chr15	+	-4732	65817870	65818319	26
PTPN1	chr20	+	30659	49157310	49157789	29
PTPN21	chr14	-	32829	88988100	88988489	52
PTPRC	chr1	+	33098	198640920	198641549	32
PTPRC	chr1	+	633698	199241700	199241969	23
PTPRC	chr1	+	30293	198638280	198638579	21
PTPRC	chr1	+	42923	198650850	198651269	21
PTPRE	chr10	+	21155	129726270	129726689	25
PTPRJ	chr11	+	74885	48076740	48077249	81
PTPRJ	chr11	+	-42415	47959470	47959919	43
PTPRJ	chr11	+	-35815	47966070	47966519	30
PTPRJ	chr11	+	-62155	47939790	47940119	23
PTPRJ	chr11	+	81005	48082980	48083249	20
PTPRJ	chr11	+	75470	48077430	48077729	19
PTPRK	chr6	-	548485	128293170	128293499	35
PTPRM	chr18	+	-149439	7417650	7418099	45
PTPRM	chr18	+	558306	8125440	8125799	26
PTPRM	chr18	+	289521	7856700	7856969	18
PTPRQ	chr12	+	-2316	80835660	80835959	21
PTPRS	chr19	-	53360	5287170	5287739	26
PTS	chr11	+	15297	112112160	112112609	28
PVT1	chr8	+	15291	128821830	128822309	62
PVT1	chr8	+	332811	129139410	129139769	21
PVT1	chr8	+	382386	129189030	129189299	21
PVT1	chr8	+	131736	128938350	128938679	19
PVT1	chr8	+	224286	129030900	129031229	17
PXK	chr3	+	51138	58369590	58369919	22
PXMP3	chr8	-	-118630	78031020	78031289	18
PXN	chr12	-	2398	120685350	120685769	48

PYGL	chr14	-	-11286	51422400	51422669	20
PYGL	chr14	-	7074	51403980	51404369	15
QDPR	chr4	-	21323	17492400	17492669	18
QPCT	chr2	+	21067	37592640	37592999	17
RAB10	chr2	+	-21689	26235150	26235419	47
RAB11A	chr15	+	8439	66170100	66170369	19
RAB11FIP5	chr2	-	9707	73330290	73330589	17
RAB20	chr13	-	36952	111176940	111177299	25
RAB27A	chr15	-	-9545	55571850	55572209	23
RAB28	chr4	-	7185	13478640	13478969	31
RAB28	chr4	-	-22725	13508550	13508879	29
RAB38	chr11	-	125045	87783360	87783749	37
RAB5A	chr3	+	25388	20013750	20014169	25
RAB8B	chr15	+	55527	63537060	63537449	18
RAB9P1	chr5	+	1213020	105648000	105648389	27
RAD51AP2	chr2	-	215002	17484480	17484929	31
RAD51L1	chr14	+	668176	68954460	68954909	56
RAD51L1	chr14	+	545716	68832030	68832419	39
RAD51L1	chr14	+	745561	69031830	69032309	32
RAD51L1	chr14	+	649141	68935470	68935829	23
RAD51L1	chr14	+	327151	68613510	68613809	22
RAD51L1	chr14	+	649771	68936130	68936429	21
RAG2	chr11	-	-1123412	37743030	37743419	38
RAI14	chr5	+	-243403	34412760	34413299	33
RALB	chr2	+	38791	121049070	121049339	18
RALGPS1	chr9	+	161372	129838200	129838649	63
RALYL	chr8	+	335202	85430490	85430819	18
RANBP17	chr5	+	431288	170720100	170720519	25
RAP1B	chr12	+	-45287	68959200	68959529	25
RAP1GDS1	chr4	+	169868	99352230	99352559	20
RAP2B	chr3	+	98746	152978490	152979059	72
RAP2B	chr3	+	94306	152974200	152974469	17
RAPGEF1	chr9	-	-24455	134609550	134609819	18
RAPGEF2	chr4	+	-38528	160150290	160150649	32
RAPGEF2	chr4	+	149347	160338180	160338509	24
RAPGEF4	chr2	+	47470	173647770	173648219	43
RARB	chr3	+	145311	25614870	25615259	33
RARB	chr3	+	26511	25496100	25496429	32
RARRES3	chr11	+	-13363	63290760	63291059	21
RASAL3	chr19	-	-2792	15578010	15578339	23
RASGRP1	chr15	-	-28887	38885670	38886119	40
RASSF10	chr11	+	-28576	13001910	13002329	17
RBM15	chr1	+	-2470	110879280	110879669	30
RBM19	chr12	-	177947	114226080	114226379	24
RBM19	chr12	-	62582	114341430	114341759	22
RBMS1	chr2	-	-258901	161609010	161609429	47
RBMS1	chr2	-	-38956	161389140	161389409	21
RBMS1	chr2	-	-253996	161604090	161604539	17
RBMS3	chr3	+	92042	29414850	29415119	19
RBPJ	chr4	+	-122362	26198730	26199209	16
RCC2	chr1	-	-37377	17802240	17802629	33

RCL1	chr9	+	73541	4866240	4866509	19
RCSD1	chr1	+	-2074	167597160	167597639	44
RCSD1	chr1	+	59501	167658840	167659109	22
RDH14	chr2	-	327450	18414300	18414719	29
RECQL5	chr17	-	21621	73641480	73641809	27
REEP3	chr10	+	190892	65471790	65472239	56
REEP3	chr10	+	-5698	65275260	65275589	34
REEP3	chr10	+	161207	65442180	65442479	18
RELN	chr7	-	270589	103359150	103359599	63
RELN	chr7	-	330739	103299060	103299389	14
REPIN1	chr7	+	-5204	150060480	150060869	36
RFC1	chr4	-	17221	39350640	39350909	17
RFC2	chr7	-	20359	73648200	73648559	34
RFC3	chr13	+	226129	34618140	34618529	29
RFC3	chr13	+	258394	34650450	34650749	22
RFC3	chr13	+	338089	34730160	34730429	18
RFESD	chr5	+	-111	94982160	94982579	43
RFTN1	chr3	-	-10612	16565670	16565999	21
RFX2	chr19	-	49480	6060960	6061409	45
RFX3	chr9	-	202119	3323700	3324029	23
RFX4	chr12	+	46752	107023590	107023979	45
RGNEF	chr5	+	-20273	72901470	72901949	37
RGNEF	chr5	+	12742	72934590	72934859	18
RGS13	chr1	+	66223	192671340	192671669	25
RGS17	chr6	-	-433720	153885930	153886289	25
RGS3	chr9	+	133319	116340150	116340509	25
RGS3	chr9	+	20894	116227770	116228039	21
RGS5	chr1	-	34688	163137960	163138409	50
RHAG	chr6	-	-31522	49635900	49636319	48
RHBDD1	chr2	+	2332	227702940	227703269	22
RHBDD2	chr7	+	1068	75509220	75509549	28
RHOB	chr2	+	63680	20710350	20710679	31
RHOBTB3	chr5	+	63270	95129910	95130329	34
RHOBTB3	chr5	+	-12810	95053860	95054219	26
RHOJ	chr14	+	72645	63743640	63743939	22
RHOT1	chr17	+	14922	30484200	30484589	34
RICS	chr11	-	-37496	128931420	128931749	22
RIMS2	chr8	+	704329	105217050	105217559	58
RIN2	chr20	+	-105355	19764660	19765049	42
RIOK1	chr6	+	16540	7419270	7419569	23
RIOK2	chr5	-	-656274	97175070	97175489	39
RIPK2	chr8	+	-190075	90579690	90580109	50
RIPPLY2	chr6	+	-37550	84525270	84525599	44
RLF	chr1	+	-20961	40605870	40606289	47
RMST	chr12	+	91621	97950210	97950629	38
RNASE7	chr14	+	2270	21512430	21512879	40
RNF103	chr2	-	2119	86848620	86849099	46
RNF144A	chr2	+	29572	7086960	7087229	19
RNF145	chr5	-	32095	158602590	158602889	18
RNF152	chr18	-	10395	59549730	59550089	27
RNF170	chr8	-	2362	42749310	42749699	43

RNF182	chr6	+	67817	13992870	13993169	25
RNF216	chr7	-	139608	5681550	5681819	22
RNF220	chr1	+	18715	44889450	44889899	37
RNF220	chr1	+	224170	45094950	45095309	26
RNF220	chr1	+	-14795	44856030	44856299	20
RNF43	chr17	-	15267	56479470	56479859	31
RNF43	chr17	-	25857	56468910	56469239	26
RNLS	chr10	-	55008	90287820	90288329	20
RNU5E	chr5	+	223408	80724720	80725019	21
ROBO1	chr3	-	-853295	79921740	79922069	30
ROBO1	chr3	-	512815	78555660	78555929	17
ROBO2	chr3	+	-166159	76923000	76923269	20
ROD1	chr9	-	91920	115003830	115004219	29
ROD1	chr9	-	-8055	115103850	115104149	25
ROPN1L	chr5	+	72526	10514400	10514669	24
ROPN1L	chr5	+	11806	10453680	10453949	21
ROR1	chr1	+	146595	64386030	64386539	77
ROR1	chr1	+	99870	64339380	64339739	20
ROR2	chr9	-	4680	94707630	94707899	22
RORB	chr9	+	59223	77171250	77171699	35
RPRD1B	chr20	+	60902	36722670	36723029	33
RPS24	chr10	+	222632	80015970	80016329	29
RPS27A	chr2	+	-74760	55384140	55384469	27
RPS29	chr14	-	80805	49972080	49972499	32
RPS6KA2	chr6	-	-99038	167139570	167139959	23
RPS6KA2	chr6	-	-51338	167091900	167092229	20
RPTOR	chr17	+	316770	78835260	78835529	16
RPTOR	chr17	+	660	78519120	78519449	13
RRP15	chr1	+	-259719	218198730	218199089	20
RSPO3	chr6	+	-294468	127145370	127145789	21
RSPO4	chr20	-	28140	954630	954899	23
RTKN2	chr10	-	33682	63994650	63994919	17
RTN4	chr2	-	-87719	55325040	55325339	17
RTP3	chr3	+	-7835	46531470	46531829	28
RUNX1	chr21	-	42708	36218130	36218429	26
RUNX2	chr6	+	70336	45366240	45366539	22
RUSC2	chr9	+	2693	35492490	35492909	30
RUSC2	chr9	+	28073	35517930	35518229	23
RXRA	chr9	+	34624	137252760	137253119	26
S100A11	chr1	-	9967	151999380	151999709	17
S1PR1	chr1	+	134630	101836680	101837189	48
SALL4	chr20	-	4679	50414190	50414549	22
SAMD12	chr8	-	284330	119349660	119350049	33
SAMD12	chr8	-	181895	119452110	119452469	31
SAMD12	chr8	-	398630	119235420	119235689	25
SAMD12	chr8	-	296210	119337840	119338109	18
SAMD13	chr1	+	-50529	84713340	84713699	59
SAMD13	chr1	+	-44724	84719190	84719459	17
SAMD4A	chr14	+	97338	55131690	55132259	97
SAMD4A	chr14	+	169713	55204140	55204559	34
SAMD4A	chr14	+	83133	55117590	55117949	31

SAMD4A	chr14	+	208893	55243320	55243739	30
SASH1	chr6	+	-277654	148385910	148386239	33
SATB1	chr3	-	-233025	18699720	18699989	17
SAV1	chr14	-	-7476	51142230	51142769	27
SBF2	chr11	-	387555	9927960	9928439	49
SBF2	chr11	-	303405	10012140	10012559	22
SCAMP5	chr15	+	19734	75307500	75307769	19
SCG2	chr2	-	113382	224353500	224353979	35
SCGB1D4	chr11	-	-868	62067270	62067539	18
SCIN	chr7	+	77232	12687270	12687599	19
SCMH1	chr1	-	-91421	41799060	41799359	20
SCN4B	chr11	-	5605	118010400	118010759	22
SCRN2	chr17	-	-3095	45921600	45921989	44
SCUBE3	chr6	+	-5855	35176140	35176529	18
SDC2	chr8	+	13933	97519650	97519979	15
SDC2	chr8	+	-91802	97413930	97414229	14
SDC4	chr20	-	6230	43970700	43970969	19
SDCCAG1	chr14	-	-285	50319660	50319989	28
SDPR	chr2	-	-80078	192791880	192792239	26
SEC23A	chr14	-	154943	39417330	39417659	34
SEC23A	chr14	-	110693	39461550	39461939	23
SEC23A	chr14	-	130073	39442230	39442499	18
SEC61A1	chr3	+	27843	127798920	127799189	20
SELE	chr1	-	11146	169691880	169692269	26
SEMA3A	chr7	-	169173	83654850	83655239	37
SEMA3A	chr7	-	-443217	84267240	84267629	35
SEMA3C	chr7	-	-240267	80788620	80789249	75
SEMA3C	chr7	-	-365982	80914470	80914829	41
SEMA3D	chr7	-	-14172	84765150	84765689	53
SEMA3D	chr7	-	139503	84611610	84611879	22
SEMA3D	chr7	-	-202602	84953700	84953999	17
SEMA3E	chr7	-	120020	83158110	83158499	20
SEMA3E	chr7	-	124880	83153310	83153579	16
SEMA3F	chr3	+	9047	50201760	50202029	21
SEMA6A	chr5	-	-289288	116199660	116200019	25
SEMA6D	chr15	+	94059	48104550	48104939	45
SERP1	chr3	-	48474	150215820	150216089	19
SERTAD2	chr2	-	-115708	64996620	64996889	24
SETBP1	chr18	+	-28553	42231420	42231749	27
SETBP1	chr18	+	109747	42369750	42370019	21
SETD2	chr3	-	145758	47059530	47059889	18
SFRP1	chr8	-	188916	40977930	40978199	20
SFRP2	chr4	-	-5096	154715130	154715519	27
SFRS3	chr6	+	22790	36584670	36585089	26
SFTA1P	chr10	-	655718	10180980	10181339	27
SFTA1P	chr10	-	-1912	10838640	10838939	25
SFTA1P	chr10	-	468608	10368120	10368419	24
SFTA1P	chr10	-	7163	10829580	10829849	22
SFXN3	chr10	+	1889	102792660	102793109	61
SFXN5	chr2	-	50876	73247880	73248299	41
SGCE	chr7	-	12307	94273080	94273349	19

SGIP1	chr1	+	30255	67029900	67030259	15
SGK1	chr6	-	-365110	134860920	134861369	71
SGMS1	chr10	-	-12762	52396290	52396709	46
SGMS1	chr10	-	111993	52271460	52272029	40
SGMS1	chr10	-	-28182	52411740	52412099	32
SGPL1	chr10	+	33956	72609450	72609869	29
SGPL1	chr10	+	791	72576360	72576629	21
SH3RF1	chr4	-	-12210	170204220	170204699	64
SH3RF3	chr2	+	7158	109752990	109753319	25
SH3RF3	chr2	+	290658	110036520	110036789	18
SHANK2	chr11	-	164584	70343100	70343579	39
SHROOM3	chr4	+	313732	77669820	77670149	39
SIGLEC10	chr19	-	9212	51912030	51912449	21
SIM1	chr6	-	166272	100745070	100745489	49
SIPA1L1	chr14	+	130033	72125940	72126209	16
SIRPA	chr20	+	-81008	1793640	1793969	29
SIX2	chr2	-	-253897	45490260	45490619	26
SKAP1	chr17	-	183440	46323960	46324349	35
SKAP1	chr17	-	185720	46321680	46322069	31
SLAMF1	chr1	-	6757	160610190	160610459	24
SLAMF7	chr1	+	8033	160716930	160717289	24
SLC12A3	chr16	+	-3774	56895090	56895599	26
SLC13A1	chr7	-	-42839	122882700	122883029	21
SLC14A1	chr18	+	-10777	43293150	43293479	24
SLC16A7	chr12	+	47689	60130650	60130979	28
SLC18A1	chr8	-	96388	19944150	19944509	37
SLC18A1	chr8	-	63463	19977090	19977419	25
SLC1A3	chr5	+	60863	36667080	36667559	34
SLC1A4	chr2	+	8726	65224110	65224499	28
SLC1A5	chr19	-	-7360	47295270	47295719	64
SLC1A5	chr19	-	-1900	47289900	47290169	21
SLC1A7	chr1	-	34830	53573190	53573729	49
SLC1A7	chr1	-	-5775	53613930	53614199	22
SLC20A1	chr2	+	-14897	113388480	113388779	22
SLC25A21	chr14	-	350021	37291620	37292069	23
SLC25A21	chr14	-	471431	37170240	37170629	18
SLC25A29	chr14	-	3671	100769010	100769369	26
SLC25A46	chr5	+	96046	110170620	110170979	30
SLC25A46	chr5	+	46036	110120610	110120969	22
SLC28A1	chr15	+	45522	85473240	85473629	16
SLC30A1	chr1	-	-13565	211765500	211765829	23
SLC30A10	chr1	-	217534	219884310	219884609	27
SLC34A2	chr4	+	38825	25696110	25696409	24
SLC35F3	chr1	+	235186	234275700	234276029	25
SLC35F3	chr1	+	76441	234116970	234117269	20
SLC38A1	chr12	-	49404	46613610	46613999	28
SLC38A2	chr12	-	-1599	46768050	46768439	22
SLC38A4	chr12	-	55296	47164290	47164679	44
SLC38A4	chr12	-	164721	47054850	47055269	39
SLC38A6	chr14	+	38515	61486200	61486589	22
SLC39A11	chr17	-	390254	70698420	70698779	32

SLC39A11	chr17	-	431174	70657470	70657889	30
SLC44A1	chr9	+	-1424	108005280	108005729	54
SLC44A2	chr19	+	2414	10715370	10715699	31
SLC44A3	chr1	+	34659	95320320	95320799	41
SLC44A3	chr1	+	52914	95338650	95338979	25
SLC44A3	chr1	+	-22716	95263020	95263349	18
SLC45A1	chr1	+	-141755	8242380	8242889	57
SLC46A2	chr9	-	-8056	115661100	115661399	25
SLC4A2	chr7	+	8718	150765210	150765539	23
SLC4A3	chr2	+	906163	221398230	221398679	76
SLC4A3	chr2	+	597733	221089800	221090249	65
SLC4A3	chr2	+	645748	221137860	221138219	30
SLC4A4	chr4	+	258757	72311610	72311909	27
SLC4A4	chr4	+	2242	72055110	72055379	15
SLC9A1	chr1	-	13017	27468300	27468569	21
SLC9A11	chr1	-	12304	173559720	173560139	62
SLCO2A1	chr3	-	21386	133727310	133727759	47
SLCO4C1	chr5	-	157329	101474700	101475149	49
SLCO6A1	chr5	-	-7524	101842050	101842439	19
SLFN5	chr17	+	14464	33584310	33584789	54
SLIT2	chr4	+	-443430	19811670	19811939	25
SLIT2	chr4	+	-270810	19984290	19984559	19
SLIT3	chr5	-	448954	168278940	168279419	45
SLIT3	chr5	-	-73811	168801750	168802139	31
SLIT3	chr5	-	-53381	168781380	168781649	21
SLMAP	chr3	+	95671	57838710	57838979	17
SLMO2	chr20	-	-5453	57623160	57623549	26
SLN	chr11	-	15718	107566860	107567279	29
SMAD3	chr15	+	-16220	67341690	67342259	45
SMAD3	chr15	+	113095	67471080	67471499	30
SMAD3	chr15	+	82525	67440570	67440869	21
SMAD3	chr15	+	41020	67399080	67399349	19
SMAD6	chr15	+	126665	67127460	67127729	24
SMAD6	chr15	+	154655	67155450	67155719	17
SMAD7	chr18	-	-46498	46523430	46523729	27
SMAD7	chr18	-	2237	46474710	46474979	21
SMARCA2	chr9	+	7378	2022540	2022899	30
SMARCA1	chr4	+	-76144	95052450	95052779	24
SMARCC2	chr12	-	-7538	56590740	56591039	30
SMCHD1	chr18	+	100049	2755770	2756099	20
SMG6	chr17	-	164820	2042040	2042459	61
SMPD3	chr16	-	30420	68451810	68452169	29
SMURF2	chr17	-	-44748	62702880	62703389	51
SMYD3	chr1	-	430367	246150120	246150569	40
SMYD3	chr1	-	645152	245935380	245935739	32
SMYD3	chr1	-	277397	246303150	246303479	22
SNAP25	chr20	+	88423	10287720	10288079	36
SNAR-H	chr2	-	-354592	78536610	78536879	18
SNCA	chr4	-	36883	90721110	90721379	18
SNIP	chr17	-	13969	36748050	36748379	31
SNRPC	chr6	+	-12117	34713030	34713359	17

SNTB1	chr8	-	-342500	122166600	122167019	37
SNTB1	chr8	-	-353270	122177370	122177789	32
SNTB1	chr8	-	-333605	122157750	122158079	28
SNTB1	chr8	-	-339080	122163210	122163569	21
SNX13	chr7	-	95467	17884530	17884799	19
SNX16	chr8	-	-860863	83615160	83615609	68
SNX19	chr11	-	68688	130717500	130717889	27
SNX19	chr11	-	87498	130698720	130699049	21
SNX25	chr4	+	59396	186190470	186190889	49
SNX6	chr14	-	-36039	35135220	35135489	19
SNX7	chr1	+	-345201	98781900	98782169	18
SOCS1	chr16	-	54395	11295480	11295809	18
SOCS5	chr2	+	-7944	46917960	46918349	42
SORBS1	chr10	-	-44242	97245000	97245359	27
SORBS3	chr8	+	-2446	22406640	22406969	33
SORCS3	chr10	+	944706	107345340	107345789	62
SORCS3	chr10	+	1193706	107594430	107594699	19
SOX11	chr2	+	-580909	5251740	5252039	19
SOX13	chr1	+	31674	204073740	204074099	29
SOX2OT	chr3	+	-151401	181176540	181176959	35
SOX5	chr12	-	168279	23569020	23569439	27
SOX9	chr17	+	149694	70266660	70267049	38
SOX9	chr17	+	-488541	69628440	69628799	33
SOX9	chr17	+	270294	70387110	70387799	25
SOX9	chr17	+	-931836	69185190	69185459	22
SOX9	chr17	+	-492471	69624510	69624869	17
SP3	chr2	-	20883	174807870	174808259	25
SPAG9	chr17	-	35217	49162800	49163219	22
SPARC	chr5	-	1773	151064610	151064879	21
SPATA7	chr14	+	-39292	88812570	88812869	19
SPATA9	chr5	-	-15480	95034000	95034389	77
SPATA9	chr5	-	-3825	95022300	95022779	52
SPATS2	chr12	+	37212	49797660	49798139	16
SPCS3	chr4	+	119915	177360840	177361169	35
SPCS3	chr4	+	116705	177357630	177357959	23
SPCS3	chr4	+	-595	177240360	177240629	18
SPEF2	chr5	+	193371	35811150	35811569	35
SPEF2	chr5	+	-128979	35488830	35489189	25
SPG20	chr13	-	4672	36915840	36916109	18
SPINK5	chr5	+	-52035	147391290	147391709	29
SPIRE2	chr16	+	14403	89909070	89909549	51
SPOCK1	chr5	-	103809	136731060	136731359	27
SPP2	chr2	+	208134	235167330	235167629	20
SPRED2	chr2	-	-163893	65757660	65757989	19
SPRY4	chr5	-	-96439	141800820	141801299	49
SPRY4	chr5	-	-104539	141808950	141809369	41
SPRY4	chr5	-	-94	141704550	141704879	29
SPRY4	chr5	-	-108064	141812550	141812819	24
SPTB	chr14	-	11367	65278350	65278649	33
SPTBN1	chr2	+	26856	54710130	54710489	24
SPTBN1	chr2	+	105876	54789180	54789479	18

SPTBN1	chr2	+	118101	54801420	54801689	18
SRBD1	chr2	-	175689	45662550	45662939	54
SRGAP1	chr12	+	76644	64315050	64315319	18
SSFA2	chr2	+	-100812	182655420	182655899	25
SSH2	chr17	-	121354	28135530	28135799	18
SSPN	chr12	+	79034	26427270	26427809	41
SSPN	chr12	+	60809	26409150	26409479	17
SSTR4	chr20	+	-113247	22902660	22902959	23
ST3GAL1	chr8	-	3029	134580990	134581319	23
ST3GAL3	chr1	+	150927	44323980	44324309	30
ST3GAL6	chr3	+	8698	98460000	98460539	20
ST6GALNAC3	chr1	+	199821	76740060	76740359	15
ST7	chr7	+	179144	116772240	116772809	44
STAC	chr3	+	-18117	36403830	36404129	24
STAM	chr10	+	11716	17697690	17697989	53
STAM	chr10	+	29506	17715480	17715779	25
STAMBPL1	chr10	+	2499	90642360	90642689	22
STARD13	chr13	-	65747	33694290	33694649	25
STARD13	chr13	-	-32188	33792210	33792599	25
STARD4	chr5	-	-18177	110866200	110866469	18
STAT1	chr2	-	35682	191843100	191843489	39
STAT3	chr17	-	40991	40499220	40499609	42
STAT4	chr2	-	-19279	192035040	192035369	23
STC2	chr5	-	34687	172721550	172722089	50
STC2	chr5	-	-122318	172878630	172879019	28
STK31	chr7	+	153028	23902830	23903219	31
STK31	chr7	+	141943	23891790	23892089	19
STK31	chr7	+	271213	24021060	24021359	17
STK39	chr2	-	36391	169067460	169067969	62
STK40	chr1	-	-289	36851610	36851939	30
STOML1	chr15	-	22726	74261760	74262059	28
STX18	chr4	-	-14049	4557600	4558049	48
STX3	chr11	+	20526	59543250	59543579	16
STYX	chr14	+	25937	53222670	53222969	15
SUCLG2	chr3	-	49354	67655520	67655849	25
SUCNR1	chr3	+	145893	151737180	151737479	20
SULT1C4	chr2	+	35284	109029540	109029869	22
SULT6B1	chr2	-	6606	37408830	37409339	100
SUMO1P1	chr20	-	47149	52444890	52445309	28
SUSD1	chr9	-	109242	114828090	114828539	68
SWAP70	chr11	+	112267	9797760	9798029	19
SYDE2	chr1	-	54004	85612530	85612919	26
SYDE2	chr1	-	-1691	85668180	85668659	24
SYDE2	chr1	-	-2966	85669530	85669859	23
SYN3	chr22	-	245127	33157500	33157859	30
SYNJ2	chr6	+	58626	158461410	158461679	48
SYPL1	chr7	-	71069	105681540	105681839	23
SYT1	chr12	+	194557	79452120	79452539	50
SYT1	chr12	+	560212	79817790	79818179	38
SYT1	chr12	+	-159653	79097940	79098299	30
TACR1	chr2	-	99031	75327420	75327809	21

TACSTD2	chr1	-	-43708	59086710	59087039	31
TACSTD2	chr1	-	-51613	59094600	59094959	30
TAF1L	chr9	-	-37032	32672400	32672999	31
TANK	chr2	+	-203841	161789490	161789759	21
TAOK3	chr12	-	15506	118795050	118795439	32
TARS	chr5	+	-119448	33321270	33321629	30
TBC1D12	chr10	+	-4506	96157410	96157949	69
TBC1D12	chr10	+	6834	96168870	96169169	27
TBC1D12	chr10	+	1539	96163590	96163859	21
TBC1D22A	chr22	+	332261	47490660	47490959	26
TBC1D22A	chr22	+	574736	47733090	47733479	25
TBC1D22B	chr6	+	32322	37257720	37258019	22
TBC1D7	chr6	-	-27064	13355640	13356029	28
TBC1D8	chr2	-	30872	101736810	101737139	17
TBCC	chr6	-	-31480	42745200	42745529	21
TBL1XR1	chr3	-	-612981	177527850	177528209	29
TBPL2	chr14	-	13049	55893960	55894469	65
TBPL2	chr14	-	-13891	55920930	55921379	27
TBX15	chr1	-	327525	119204490	119204819	21
TBX18	chr6	-	-133890	85607610	85607969	21
TBX19	chr1	+	116067	168366180	168366509	37
TBX3	chr12	-	-412845	115534620	115535009	33
TBX3	chr12	-	-235575	115357410	115357679	18
TCF12	chr15	+	60862	57271470	57271919	44
TCF12	chr15	+	45757	57256380	57256799	23
TCF7L1	chr2	+	-26619	85333890	85334339	48
TCF7L2	chr10	+	136201	114846030	114846389	26
TCF7L2	chr10	+	6196	114716070	114716339	18
TCN1	chr11	-	8352	59625450	59625929	32
TCN1	chr11	-	-32268	59666130	59666489	27
TCP10L	chr21	-	63996	33893670	33894029	29
TDG	chr12	+	3747	104363190	104363489	22
TDRG1	chr6	+	-25293	40320660	40321079	21
TEAD1	chr11	+	140596	12836400	12836729	27
TECRL	chr4	-	691839	64583160	64583519	39
TES	chr7	+	50350	115913250	115913519	17
TET2	chr4	+	-58488	106009290	106009619	17
TEX2	chr17	-	60594	62279910	62280209	22
TEX9	chr15	+	806	56658300	56658599	16
TFAP2A	chr6	-	64453	10347960	10348349	31
TFEC	chr7	-	358524	115312140	115312409	19
TFPI	chr2	-	49760	188369280	188369639	35
TFPI	chr2	-	7415	188411640	188411969	22
TFPI2	chr7	-	153496	93366360	93366779	48
TG	chr8	+	166180	134045220	134045549	20
TGFB2	chr1	+	147024	218666160	218666669	84
TGFB2	chr1	+	246804	218765910	218766479	61
TGFB2	chr1	+	134859	218654070	218654429	28
TGFB2	chr1	+	8004	218527260	218527529	18
TGFB2	chr1	+	150099	218669340	218669639	16
TGFB2	chr3	+	4671	30652470	30652859	39

TGFBR2	chr3	+	-258759	30389040	30389429	25
TGFBR2	chr3	+	20481	30668310	30668639	23
TGFBR2	chr3	+	77496	30725250	30725729	23
TGFBR3	chr1	-	90028	92261460	92262059	45
TGFBR3	chr1	-	-26222	92377860	92378159	25
TGIF1	chr18	+	-44227	3367710	3367979	15
TGIF2	chr20	+	18695	35220570	35220839	25
TGM2	chr20	-	47496	36745950	36746459	60
TGM2	chr20	-	55476	36737970	36738479	55
TGM2	chr20	-	55011	36738510	36738869	24
TGM2	chr20	-	30486	36763080	36763349	20
THADA	chr2	-	169364	43653540	43653959	36
THADA	chr2	-	127814	43695150	43695449	18
THOC3	chr5	-	46356	175348950	175349429	17
THRB	chr3	-	151172	24384900	24385289	42
THRB	chr3	-	-15478	24551550	24551939	36
THRB	chr3	-	-19513	24555630	24555929	20
THSD4	chr15	+	153487	71586900	71587649	30
THSD4	chr15	+	154027	71587680	71587949	15
THSD7B	chr2	+	444573	138192810	138193259	44
THSD7B	chr2	+	189333	137937660	137937929	18
THUMPD2	chr2	-	-315983	40322160	40322639	43
TIAM2	chr6	+	111832	155523120	155523389	19
TIGD2	chr4	+	23827	90057660	90057929	32
TIMM22	chr17	+	4578	904710	905159	37
TJP1	chr15	-	-2878	30117420	30117749	29
TJP1	chr15	-	-31948	30146490	30146819	19
TJP2	chr9	+	6731	71795580	71796059	66
TJP2	chr9	+	4496	71793360	71793809	38
TJP2	chr9	+	49541	71838480	71838779	29
TLE1	chr9	-	2202	84301200	84301589	41
TLE3	chr15	-	5937	70384170	70384469	44
TLE3	chr15	-	109992	70280070	70280459	19
TLL1	chr4	+	32290	166826520	166826879	30
TLN2	chr15	+	92470	63031710	63032249	59
TLN2	chr15	+	94855	63034200	63034529	18
TLR9	chr3	-	8955	52251030	52251419	25
TM2D1	chr1	-	228781	61962180	61962449	37
TMED10	chr14	-	19530	75623670	75623969	29
TMEM111	chr3	-	16263	10012050	10012469	32
TMEM131	chr2	-	45105	98567070	98567429	44
TMEM135	chr11	+	-37740	86711040	86711609	81
TMEM135	chr11	+	-32550	86716350	86716679	28
TMEM135	chr11	+	137490	86886390	86886719	20
TMEM156	chr4	-	-3728	39037620	39037919	18
TMEM158	chr3	-	7685	45259950	45260309	37
TMEM158	chr3	-	2300	45265320	45265709	33
TMEM163	chr2	-	227092	135249330	135249629	18
TMEM196	chr7	-	-185070	19997280	19997669	27
TMEM20	chr10	+	-20405	95633190	95633459	16
TMEM200A	chr6	+	163988	130922040	130922459	33

TMEM200A	chr6	+	121223	130879350	130879619	20
TMEM200A	chr6	+	45863	130803960	130804289	17
TMEM207	chr3	-	2991	190164540	190164809	19
TMEM211	chr22	-	-30705	25365810	25366229	42
TMEM26	chr10	-	169079	63043950	63044309	31
TMEM26	chr10	-	14684	63198360	63198689	21
TMEM39B	chr1	+	28492	32566830	32567159	25
TMEM44	chr3	-	71849	194282100	194282429	17
TMEM45A	chr3	+	8157	100219470	100219769	61
TMEM45A	chr3	+	40722	100252020	100252349	17
TMEM48	chr1	-	39156	54264810	54265229	30
TMEM49	chr17	+	125532	57910230	57910559	26
TMEM64	chr8	-	99949	91557900	91558469	96
TMEM64	chr8	-	108994	91548960	91549319	25
TMEM72	chr10	+	-49764	45356790	45357209	63
TMEM72	chr10	+	-186339	45220230	45220619	31
TMOD1	chr9	+	40328	100326600	100326929	34
TMOD3	chr15	+	-564	52121130	52121519	25
TMPO	chr12	+	-512529	98396640	98397119	52
TMTC2	chr12	+	-131249	82949490	82949879	22
TMTC3	chr12	+	157502	88693380	88693769	26
TMX1	chr14	+	102214	51808890	51809309	41
TMX3	chr18	-	536764	65845410	65845769	20
TNFAIP3	chr6	+	-123691	138064710	138065069	20
TNFAIP8	chr5	+	18987	118623240	118623569	22
TNFRSF10D	chr8	-	2706	23018610	23019059	19
TNFRSF11B	chr8	-	155169	119809050	119809379	17
TNFSF11	chr13	+	-75637	43061040	43061429	54
TNFSF18	chr1	-	116844	172903020	172903499	66
TNFSF18	chr1	-	81204	172938690	172939109	57
TNFSF18	chr1	-	95394	172924530	172924889	26
TNFSF18	chr1	-	45249	172974720	172974989	25
TNFSF18	chr1	-	-11451	173031420	173031689	18
TNFSF4	chr1	-	-62188	173238420	173238899	27
TNFSF8	chr9	-	22136	117670380	117670889	67
TNIP1	chr5	-	-12342	150473160	150473519	27
TNNI3K	chr1	+	164198	74827980	74828309	31
TNNI3K	chr1	+	155078	74818890	74819159	20
TNP1	chr2	-	-358897	218083530	218083829	14
TNPO2	chr19	-	-1884	12834780	12835199	18
TNRC18	chr7	-	4948	5458080	5458379	17
TNS4	chr17	-	-15340	38673000	38673389	31
TOP1	chr20	+	49803	39707070	39707459	43
TOPBP1	chr3	-	16953	133363590	133363979	27
TOX	chr8	-	262918	59768640	59769059	35
TOX	chr8	-	268453	59763150	59763479	30
TOX	chr8	-	239593	59792010	59792339	23
TOX2	chr20	+	35553	42578790	42579299	57
TPH2	chr12	+	61094	72393570	72393869	28
TPM1	chr15	+	2947	63337530	63338039	51
TPM1	chr15	+	-196688	63137940	63138359	29

TPM1	chr15	+	-145628	63189060	63189359	16
TPRG1	chr3	+	153032	189042630	189042959	32
TPRG1	chr3	+	152552	189042150	189042479	21
TRAK1	chr3	+	67069	42199680	42199949	21
TRAK2	chr2	-	19450	202296720	202297019	20
TRAM2	chr6	-	29253	52412430	52412789	25
TRAP1	chr16	-	20539	3746910	3747209	20
TRAPPC9	chr8	-	-8103	141475800	141476129	22
TRDMT1	chr10	-	2007	17241540	17241809	18
TRERF1	chr6	-	-7146	42426780	42427079	23
TRHR	chr8	+	-138416	109961130	109961489	21
TRIB1	chr8	+	250302	126692700	126693029	25
TRIB1	chr8	+	331767	126774180	126774479	18
TRIM44	chr11	+	133967	35818110	35818529	21
TRIM9	chr14	-	97873	51464400	51464699	30
TRIO	chr5	+	124756	14268300	14268869	52
TRIO	chr5	+	-25124	14118510	14118899	29
TRIO	chr5	+	123811	14267490	14267789	17
TRPC6	chr11	-	152985	101301480	101301869	30
TRPM1	chr15	-	76430	31317270	31317719	32
TRPM6	chr9	-	83856	77418960	77419349	28
TRPS1	chr8	-	-23136	116704080	116704649	97
TRPS1	chr8	-	-368406	117049440	117049829	40
TRPS1	chr8	-	-309381	116990430	116990789	24
TRPS1	chr8	-	-190146	116871180	116871569	23
TSC22D2	chr3	+	35572	150162180	150162539	26
TSC22D2	chr3	+	38902	150165480	150165899	24
TSC22D2	chr3	+	40402	150167040	150167339	19
TSEN15	chr1	+	108899	184129500	184129919	28
TSEN2	chr3	+	-23056	12502740	12503009	21
TSHZ2	chr20	+	359378	51948060	51948449	37
TSHZ2	chr20	+	374123	51962820	51963179	33
TSLP	chr5	+	-155185	110252040	110252369	15
TSNAX	chr1	+	-16294	231647970	231648239	21
TSPAN14	chr10	+	52157	82266060	82266329	21
TSPAN15	chr10	+	-4411	71206590	71207039	33
TSPAN2	chr1	-	-90274	115722090	115722689	83
TSPAN2	chr1	-	26846	115605090	115605449	40
TSPAN2	chr1	-	26156	115605780	115606139	30
TSPAN5	chr4	-	-63622	99643170	99643529	34
TTC13	chr1	-	5888	231108570	231108869	31
TTC18	chr10	-	96523	75021900	75022289	29
TTC28	chr22	-	252949	28822740	28823069	30
TTC28	chr22	-	350164	28725540	28725839	18
TTC29	chr4	-	-214810	148081710	148081979	17
TTLL7	chr1	-	1036289	83428350	83428739	43
TTLL7	chr1	-	946079	83518590	83518919	30
TTLL7	chr1	-	943439	83521110	83521679	27
TTLL7	chr1	-	157139	84307560	84307829	18
TUBD1	chr17	-	39622	57930510	57930839	24
TUFT1	chr1	+	8654	151521180	151521689	82

TUFT1	chr1	+	5879	151518450	151518869	39
TUFT1	chr1	+	2489	151515120	151515419	21
TUSC3	chr8	+	21925	15419490	15419819	24
TXK	chr4	-	54059	48081990	48082439	50
TXK	chr4	-	53009	48083130	48083399	18
TXNDC2	chr18	+	13156	9898980	9899309	21
TXNIP	chr1	+	-817	145437510	145437779	16
TXNRD1	chr12	+	4416	104613780	104614169	67
TXNRD1	chr12	+	-37719	104571630	104572049	35
TXNRD1	chr12	+	82251	104691600	104692019	23
TXNRD2	chr22	-	60495	19868610	19869119	61
UACA	chr15	-	239231	70755120	70755659	53
UBA3	chr3	-	-1680	69131010	69131399	37
UBAP1	chr9	+	43219	34222050	34222409	25
UBE2D4	chr7	+	17103	43983000	43983299	24
UBE2E1	chr3	+	-53569	23793690	23794049	38
UBE2E1	chr3	+	-55009	23792250	23792609	25
UBE2E2	chr3	+	-88099	23156340	23157029	54
UBE2E2	chr3	+	-219874	23024730	23025089	31
UBE2E3	chr2	+	-218947	181626000	181626329	23
UBE2E3	chr2	+	223023	182068230	182068499	18
UBE2G1	chr17	-	62035	4207740	4208129	28
UBE2H	chr7	-	46055	129546570	129546899	28
UBE2I	chr16	+	-11985	1347060	1347329	27
UBE2J1	chr6	-	6265	90056220	90056489	20
UBL3	chr13	-	8991	30415680	30415979	26
UGCG	chr9	+	133139	114792180	114792509	33
UGCG	chr9	+	-4561	114654480	114654809	25
UGP2	chr2	+	-98948	63968970	63969329	26
UGT1A8	chr2	+	72914	234599070	234599339	19
UHRF1BP1	chr6	+	18126	34777710	34778129	26
UIMC1	chr5	-	76554	176356710	176357069	27
UNC13B	chr9	+	205611	35367360	35367839	44
UNC13B	chr9	+	178191	35340030	35340329	24
UNC5B	chr10	+	29132	73001250	73001609	23
UNC5CL	chr6	-	16654	40990140	40990409	19
UOX	chr1	-	13557	84849810	84850229	38
UPB1	chr22	+	-8711	24882330	24882749	40
UPP1	chr7	+	-21150	48107010	48107399	31
USP10	chr16	+	30770	84764100	84764549	37
USP12	chr13	-	-11260	27757140	27757439	22
USP14	chr18	+	-50123	108000	108719	13.68547419
USP14	chr18	+	-49328	108990	109319	12.78511405
USP15	chr12	+	111888	62765940	62766209	20
USP24	chr1	-	36043	55644570	55644869	24
USP25	chr21	+	59094	17161350	17161829	25
USP3	chr15	+	-7570	63789030	63789449	46
USP34	chr2	-	214965	61482720	61483049	27
USP40	chr2	-	8442	234465660	234465929	16
USP44	chr12	-	19691	95922780	95923079	19
USP49	chr6	-	48515	41814390	41814779	23

UTRN	chr6	+	268692	144881400	144881729	43
UVRAG	chr11	+	-1422	75524550	75525029	57
VANGL1	chr1	+	-144649	116039730	116040119	42
VANGL1	chr1	+	27176	116211540	116211959	40
VANGL1	chr1	+	-96139	116088300	116088569	16
VCAM1	chr1	+	-88342	101096760	101097149	28
VCAM1	chr1	+	-5122	101180040	101180309	19
VCAM1	chr1	+	80858	101265960	101266349	19
VCAN	chr5	+	92147	82859490	82859789	16
VEGFC	chr4	-	17191	177696510	177696899	33
VEGFC	chr4	-	-8624	177722370	177722669	22
VEZT	chr12	+	41198	95652570	95652869	23
VIM	chr10	+	3322	17273430	17273729	25
VLDLR	chr9	+	-210063	2411460	2411999	46
VLDLR	chr9	+	-207063	2414520	2414939	34
VPS13B	chr8	+	692726	100718010	100718429	23
VPS24	chr2	-	7141	86783310	86783609	17
VPS33B	chr15	-	-24946	91590630	91590929	30
VPS45	chr1	+	75048	150114240	150114539	22
VPS53	chr17	-	20392	597450	597959	36
VPS54	chr2	-	34620	64211460	64211729	25
VRK2	chr2	+	-164479	58109070	58109429	29
VSTM2L	chr20	+	4931	36536250	36536609	15
WAC	chr10	+	-88327	28732920	28733279	26
WAPAL	chr10	-	14102	88267200	88267679	31
WBSCR27	chr7	-	-9084	73265790	73266089	60
WBSCR28	chr7	+	38781	73314120	73314419	19
WDFY2	chr13	+	-41464	52116840	52117199	28
WDFY3	chr4	-	26324	85791660	85792019	25
WDR19	chr4	+	73076	39256920	39257279	28
WDR19	chr4	+	-12934	39170940	39171239	21
WDR27	chr6	-	130680	169971330	169971629	18
WDR33	chr2	-	40016	128528580	128528879	19
WDR36	chr5	+	35565	110463300	110463569	18
WDR37	chr10	+	5334	1107930	1108289	23
WDR78	chr1	-	62711	67327680	67328039	33
WDR89	chr14	-	-36674	64144620	64144979	43
WDSUB1	chr2	-	-4181	160147020	160147469	32
WIPF2	chr17	+	62541	38437950	38438279	24
WSB1	chr17	+	46054	25666950	25667369	22
WSCD1	chr17	+	-199789	5773950	5774339	23
WT1	chr11	-	14368	32442330	32443109	50
WWC1	chr5	+	117990	167836860	167837249	19
WWC2	chr4	+	31682	184051950	184052339	25
WWP1	chr8	+	1601	87356430	87356759	26
XCL2	chr1	-	74081	168438990	168439319	15
XRCC5	chr2	+	108335	217082220	217082489	28
XRCC6BP1	chr12	+	369275	58704450	58704989	47
XRCC6BP1	chr12	+	206915	58542150	58542569	26
XRCC6BP1	chr12	+	126125	58461420	58461719	23
XRCC6BP1	chr12	+	-36040	58299270	58299539	19

YEATS2	chr3	+	67724	183483150	183483509	37
YIPF4	chr2	+	12662	32515470	32515769	23
YPEL2	chr17	+	29907	57438810	57439109	22
YWHAZ	chr8	-	2325	101960340	101960609	18
ZAP70	chr2	+	-2406	98327400	98327849	46
ZBBX	chr3	-	544417	166553520	166553789	20
ZBTB38	chr3	+	36955	141079830	141080189	31
ZBTB38	chr3	+	-29240	141013620	141014009	23
ZC3H15	chr2	+	-683070	186667680	186667949	23
ZCCHC7	chr9	+	80341	37200570	37201049	50
ZDHHC7	chr16	-	6767	85038150	85038599	57
ZEB2	chr2	-	-574253	145851960	145852379	52
ZEB2	chr2	-	-863708	146141460	146141789	30
ZEB2	chr2	-	38452	145239330	145239599	27
ZEB2	chr2	-	-803288	146081040	146081369	27
ZEB2	chr2	-	-187928	145465710	145465979	16
ZFAND3	chr6	+	261708	38048880	38049149	17
ZFAND6	chr15	+	31329	80383170	80383529	28
ZFHX4	chr8	+	117255	77710590	77710949	37
ZFP36L1	chr14	-	164191	69095430	69095759	25
ZFP36L2	chr2	-	95451	43358070	43358519	34
ZFP36L2	chr2	-	112131	43341480	43341749	17
ZFP42	chr4	+	12725	188929440	188929859	33
ZFP42	chr4	+	-387415	188529330	188529689	27
ZFP64	chr20	-	143820	50664510	50664899	21
ZFPM1	chr16	+	1616	88521390	88521869	51
ZFPM1	chr16	+	2246	88522020	88522499	35
ZFPM1	chr16	+	476	88520310	88520669	29
ZFPM2	chr8	+	315058	106646040	106646369	21
ZG16B	chr16	+	-143	2879790	2880269	45
ZHX2	chr8	+	-91751	123701910	123702389	36
ZKSCAN1	chr7	+	19796	99632850	99633179	34
ZNF184	chr6	-	-6867	27447540	27447989	23
ZNF248	chr10	-	122387	38023950	38024249	24
ZNF281	chr1	-	78887	200300130	200300429	18
ZNF326	chr1	+	225242	90685770	90686069	18
ZNF326	chr1	+	206357	90666870	90667199	15
ZNF337	chr20	-	-42970	25720260	25720619	26
ZNF365	chr10	+	10074	64143690	64144289	87
ZNF365	chr10	+	7794	64141500	64141919	37
ZNF385B	chr2	-	-148894	180576030	180576389	28
ZNF395	chr8	-	-13637	28257450	28257779	19
ZNF462	chr9	+	-459848	109165380	109165679	26
ZNF462	chr9	+	-419363	109205820	109206209	26
ZNF462	chr9	+	-486248	109138980	109139279	25
ZNF474	chr5	+	51495	121516530	121516889	21
ZNF488	chr10	+	-28794	48326160	48326429	17
ZNF608	chr5	-	171071	123909600	123909869	48
ZNF608	chr5	-	-581164	124661760	124662179	34
ZNF608	chr5	-	-529324	124609920	124610339	29
ZNF652	chr17	-	-23088	47462370	47462759	24

ZNF669	chr1	-	-155	247267680	247267979	25
ZNF670	chr1	-	-140	247242060	247242359	20
ZNF697	chr1	-	-26284	120216480	120216869	41
ZNF697	chr1	-	-26704	120216930	120217259	30
ZNF704	chr8	-	66792	81720060	81720389	22
ZNF704	chr8	-	108432	81678420	81678749	19
ZNF704	chr8	-	7167	81779700	81779999	18
ZNF716	chr7	+	448017	57957690	57958109	25
ZNF717	chr3	-	-256004	76090110	76090409	31
ZNF717	chr3	-	-266249	76100370	76100639	20
ZNRF1	chr16	+	31000	75063720	75064109	28
ZNRF3	chr22	+	146360	29426070	29426429	28
ZP4	chr1	-	-245394	238299150	238299509	19
ZSWIM7	chr17	-	-31598	15934440	15934769	19
ZWINT	chr10	-	129715	57991110	57991529	22

Table 5.2. GATA3 ChIP-seq in murine 4T1-Gata3 breast cancer cells (O'Geen Dataset-2)

GeneID	chromosome	strand	distance	peak start	peak end	peak value
A1bg	chr15	-	-301881	61054650	61054919	18
A830043J08Rik	chr1	-	41970	80168370	80168729	20
abParts	chr6	+	687010	68192490	68192789	23
abParts	chr6	+	695560	68201040	68201339	20
Acacb	chr5	+	7618	114622950	114623339	17
Adamts16	chr13	-	52199	70928340	70928639	19
Adamtsl4	chr3	-	-8583	95500230	95500499	15
Adcy8	chr15	-	153539	64600140	64600499	30
Afap1	chr5	+	31062	36266850	36267209	18
Aff1	chr5	+	44876	104166120	104166509	28
Agpat3	chr10	-	-19572	77834580	77834939	19
Ahnak	chr19	+	68661	9132300	9132569	13
AK002250	chr10	-	78235	95106750	95107049	19
AK005707	chr3	-	-8485	57976080	57976349	15
AK006031	chr13	-	12814	5844360	5844689	17
AK006310	chr2	-	-98963	109989420	109989779	20
AK006454	chr5	+	-21765	152149380	152149649	15
AK006570	chr10	-	-4716	116170800	116171099	18
AK014626	chr14	-	20249	21066210	21066569	19
AK016040	chr13	+	12099	115409370	115409699	17
AK016730	chr1	+	-4033	87820890	87821159	16
AK017306	chr1	+	-20450	193620660	193620989	24
AK019072	chr3	-	11228	92413740	92414009	16
AK020502	chr13	-	37521	51158430	51158699	10
AK021007	chr7	+	153483	97779000	97779329	18
AK033051	chr4	+	-77547	150122100	150122489	12
AK036632	chr6	-	17358	129166440	129166829	17
AK038386	chr3	+	109997	102176160	102176489	23
AK038843	chr17	-	4752	27374640	27375059	39
AK038843	chr17	-	-20388	27399840	27400139	15
AK039060	chr12	-	-52799	118628130	118628459	13
AK039577	chr12	+	-19700	101508570	101508899	17
AK040027	chr10	+	5612	127941750	127942079	18
AK040143	chr17	-	23422	84515580	84515939	25
AK040166	chr1	-	-44167	60882540	60882869	19
AK040684	chr5	+	8641	75107100	75107429	17
AK042353	chr8	+	7128	8342070	8342339	18
AK043869	chr1	-	2728	51650700	51650969	17
AK044430	chr18	+	7486	44828670	44828939	15
AK048739	chr5	-	229072	99928860	99929159	20
AK050164	chr8	-	24304	129375900	129376169	15
AK052569	chr16	-	-410228	51142950	51143219	13
AK052888	chr6	+	4108	17152080	17152349	18
AK054074	chr11	+	35492	83430870	83431229	14
AK054493	chr17	+	2544	73602690	73603049	16
AK076734	chr3	+	43261	131465790	131466269	31

AK077143	chr4	+	5460	63300750	63301079	19
AK078931	chr11	-	1545	79594680	79595009	17
AK078949	chr11	-	36570	22596270	22596569	18
AK082857	chr9	-	130022	9108930	9109229	19
AK084170	chr9	+	39445	41322270	41322659	30
AK084170	chr9	+	50215	41332980	41333489	24
AK085480	chr16	-	-14635	86820900	86821589	57
AK087392	chr11	+	9019	22526250	22526609	21
AK087392	chr11	+	12394	22529670	22529939	15
AK087725	chr16	-	57886	78237240	78237509	12
AK131742	chr9	-	18501	103843830	103844099	14
AK134047	chr9	+	8987	103334760	103335029	16
AK138371	chr11	-	3974	53296710	53297039	14
AK139772	chr5	-	254082	101645910	101646239	23
AK141227	chr1	-	28294	71046780	71047109	16
AK143195	chr17	-	1010	13819350	13819619	12
AK144610	chr3	-	-9546	65685570	65685929	27
AK144610	chr3	-	-32601	65708670	65708939	17
AK144910	chr4	+	-41	88695990	88696349	23
AK146861	chr10	+	36833	56188470	56188769	17
AK148321	chr4	+	166611	89107560	89107829	18
AK157218	chr2	-	-115619	128370540	128370869	14
AK157916	chr11	+	15241	50065560	50065949	38
AK162833	chr6	+	-15150	117494370	117494759	20
AK163289	chr15	-	130953	61439610	61440029	33
AK163289	chr15	-	298758	61271700	61272329	23
AK163289	chr15	-	292698	61277940	61278209	35
AK163289	chr15	-	287748	61282890	61283159	12
AK163289	chr15	-	286878	61283700	61284089	40
AK163289	chr15	-	226308	61344240	61344689	30
AK188320	chr2	+	-16978	19563360	19563629	12
AK195420	chr7	+	2867	59005560	59005919	26
AK206957	chr17	+	-123	25321260	25321559	22
Aldh3a2	chr11	-	-11820	61092300	61092599	18
AM295304	chr11	-	1293	96886290	96886589	13
Amotl1	chr9	-	67865	14351430	14351729	14
Ank	chr15	+	-47907	27348360	27348689	29
Ankrd1	chr19	-	-23720	36217830	36218279	15
Ankrd11	chr8	-	135058	125430660	125431019	28
Antxr2	chr5	-	-45783	98505630	98505899	12
Aplp2	chr9	-	-51359	31012440	31012739	15
Aqp3	chr4	-	19272	41025810	41026079	13
Arhgap10	chr8	-	89212	79952460	79952729	14
Arhgap21	chr2	-	-21576	20824920	20825189	13
Arid1b	chr17	+	99691	5094630	5094899	15
Arid1b	chr17	+	-368759	4626150	4626479	16
Asah2	chr19	-	31956	32145480	32145869	21
Asb4	chr6	+	46409	5379540	5380049	17
Atp6v1g3	chr1	+	70010	140240070	140240579	40
Atxn7l4	chr12	+	34250	33866760	33867029	15
AY139115	chr14	+	144670	74882730	74882999	14

B230339M05Rik	chr2	+	-454	158234940	158235329	36
Bag5	chr12	-	14254	112936740	112937009	15
Bbs4	chr9	-	-15897	59216970	59217359	16
BC006965	chr11	-	96221	112476240	112476629	33
BC023070	chr15	+	-49731	5396400	5396669	15
BC023070	chr15	+	121629	5567730	5568059	16
BC028660	chr5	-	-12425	138144270	138144539	16
BC038822	chr15	+	126227	97203600	97203929	18
BC048584	chr9	+	51436	41065380	41065979	38
BC049560	chr15	+	2328	81023370	81023699	16
BC055918	chr17	+	29366	36160680	36160949	17
BC057611	chr3	+	-114988	34344150	34344479	22
BC098222	chr16	+	-6758	23168460	23168729	14
Bcar3	chr3	+	-27950	122094660	122094929	16
Bmper	chr9	+	81630	23109000	23109299	14
Bmper	chr9	+	82095	23109480	23109749	12
Bmpr1b	chr3	-	-186997	141781320	141781649	21
Bpag1	chr1	+	132990	34097910	34098209	24
Bpag1	chr1	+	152580	34117500	34117799	13
C130090K23Rik	chr5	+	-32792	73764360	73764689	18
Calcoco2	chr11	-	-2228	95975340	95975669	22
Capns1	chr7	-	-5738	30983490	30983849	21
Capza2	chr6	+	-35283	17551650	17551979	18
Capzb	chr4	+	-12991	138735720	138735989	14
Cc2d1a	chr8	-	2823	86668650	86669009	29
Ccdc12	chr9	+	29213	110588040	110588399	28
Ccdc25	chr14	+	46136	66502110	66502439	18
Ccdc80	chr16	+	17046	45110700	45110969	23
Ccrn4l	chr3	+	-3484	51024720	51025049	20
Cd44	chr2	-	-80117	102821760	102822119	14
Cdc14a	chr3	-	2936	116123850	116124179	23
Cdh1	chr8	+	20607	109147740	109148009	13
Cdh18	chr15	+	184213	22662840	22663139	18
Cdkn2aipnl	chr11	+	-15385	51765630	51765929	20
Chd3	chr11	-	-20778	69180000	69180329	19
Chek1	chr9	-	-1255	36535050	36535379	13
Chek2	chr5	+	-131	111268740	111269069	19
Chic2	chr5	-	122292	75318210	75318509	16
Clcf1	chr19	+	-5617	4208610	4208939	23
Cldn23	chr8	-	-88026	36977490	36977789	20
Clic6	chr16	+	53603	92551860	92552129	15
Col18a1	chr10	-	28976	76547310	76547639	19
Col4a1	chr8	-	54256	11258250	11258699	18
Col4a2	chr8	+	58596	11371440	11371799	32
Coro2a	chr4	-	26514	46588140	46588409	20
Coro2a	chr4	-	17919	46596690	46597049	25
Cpt1a	chr19	+	10867	3334050	3334379	21
Crim1	chr17	+	40783	78640470	78640769	20
Crybb1	chr5	+	8359	112692990	112693409	35
D15Ert621e	chr15	+	127578	58374450	58374779	20
D630040G17Rik	chr8	+	-71661	81480540	81480809	14

D8Ert82e	chr8	+	-89060	37068690	37069019	19
D9Ert8280e	chr9	-	3605	99255210	99255479	14
Dab2	chr15	+	-62259	6187380	6187679	10
Dck1	chr3	+	197535	55243950	55244219	15
Ddc	chr11	-	2270	11720520	11720789	20
Ddef1	chr15	-	86538	64057290	64057709	44
Ddef1	chr15	-	76743	64067130	64067459	23
Ddef2	chr12	+	22798	21140280	21140549	18
Ddx28	chr8	-	4169	108531000	108531329	17
Ddx39	chr8	+	1602	86240520	86240879	25
Depdc6	chr15	+	29252	54973140	54973439	17
Depdc6	chr15	+	74252	55018110	55018469	31
Depdc6	chr15	+	76877	55020720	55021109	18
Disc1	chr8	+	199705	127777650	127777949	16
Dlgap1	chr17	+	740507	71059080	71059349	22
Dmbt1	chr7	+	-19860	138155580	138155879	15
Dnmt1	chr9	-	13498	20750700	20750969	21
Dusp10	chr1	+	-269895	185588310	185588579	17
Dusp10	chr1	+	226845	186085050	186085319	20
EBF	chr11	+	333494	44764980	44765279	18
Elf5	chr2	+	1825	103253520	103253819	12
Emb	chr13	+	-76634	117932610	117932999	22
Eml4	chr17	+	-38307	83711880	83712149	16
Epb4.112	chr10	+	-43072	25036410	25036769	26
Epha4	chr1	-	52729	77458800	77459069	15
Epha4	chr1	-	29719	77481810	77482079	14
Epha4	chr1	-	-331241	77842740	77843069	25
Ets1	chr9	+	70359	32514000	32514329	20
Ets1	chr9	+	72129	32515740	32516129	21
Evl	chr12	+	121160	109913910	109914269	23
Faim2	chr15	-	28989	99329280	99329639	19
Fam50b	chr13	+	-21619	34804950	34805249	14
Fam76b	chr9	+	50598	13691580	13691909	17
Fbxo24	chr5	-	9622	138056490	138056879	30
Fert2	chr17	+	-64765	64148430	64148699	17
Fez2	chr17	-	25624	78791640	78791999	27
Fez2	chr17	-	16339	78800970	78801239	20
Flnb	chr14	+	53039	8703060	8703959	15
Fmnl2	chr2	+	94678	52811430	52811729	21
Fndc3a	chr14	-	263320	72845850	72846209	20
Foxc1	chr13	+	3800	31902180	31902449	17
Frmf6	chr12	+	-169741	71756610	71756909	15
Gadd45g	chr13	+	123641	52065540	52065869	23
Galnt2	chr8	+	-28289	126726840	126727169	21
Gbp4	chr5	-	-15018	105583440	105583739	14
Gm106	chr1	-	-44151	15926790	15927119	20
Gpa33	chr1	+	2169	168062550	168062969	26
Gpr35	chr1	+	-1116	94868400	94868759	13
Gpr97	chr8	+	833	97542330	97542599	12
Grhl3	chr4	-	-25549	135154950	135155219	18
Gtdc1	chr2	-	-35698	44818320	44818589	18

Hand2	chr8	+	63000	59862600	59862959	19
Has2	chr15	-	-11916	56537820	56538119	16
Has2	chr15	-	-27111	56552940	56553389	25
Has2	chr15	-	-71856	56597730	56598089	28
Hdh	chr5	+	5886	35110020	35110529	28
Hs3st1	chr5	-	-71721	40107420	40107719	13
Hs6st1	chr1	+	-91054	36034020	36034379	23
Hsf2bp	chr17	-	3533	32167770	32168039	11
Htra3	chr5	-	16867	36005430	36005699	14
Ikzf3	chr11	-	8135	98331540	98331809	17
Il18rap	chr1	+	27528	40599600	40599869	14
Il6ra	chr3	-	40680	89676180	89676629	21
Itgb1	chr8	+	-18781	131190750	131191049	19
Jag1	chr2	-	-450642	137392560	137392859	18
Jak2	chr19	+	-65893	29260200	29260649	23
Kalrn	chr16	-	99682	34163220	34163519	15
Kif6	chr17	+	236258	49990590	49990919	23
Kirrel3	chr9	+	507026	34803180	34803449	14
Kirrel3	chr9	+	-7446	34286130	34286399	17
Krr1	chr10	+	-107203	111302400	111302819	20
Lactb	chr9	-	-13573	66836670	66837059	21
Lamb1-1	chr12	+	-78669	31871310	31871669	11
Lipl3	chr19	+	19347	34194630	34194929	17
Lmo7	chr14	+	68121	102197100	102197489	28
Lphn2	chr3	-	-127830	148780110	148780439	25
Lrrc6	chr15	-	-7271	66339390	66339959	38
Lrrfp1	chr1	+	79021	92974200	92974469	18
Lrrn2	chr1	+	-14622	134762160	134762459	20
Lsm4	chr8	+	354	73197480	73197809	28
Ly6g6c	chr17	+	1120	35205240	35205539	17
Map3k5	chr10	+	4951	19659090	19659389	18
Mapkapk3	chr9	-	10606	107180970	107181299	24
Mbnl2	chr14	+	32599	120707340	120707639	15
Med13	chr11	-	10818	86160030	86160389	24
Mif1	chr10	+	55535	81872310	81872759	33
mKIAA0650	chr17	-	79439	72345570	72345869	15
Mllt4	chr17	+	18044	14053740	14054009	13
Mtap2	chr1	+	-42098	66179640	66179969	20
Myh9	chr15	-	21326	77651010	77651429	15
Nav1	chr1	-	56223	137528280	137528639	18
Nck2	chr1	+	-17776	43484520	43485119	17
Ndufs4	chr13	-	-32338	115210410	115210769	20
Neu2	chr1	+	-34997	89445480	89445749	16
Nfia	chr4	+	450748	97699170	97699439	14
Nmnat2	chr1	+	130384	154932450	154932779	21
Nmu	chr5	-	-63087	76855620	76856159	13
Nrg1	chr8	-	15204	33104610	33104909	21
Nrip1	chr16	-	-136290	76509390	76509779	43
Nsmce1	chr7	-	19135	132615780	132616049	14
Nt5c1b	chr12	+	-234135	10142520	10142909	13
Nudt12	chr17	-	-63294	59215890	59216189	18

Ocln	chr13	-	35659	101286630	101286959	22
Olfir876	chr9	+	3042	37614270	37614809	34
Osr1	chr12	+	100822	9681990	9682259	10
Otud4	chr8	+	-32710	82130730	82130999	19
Pcdh12	chr18	-	1222	38436240	38436539	22
Phldb2	chr16	-	11172	45942390	45942689	20
Pigt	chr2	+	21655	164344530	164344829	14
Pla2g4a	chr1	-	147300	151660860	151661369	15
Pla2g4a	chr1	-	-51375	151859640	151859939	12
Plcb4	chr2	+	107232	135592380	135592649	14
Pmp22	chr11	+	38638	62983410	62983889	17
Pold4	chr19	+	12547	4244850	4245119	14
Ppp1r9a	chr6	+	179376	5032170	5032529	12
Ppp2r5c	chr12	+	595289	111735750	111736049	15
Ppp3ca	chr3	+	-103899	136229700	136229969	16
Prdm10	chr9	+	-39028	31048980	31049309	18
Prdm8	chr5	+	-72113	98537550	98537999	21
Prkcn	chr17	-	12981	79394730	79394999	18
Ptges	chr2	-	-17117	30775800	30776069	16
Ptpn21	chr12	-	20515	99952110	99952469	19
Ptpru	chr4	-	11219	131382810	131383139	33
Pvrl3	chr16	-	-361932	46858860	46859159	19
Pycr2	chr1	+	12950	182847210	182847569	34
Rad51l1	chr12	+	300231	80698350	80698649	16
Raet1a	chr10	+	198645	22076910	22077209	14
Rasgef1b	chr5	-	-71367	99722640	99723029	20
Rbks	chr5	-	5554	31994280	31994579	16
Rbms1	chr2	-	-232417	60959310	60959699	14
Rgs2	chr1	-	-2630	145853760	145854059	23
Rps6ka2	chr17	+	-5999	7368330	7368599	17
Rpusd4	chr9	+	56724	35132010	35132369	14
Rspo2	chr15	-	11615	42990570	42990929	27
Runx1	chr16	-	43234	92654130	92654549	46
Runx1	chr16	-	-266861	92964240	92964629	21
Runx2	chr17	-	16808	44855610	44855909	11
Scgb1a1	chr19	-	7827	9154380	9154859	24
Scgb1a1	chr19	-	-19038	9181320	9181649	23
Sdc4	chr2	-	9509	164259030	164259329	23
Sec14l2	chr11	-	16658	4001940	4002209	17
Sel1h	chr12	-	-1034477	94121880	94122269	29
Sel1h	chr12	-	-1043027	94130460	94130789	27
Sema3e	chr5	+	-37986	13987140	13987439	14
Sema3e	chr5	+	-20601	14004540	14004809	13
Sema3f	chr9	-	3938	107603850	107604119	14
Sfrp1	chr8	+	-24905	24496950	24497219	13
Sgk	chr10	+	-48295	21553560	21553829	13
Sgms1	chr19	-	23931	32261370	32261729	20
Sh3bp4	chr1	+	-50327	90916560	90916859	18
Shank2	chr7	+	165140	151526400	151526729	21
Slc15a1	chr14	-	8652	121882740	121883339	12
Slc19a2	chr1	+	-5991	166173060	166173329	15

Slc20a2	chr8	+	84088	23671200	23671499	20
Slc24a2	chr4	-	-203850	87080160	87080429	13
Slc25a1	chr16	-	-21323	17949420	17949749	24
Slc29a2	chr19	+	-10786	5013120	5013449	22
Slc45a1	chr4	-	-56186	150082260	150082679	26
Slc7a1	chr5	-	31222	149172840	149173499	19
Slc7a1	chr5	-	1161	149210160	149210429	12
Slitrk1	chr14	-	-83778	109397070	109397399	24
Slitrk1	chr14	-	-84888	109398150	109398539	25
Smarce1	chr11	-	30637	99061530	99061859	22
Smyd3	chr1	-	386435	181061550	181061849	20
Snag1	chr13	-	122432	114286200	114286469	15
Spint1	chr2	+	13567	119076630	119076929	20
St3gal1	chr15	-	46735	66961560	66961859	16
Stc1	chr14	+	24054	69671250	69671549	15
Stk25	chr1	-	-18700	95550840	95551169	18
Stk25	chr1	-	-21100	95553240	95553569	15
Stx8	chr11	+	105860	67885710	67885979	14
Synpr	chr14	+	76576	14193660	14194079	19
Tbl1xr1	chr3	+	-136744	21838680	21838979	14
Tbx20	chr9	-	-36478	24615060	24615389	18
Tcf3	chr10	-	-2921	79896570	79896899	29
TCR-alpha	chr14	+	18175	54847080	54847409	27
Tdrd1	chr19	+	21526	56922030	56922419	17
Tg	chr15	+	136713	66638880	66639209	14
Tgfa	chr6	+	24295	86169540	86169869	24
Thns1	chr2	+	55421	21182670	21182999	21
Tll1	chr8	-	-59824	66744480	66744749	16
Tm9sf3	chr19	-	15520	41322840	41323109	16
Tmcc3	chr10	+	-33944	93943530	93943829	21
Tmem16f	chr15	+	33201	95654310	95654639	23
Tmem26	chr10	+	120409	68306790	68307059	17
Tmem44	chr16	-	58620	30491910	30492179	12
Tmem66	chr8	+	70653	35288160	35288429	16
Tnfsf11	chr14	-	63031	78644670	78644969	16
Tns3	chr11	-	-57353	8425710	8426009	16
Tpbp	chr9	+	-32917	85702920	85703279	22
Tslp	chr18	+	-88137	32886750	32887049	26
Tspan12	chr6	-	46111	21756270	21756539	18
Tspan5	chr3	+	3426	138408450	138408719	16
Ttc8	chr12	+	84416	100243080	100243349	14
Ttll7s	chr3	+	69084	146625900	146626199	16
Txinb	chr10	+	65115	17580990	17581289	17
Wdr51b	chr10	+	-25160	98544450	98544839	26
Wdr60	chr12	-	8224	117493110	117493439	17
Wdr70	chr15	-	100875	7948200	7948469	17
Wdr70	chr15	-	93060	7956000	7956299	23
Wsb1	chr11	-	-12677	79069620	79069889	19
Ywhaz	chr15	-	-82055	36805860	36806249	21
Zfand3	chr17	+	-24912	30116940	30117299	29
Zfp169	chr13	-	-35078	48643740	48644009	20

Zfp202	chr9	+	14254	40014030	40014299	16
Zfp326	chr5	+	-16956	106288530	106288799	17
Zfp414	chr17	+	558	33766470	33766739	14
Zfp791	chr8	-	-9495	87656340	87656639	21
Zfpm1	chr8	+	754	124806660	124806929	21
Zfpm1	chr8	+	2359	124808190	124808609	13
Zfpm1	chr8	+	14569	124820460	124820759	16

Chapter 6: DGCR8 controls branching morphogenesis and cell fate in mammary epithelial cells and promotes breast cancer progression

Source: The following chapter contains unpublished data that is being prepared as a manuscript.

Contributions: For this project, I conceived the idea and performed most of the experiments, with help from Audrey Brenot, Jeffrey Lin, Jennifer Tai and Joanne Dai. Pengfei Lu provided helpful discussions and taught me how to perform mammary reconstitution experiments. Robert Blelloch provided the *Dgcr8* floxed mice. Zena Werb and I wrote the manuscript. Zena Werb supervised the project.

DGCR8 controls branching morphogenesis and cell fate in mammary epithelial cells and promotes breast cancer progression

Jonathan Chou^{1,2}, Audrey Brenot¹, Jeffrey Lin¹, Jennifer Tai¹, Joanne Dai¹ and Zena Werb^{1,2,3}

¹Department of Anatomy, and ²Biomedical Sciences Program, University of California, San Francisco CA 94143-0452

³Corresponding Author:

Zena Werb, Ph.D.

Department of Anatomy, HSW 1323

University of California

San Francisco, CA 94143-0452 USA

Tel.: (415) 476-4622

Fax: (415) 476-4565

zena.werb@ucsf.edu

Keywords:

DGCR8; microRNA; cell differentiation; branching morphogenesis; MMTV-PyMT

ABSTRACT

MicroRNAs (miRNAs) are a class of small non-coding RNAs that play roles in numerous aspects of normal development and cancer. DGCR8 is an RNA-binding protein essential for miRNA processing. Conditional knockout (cKO) studies of *Dgcr8* have been performed to interrogate the global function of miRNAs during the development of embryonic stem cells, neurons, skin, neural crest cells and thymocytes. However, the requirement for miRNA biogenesis in the mammary gland has not been investigated. Here, using a cKO of *Dgcr8*, we studied the role of miRNAs in mammary epithelial cells (MECs). Loss of *Dgcr8* in fully-differentiated MECs impaired branching morphogenesis and resulted in a loss of cell identity. *Dgcr8*-null MECs expressed higher levels of mesenchymal markers, characteristic of a progenitor-like phenotype, and failed to reconstitute the mammary gland. In vivo, MMTV-Cre-mediated deletion of *Dgcr8* resulted in a modest defect in ductal invasion during puberty that recovered over time. In addition, loss of one miRNA, miR-29b, delayed branching morphogenesis. Because global miRNA reduction has been observed in various cancer types, we investigated the effects of global miRNA down-regulation in breast cancer progression and metastasis. Contrary to the effects observed in lung cancer, we found that loss of miRNAs is not well-tolerated in breast cancer cells, resulting in impaired differentiation and growth, and decreased metastasis. Together, these results suggest a critical role for miRNAs in governing essential aspects of normal mammary gland development and breast cancer. Future work will be aimed at uncovering the roles of individual miRNAs in these processes.

INTRODUCTION

Our knowledge and understanding of microRNAs (miRNAs) has greatly increased over the past several years. miRNAs are a class of non-coding RNAs of approximately 21 nucleotides in length, essential for various aspects of animal development (Stefani and Slack 2008). They regulate gene expression post-transcriptionally by guiding the RNA-induced silencing complex (RISC) to the complementary sites in the 3'-untranslated region (UTR) of target mRNAs. Based on bioinformatics predictions, more than one third of mammalian mRNAs are potential targets of microRNAs (Lewis et al. 2005).

Canonical microRNA biogenesis requires the microprocessor, composed of Drosha and DGCR8 (also known as Pasha), to generate precursor-miRNAs (pre-miRNAs). These are exported out of the nucleus in an exportin-5-dependent manner (Yi et al. 2003; Denli et al. 2004), and further processed by an additional RNase enzyme, Dicer, to generate the mature miRNA (Hutvagner et al. 2001). Nucleotides 2 – 7 of the miRNA are known as the “seed” sequence, which specify the set of mRNAs targets for a particular miRNA (Bartel 2009).

The role of *Dgcr8* and miRNA biogenesis has been widely investigated in the development of many tissues and cell types. For example, embryonic stem (ES) cells lacking *Dgcr8* fail to silence pluripotency markers and the self-renewal program, and are thus unable to properly differentiate (Wang et al. 2007). In addition, mice with *Dgcr8*-deficient cardiac neural crest cells display a wide range of cardiac malformations, including persistent truncus arteriosus and ventral septal defects (Chapnik et al. 2012). Moreover, loss of *Dgcr8* in vascular smooth

muscle cells results in extensive liver hemorrhaging, dilated blood vessels, and chaotic vascular architecture (Chen et al. 2012). In the skin, *Dgcr8*-deletion results in multiple morphological abnormalities including breakdown of the epidermal barrier, whereas *Dgcr8*-null thymocytes are skewed towards a Th1 interferon gamma (IFN γ) phenotype (Steiner et al. 2011). These studies underscore the importance of *Dgcr8* and miRNAs in normal development. However, the role of *Dgcr8* in the mammary gland has not yet been investigated.

The mammary gland is a remarkable organ that develops mostly during postnatal life. In mice, an epithelial placode is formed during embryogenesis, and a rudimentary ductal structure in the nipple area is present at birth. During pubertal development, these ducts extend to fill the mammary fat pad in response to hormones. This normal physiologic invasion occurs at the distal tips, called terminal end buds (TEBs), and is completed at about 9 weeks of age when an intricate epithelial network is fully established. Additional ductal side branches are formed during each estrus cycle. During pregnancy, these ductal structures differentiate into lobuloalveolar cells, which produce and secrete milk during the lactation period. After weaning, the mammary gland involutes and resembles the virgin gland morphologically (Khokha and Werb 2010; McNally and Martin 2011).

Understanding the cellular and molecular mechanisms that control ductal invasion and branching morphogenesis has been a major focus of mammary gland biology. Several hormonal and mitogenic signaling pathways that promote or inhibit branching morphogenesis have been studied, including the epidermal growth factor (EGF), fibroblast growth factor (FGF), transforming growth factor

(TGF β), insulin-like growth factor (IGF), Wnt and Hedgehog pathways (Wiseman and Werb 2002; Lu and Werb 2008; Khokha and Werb 2010). These studies further highlight the intricate crosstalk between the epithelial and stromal compartments in regulating mammary development.

Recently, the role of individual miRNAs in mammary development has been investigated (Piao and Ma 2012). A comprehensive miRNA profiling of the mammary gland at different developmental stages (including juvenile, puberty, mature virgin, gestation, lactation, and involution) was recently published, showing that miRNAs are differentially regulated in seven major clusters and contribute to various functions during development (Avril-Sassen et al. 2009). For example, the let-7 family inhibits self-renewing properties, whereas miR-22 and miR-205 promote the expansion of mammary progenitors (Ibarra et al. 2007; Greene et al. 2010). In addition, miRNAs such as miR-126-3p directly target the progesterone receptor, leading to reduced proliferation and expression of β -casein (Cui et al. 2011). Moreover, miRNAs expressed in the stroma, such as the miR-212/132 family, regulate ductal outgrowth by direct targeting of *Mmp9*. Transplantation experiments with wildtype mammary cells into cleared fat pads of miR-212/132 knockout mice had impaired outgrowths, whereas mammary cells isolated from miR-212/132 knockout grew normally in cleared wildtype fat pads (Ucar et al. 2010).

In addition to playing a role in normal development, miRNAs also have positive and negative roles in the pathogenesis of cancer. For example, miR-9 and miR-10b promotes breast cancer invasion by targeting *CDH1* and *HOXD10*,

respectively, (Ma et al. 2007; Ma et al. 2010b), whereas miR-31 inhibits invasion by inhibiting *RHOA* and *ITGA5* (Valastyan et al. 2009). Mutations in miRNA biogenesis genes that promote cancer have also been identified, including those in *XPO5*, *TARBP*, and *DICER1* (Merritt et al. 2008; Melo et al. 2009; Melo et al. 2010). Therefore, understanding the role of miRNAs in the mammary gland will help elucidate molecular mechanisms in breast cancer.

Using a conditional *Dgcr8* knockout mouse model, we have investigated the effects of perturbing global miRNA expression in mammary development and breast cancer progression. Our findings underscore the importance of microRNAs in branching morphogenesis and maintenance of cell identity, and demonstrate that general down-regulation of miRNA biogenesis in breast cancer is not well-tolerated, thereby opening up new opportunities for therapeutic intervention.

RESULTS

DGCR8 controls differentiation and is essential for mammary branching morphogenesis

DGCR8 is part of the microprocessor complex, which is essential for the biogenesis of miRNAs (Landthaler et al. 2004; Wang et al. 2007). To confirm that DGCR8 is present in mammary epithelial cells (MECs), we stained mammary glands from adult virgin female mice. We found that DGCR8 was expressed in both luminal and basal cells along the mammary duct (**Fig. 6.1A**). In addition, we fractionated MECs into luminal and basal fractions using the previously described markers CD24 and CD49f (Stingl et al. 2006). We confirmed successful fractionation of luminal and basal cells by analyzing differentiation markers by quantitative PCR (qPCR), and found that *Dgcr8* was expressed equally in both populations (**Fig. 6.1B–6.1C**). These results indicate that DGCR8 is expressed in both epithelial compartments of the adult mouse mammary gland.

Several miRNAs have been reported to be expressed in the mammary gland, both in the epithelial and the stromal compartments (Avril-Sassen et al. 2009; Ucar et al. 2010; Wright et al. 2010). However, the effect of deleting the entire repertoire of miRNAs in mammary epithelial cells has not been addressed. To investigate the role of DGCR8 and miRNAs in mammary gland development and branching morphogenesis, we used a conditional *Dgcr8* floxed mouse in which excision of exon 3 from the *Dgcr8* mRNA by Cre recombinase generates a frame shift, leading to several premature stop codons (Wang et al. 2007). We crossed the *Dgcr8* floxed mice to a Cre reporter line (lox-STOP-lox-YFP (LSL-

YFP)), in which excision of the STOP elements by Cre recombinase turns on expression of a yellow fluorescent protein (YFP) reporter driven off the Rosa26 promoter. This allowed us to mark cells that have been exposed to Cre. We isolated primary MECs from *Dgcr8 fl/+* and *fl/fl* mice, infected the cells with an adenovirus containing Cre recombinase-GFP (adenoCre-GFP) ex vivo, and then sorted the cells for YFP expression to obtain a pure population of control and *Dgcr8*-null cells (**Fig. 6.2A**). As an alternate control, we infected *fl/fl* MECs with an adeno-GFP virus and sorted them for GFP expression (*fl/fl*; GFP). We confirmed that the *Dgcr8 fl/fl*; YFP cells had decreased *Dgcr8* mRNA and DGCR8 protein levels by quantitative PCR (qPCR) and western blot, respectively (**Fig. 6.2B–6.2C**). In addition, we confirmed that several miRNAs expressed by MECs were decreased following *Dgcr8* deletion (**Fig. 6.2D**).

We embedded the MECs into Matrigel, a laminin-rich extracellular matrix, following re-aggregation overnight. Primary MECs and organoids undergo branching morphogenesis in Matrigel culture that recapitulates many features branching and ductal elongation in vivo (Ewald et al. 2008). Whereas the *Dgcr8 +/+* and *fl/+* MECs formed branched structures by day 7 in culture, *Dgcr8 fl/fl* MECs failed to branch. Instead, *Dgcr8*-null MECs formed solid, compact cellular structures (**Fig. 6.2E, Supplemental Movies 1 and 2**). We further characterized these cells by staining for epithelial markers. Although deletion of *Dgcr8* did not inhibit expression of epithelial markers, *Dgcr8*-null MEC aggregates failed to properly organize keratin-14 (K14)-positive basal cells to surround keratin-8 (K8)-positive luminal cells (**Fig. 6.2F**). Instead, K8 and K14 cells seemed to be

randomly dispersed within the aggregate. We also noted several aggregates with K8+K14+ double positive cells, associated with a de-differentiated, progenitor-like phenotype. Finally, we analyzed *Dgcr8*-null aggregates for differentiation markers by qPCR and found that there was a general increase in basal/mesenchymal marker expression, and a general decrease in luminal marker expression (**Fig. 6.2F**). These data are consistent with a more de-differentiated phenotype in MECs lacking *Dgcr8*. Taken together, our results demonstrate that DGCR8 maintains cell differentiation and identity and is necessary for MECs to undergo branching morphogenesis.

DGCR8 deletion impairs mammary gland reconstitution in vivo

Because *Dgcr8*-null MECs cannot undergo branching in culture, we sought to determine whether *Dgcr8* was required to reconstitute the mammary gland in vivo. We isolated *Dgcr8* *+/+*, *fl/+*, and *fl/fl* MECs from adult virgin mice, infected the cells ex vivo with adeno-Cre, and sorted the cells by FACS, as previously described. We injected these cells into cleared fat pads of the fourth mammary gland, and allowed the cells to grow for six weeks. Because the MECs were marked with the LSL-YFP reporter, we were able to trace our injected cells. Whereas 90% (9/10) of the control MECs (*+/+* and *fl/+*) were able to give rise to a fully reconstituted mammary gland (**Fig. 6.3A–6.3B**), only 30% (3/10) of the *Dgcr8*-null MECs were capable to do so (**Fig. 6.3C–6.3D**). The few structures that did grow from *Dgcr8* *fl/fl* MECs were stunted (**Fig. 6.3D**). These results demonstrate that *Dgcr8* is required for mammary reconstitution in vivo.

MMTV-Cre-mediated deletion of DGCR8 delays ductal invasion during puberty

We have thus far isolated pure populations of recombined cells by FACS, following transduction with adeno-Cre virus. However, to investigate the effects of deleting *Dgcr8* during development in vivo, we crossed the *Dgcr8* mice to the murine mammary tumor virus (MMTV) promoter-Cre line. MMTV-Cre has been previously reported to target cells in the luminal and myoepithelial lineages, with little or no expression in the stroma, apidocytes or fibroblasts. Its expression is hormonally controlled, increases during puberty, and cycles with estrus (Wagner et al. 2001). Interestingly, MMTV-Cre recombination is mosaic, such that not all cells will express Cre. We used flow cytometry to characterize the expression of the LSL-YFP reporter in the luminal and basal compartments using previously described the previously described markers, CD49f and CD24 (Stingl et al. 2006), and found that within the luminal population of adult mice (10 – 14 weeks old), ~80% of cells expressed YFP (data not shown), confirming previous studies showing mosaicism in the MMTV-Cre model (Lu et al. 2008). In addition, within the basal population, ~20% of cells expressed YFP. Thus MMTV-Cre expression is not restricted to luminal cells per se, but also targets a minor population of basal cells.

The mammary gland undergoes rapid proliferation and invasion during pubertal development. In mice, this occurs between weeks 4 – 9. We analyzed whole-mounts of mammary glands isolated from MMTV-Cre; *Dgcr8* *fl/+* and *fl/fl*

mice at various ages (5, 8, and 12 weeks). At five weeks, ductal invasion into the stromal fat pad was delayed in *Dgcr8*-null mice (**Fig. 6.4A**). There were also fewer terminal end buds (TEBs) in *Dgcr8*-null mice (data not shown). However, ductal invasion between the two genotypes was indistinguishable at ten weeks, despite the initial delay (**Fig. 6.4B**).

We next evaluated YFP expression in adult MMTV-Cre; *Dgcr8 fl/fl*; LSL-YFP mice to verify that the MMTV-Cre was being efficiently expressed in the *Dgcr8 fl/fl* glands. Whereas 80% of the luminal cells from MMTV-Cre; *Dgcr8 fl/+*; LSL-YFP control mice were YFP-positive, only ~40% of cells from MMTV-Cre; *Dgcr8 fl/fl*; LSL-YFP were YFP-positive (**Fig. 6.4C**). This decrease of YFP-positive cells in the mutant glands may stem from at least two possible non-mutually exclusive reasons.

One possible explanation for this difference is that because MMTV-Cre is expressed with each hormonal cycle, the YFP-positive luminal cells in the MMTV-Cre; *Dgcr8 fl/fl*; LSL-YFP mice may have transiently expressed Cre, but then subsequently eliminated (e.g. by apoptosis). This would account for the fewer number of YFP-positive cells detected, which represents the equilibrium between the expressing Cre and losing recombined cells. Another possibility is that the cells that persist expressed MMTV-Cre, and were thus marked by YFP, but failed to delete both alleles of *Dgcr8*. This hypothesis would imply that YFP-positive cells from MMTV-Cre; *Dgcr8 fl/fl*; LSL-YFP mice are heterozygous, and thus capable of branching.

To address these possibilities, we isolated YFP-positive cells from MMTV-Cre; *Dgcr8 fl/+*; LSL-YFP and MMTV-Cre; *Dgcr8 fl/fl*; LSL-YFP mice, as well as the YFP-negative cells as a control, and embedded the cells in 3D Matrigel to ask if the YFP-positive cells from the *Dgcr8 fl/fl* mice would branch. We found that whereas YFP-positive cells from *Dgcr8 fl/+* mice were capable of branching, the YFP-positive cells from *Dgcr8 fl/fl* mice were unable to branch (**Fig. 6.4D**). In addition, we confirmed recombination of both loci in the YFP-positive cells by PCR (data not shown). Importantly, the YFP fluorescence in MMTV-Cre; *Dgcr8 fl/fl*; LSL-YFP mice was dimmer than in control mice throughout the entire mammary gland, suggesting that recombination was not localized specifically to the proximal or distal region, and is consistent with our observation of fewer YFP-positive cells by flow cytometry.

We stained glands from 5 week old *Dgcr8 fl/+* and *fl/fl* mice for luminal and basal keratin markers, as well as for E-cadherin, which localizes to the membranes of luminal epithelial cells. We found that whereas *Dgcr8 fl/+* control glands contained E-cadherin/K8-positive luminal cells surrounded by K14-positive basal cells, mutant glands contained some K14-positive basal cells in the luminal layer (**Fig. 6.4E**). In addition, although luminal cells from *Dgcr8 fl/fl* glands expressed E-cadherin, they were more disorganized morphologically, and contained areas of multi-layered epithelium. Moreover, some *Dgcr8 fl/fl* glands were more dilated than their control counterparts. Taken together, these results show that *Dgcr8*-mutant mosaic glands have delayed ductal invasion as well as cell differentiation defects. However, MMTV-Cre-mediated deletion of *Dgcr8* is

tolerated in mosaic mammary glands, so that *Dgcr8*-null cells can contribute to the final epithelial ductal network.

***Dgcr8*-sufficient cells cannot compensate for *Dgcr8*-null cells in a mammary reconstitution assay**

Because MMTV-Cre-mediated deletion of *Dgcr8* is incomplete in vivo, and because *Dgcr8*-null cells seem to be tolerated in the developing mammary gland, we wanted to explore the possibility that *Dgcr8*-null MECs could be rescued by *Dgcr8*-sufficient cells. To investigate this, we isolated *Dgcr8 fl/fl*; LSL-YFP MECs, *Dgcr8 fl/+*; LSL-RFP MECs or *Dgcr8 +/+*; LSL-YFP MECs. This allowed us to label the MECs from different genotypes with different colors (YFP or RFP). After infecting these cells with adeno-Cre virus and sorting, we mixed *Dgcr8 fl/fl*; LSL-YFP and the RFP *Dgcr8 fl/+* control cells at a 1:1 ratio, injected the cells in cleared fat pads, and allowed the cells to grow for 6 weeks. As a control, we mixed *Dgcr8 +/+*; LSL-YFP MECs and *Dgcr8 fl/+*; LSL-RFP MECs.

If the control RFP MECs were able to rescue the *Dgcr8*-null YFP MECs, we would expect dual-colored YFP and RFP ducts. However, we found that *Dgcr8 fl/fl*; LSL-YFP MECs failed to contribute to any part of the mammary gland, which were instead completely reconstituted by *Dgcr8 fl/+*; LSL-RFP MECs (**Fig. 6.4F**). In contrast, mice that were reconstituted with a mixture of *Dgcr8 fl/+*; LSL-RFP and *Dgcr8 +/+*; LSL-YFP MECs showed dual-colored ductal structures, suggesting that both genotypes contributed to final ductal outgrowth (**Fig. 6.4F**). We did not, however, observe intermingling between the RFP and YFP control

MECs, which suggests that the mammary trees regenerated independently. Taken together, our results indicate that *Dgcr8*-sufficient MECs cannot rescue *Dgcr8*-null MECs when mixed together in a mammary reconstitution assay; *Dgcr8*-null cells are most likely out-competed by *Dgcr8*-sufficient cells when forming a de novo mammary ductal network.

Loss of miR-29b delays branching morphogenesis and causes an increase in reconstitution efficiency

We previously showed that miR-29b is downstream of GATA3 (Chapter 2) and that miR-29b promotes differentiation along the luminal epithelial lineage. Expression of miR-29b is enriched in normal luminal cells in the mouse mammary gland and in luminal type breast cancers. In addition, loss of miR-29b in breast cancer cells results in a de-differentiated, more mesenchymal phenotype, and increases metastasis. We sought to determine whether miR-29b plays a role in normal MECs differentiation and branching morphogenesis. We utilized a miRZip-29b lentivirus to stably knockdown miR-29b expression in normal MECs, and selected transduced cells by flow cytometry using the GFP marker on the lentivirus. Interestingly, in the 3D Matrigel assay, loss of miR-29b delayed branching at Day 5 (**Fig. 6.5A**). However, by Day 7, Control MECs and Zip-29b MECs were both branched, although the Zip-29b organoids tended to be slightly smaller than the Controls organoids (**Fig. 6.5B**).

To test what happens in vivo, we sorted transduced MECs, transplanted them into contralateral mammary fat pads, and allowed the cells to grow for 6

weeks. We found that Zip-29b MECs reconstituted the fat pad very efficiently, forming fuller and even more complex epithelial networks than the Control MECs on the contralateral side, as examined by fluorescence whole-mount microscopy (**Fig. 6.5C**) and Carmine staining (**Fig. 6.5D**). These results are consistent with data showing that loss of miR-29b increases expression of many stem cell markers in breast cancer cells, including *Bmi1*, *Klf4*, *Lgr5*, *Lrp6*, *Snail* and *Slug* (**Fig. 6.5E**).

DGCR8 is expressed in multiple breast cancer mouse models but is lost in the basal C3(1)-Tag model

miRNAs have previously been reported to be generally down-regulated in multiple cancer types compared to normal tissue counterparts, and can be used to classify poorly differentiated tumors (Lu et al. 2005). In addition, *DICER1* has been shown to be a haploinsufficient tumor suppressor in a lung cancer model. Analysis of human cancer genome copy number data reveals frequent deletion of *DICER1* (Kumar et al. 2009). Moreover, expression of miRNAs that target Dicer are associated with metastasis and poor outcome (Martello et al. 2010), and a low Droscha and low Dicer levels are associated with poor outcome in ovarian cancer patients (Merritt et al. 2008; Faggad et al. 2010). These results suggest that dysregulation of enzymes involved in miRNA biogenesis contribute to tumorigenesis and metastasis. Indeed, mutations or deletions in other members of the miRNA biogenesis pathway (e.g. exportin-5 and TARBP2) have been identified in many types of cancer (Melo et al. 2009; Melo et al. 2010).

In breast cancer, there have been conflicting reports whether miRNAs are generally down-regulated or up-regulated during tumorigenesis. For example, Dicer was decreased in approximately 50% of human breast cancer cases and was associated with lack of expression of estrogen and progesterone receptors, high grade, high Ki-67, triple-negative and basal-like phenotypes. However, no significant associations between down-regulation of Dicer and outcome were observed (Dedes et al. 2011; Yan et al. 2012). In a separate study, Grelier and colleagues found that Dicer protein expression was significantly associated with hormone receptor status and enriched in the good prognostic luminal A subtype, and that lower Dicer expression was found in cells with a mesenchymal phenotype and in metastatic bone derivatives of a breast cancer cell line (Grelier et al. 2009). These data argue that restoring microRNA biogenesis machinery may reduce aggressiveness and improve patient outcomes.

In contrast, increased miRNA activity has also been reported in breast cancer (Israel et al. 2009), and higher Dicer protein expression has been found in triple-negative breast cancers (Passon et al. 2012). Furthermore, knockdown of Dicer by siRNA leads to G1 arrest and increased sensitivity to cisplatin, a DNA damaging agent, in MCF7 cells, a luminal breast cancer cell line (Bu et al. 2009). These results suggest that decreasing microRNA biogenesis machinery may lead to better drug sensitivity and improve patient outcomes. Whether restoring or deleting microRNA biogenesis components like DGCR8 or DICER likely lies in the balance of pro-tumorigenic and anti-tumorigenic miRNAs expressed by each individual tumor.

To determine whether DGCR8 is lost during tumorigenesis, we stained tumors for DGCR8 in various mouse breast cancer models, including two luminal models (the MMTV-promoter driven polyoma middle T antigen (PyMT) and MMTV-ErbB2/Neu models) and two basal models (MMTV-Wnt1 and the C3(1)-promoter driven SV40 large T antigen (C3(1)-Tag) models). We found that while nuclear expression of DGCR8 was present in all the tumors, expression was very heterogeneous (**Fig. 6.6A–6.6D**). In some tumor regions, only 50% of the tumor cells expressed DGCR8 by the late carcinoma stage. In contrast, nearly all early adenoma tumor cells expressed DGCR8. Therefore, although DGCR8 expression is maintained during tumor progression, its expression seems to be more heterogeneous over time.

We generated cell lines from primary tumors from three of the mouse tumor models (MMTV-PyMT, MMTV-Neu and C3(1)-Tag). Interestingly, we found that *Dgcr8* expression was absent in the basal C3(1)-Tag cell line, compared to the two luminal PyMT and Neu lines (**Fig. 6.6E**). We also conducted gene expression profiling to confirm more basal characteristics of the C3(1)-Tag line (**Fig. 6.6F**). Because basal-type tumors represent the more aggressive subtype, these data suggest that loss of *Dgcr8* may promote tumorigenesis and more aggressive behavior.

We next examined whether *DGCR8* expression was associated with breast cancer subtype in a collection of human cancer cell lines. However, using a previously published dataset (Neve et al. 2006), we found no difference in *DGCR8* expression in luminal cancer cells compared to basal cancer cells (**Fig.**

6.7A). In addition, we found no difference in *DGCR8* expression in primary human breast tumors by subtype. We also did not find any correlation with estrogen receptor (ER) expression (Chin et al. 2006) (**Fig. 6.7B–6.7C**). These results in human cells are contrary to our findings in the primary mouse breast tumors, but may more representative of human tumors. Furthermore, while higher *DGCR8* expression trended to higher overall survival and increased metastasis-free survival, this result reached statistical significance in only one of two microarray cohorts examined (**Fig. 6.7D–6.7E**). Taken together, our data suggest that regulation of *DGCR8* is complex, and that global miRNA repression may play a contributory role, especially in basal type cancers.

***Dgcr8* deletion in MMTV-PyMT breast cancer cells impairs branching morphogenesis and tumor growth**

Because previous studies suggested that general miRNA down-regulation occurs in cancer, we decided to investigate whether decreasing miRNAs globally via *Dgcr8* deletion would promote or inhibit tumor growth and metastasis, both in vitro and in a mouse model of breast cancer. Perturbing *Dgcr8* expression in breast cancer cells has not been previously reported. We chose to cross the *Dgcr8* floxed mice to the MMTV-PyMT mice. This model mimics progressive stages of human luminal breast cancer, and develops lung metastases naturally (Guy et al. 1992; Lin et al. 2003). In addition, the expression of biomarkers in PyMT-induced tumors is parallels those associated with poor outcome in humans, including loss of ER and PR, and the persistent expression of

ErbB2/Neu and cyclinD1. Thus, this model of breast cancer is clinically relevant. Moreover, the *Dgcr8* floxed mice are on C57BL/6 background, and we have back-crossed the MMTV-PyMT mice onto C57BL/6 in our lab, allowing us to assess tumor development in mice of pure genetic background. In order to visualize successful Cre-mediated recombination, these mice also carry the LSL-YFP reporter as previously described.

We generated MMTV-PyMT; *Dgcr8 fl/fl*, *fl/+*, or *+/+* mice with the LSL-YFP reporter and isolated late stage primary carcinoma cells as previously described. We infected the tumor cells with adeno-Cre and used FACS to purify cells expressing the YFP reporter. After re-aggregation overnight, the tumor organoids were embedded into Matrigel. *Dgcr8* deletion in tumor cells also impaired branching compared to *Dgcr8*-sufficient cells, similar our observations in normal MECs (**Fig. 6.8A**). Additionally, we confirmed successful knockdown of *Dgcr8* mRNA in the 3D cultures by qPCR (**Fig. 6.8B**). We then stained the 3D organoids for E-cadherin, K8 and K14. Similar to normal MECs, PyMT organoids lacking *Dgcr8* also failed to properly organize K8 and K14 cells (**Fig. 6.8C**). Only *Dgcr8*-null PyMT organoids contained K8/K14 double-positive cells, characteristic of a progenitor-like phenotype. Furthermore, *Dgcr8*-null PyMT organoids displayed impaired cell adhesion, as evidenced by the rounded morphology of E-Cadherin staining (**Fig. 6.8D**). This was further confirmed by Western blot (data not shown)

We next analyzed the 3D organoids by qPCR and found an increase in basal and mesenchymal marker expression in *Dgcr8*-null PyMT tumor cells,

including *Slug*, *Zeb1*, *Zeb2*, and *Krt14* (**Fig 6.8E**). Interestingly, there was persistent expression of luminal markers, and even increased levels of *Gata3* and *Esr1* (which encodes ER α), suggesting that the *Dgcr8*-null organoids did not lose luminal epithelial characteristics, but gained basal/mesenchymal properties. This is consistent with a general “differentiation crisis.” Consistent with these observations, we found increased pluripotency/stem cell marker expression in *Dgcr8*-null organoids (**Fig. 6.8F**). Taken together, our results demonstrate that *Dgcr8* is required for PyMT tumor organoids to maintain luminal differentiation and undergo branching morphogenesis.

Dgcr8-deletion in PyMT tumors impairs tumor growth and results in fewer metastases

To determine whether *Dgcr8* deletion affected tumor growth, we utilized a single-tumor transplant system into nude mice recipients. This approach was taken because of the tumor heterogeneity of the MMTV-PyMT model (i.e., tumors from each of the different ten mammary glands develop with slightly different kinetics), and because of the mosaicism with the Cre transgenics. (In addition, this approach allowed us to ask what happens when *Dgcr8* is deleted in cells after a tumor has already formed, as opposed to asking how *Dgcr8* deletion affects tumor latency and development. However, we acknowledge the importance of the latter question. Breeding the PyMT; *Dgcr8* floxed mice with a Cre transgenic line will be interesting for future studies.) Therefore, we collected late carcinoma cells from *Dgcr8* +/+, fl/+ or fl/fl mice that were generally at the same stage,

infected them with Adeno-Cre virus, sorted them by FACS, and injected them into the fourth mammary gland of recipient mice.

We found that *Dgcr8*-null tumors were capable of growing, but were always significantly smaller than the control tumors, exhibiting a lag in growth (**Fig. 6.9A**). When we collected the tumors, we were surprised to find that whereas the control tumors were YFP-positive, indicating that the cells had expressed the Adeno-Cre virus and successfully underwent recombination, the *Dgcr8 fl/fl* tumors had only small patches with YFP-positive cells (**Fig. 6.9B**). This suggests that despite using FACS to obtain relatively pure populations of Adeno-Cre-transduced cells, the *Dgcr8*-null cells were out-competed by the *Dgcr8*-sufficient cells so that they contributed to more than 85% of the tumor. Consistent with this, analysis of tumors by qPCR showed that *Dgcr8* levels in PyMT; *Dgcr8* *+/+*, *fl/+* and *fl/fl* were essentially equal (**Fig. 6.9C**). We also examined the tumors histologically and found that in general, they were similar, although the *Dgcr8 fl/fl*; PyMT tumors displayed more areas of lower grade adenoma/carcinoma, suggesting that they were less advanced (**Fig. 6.9D**).

Since the PyMT model spontaneously metastasizes to the lungs, we also examined metastases in a preliminary cohort of mice. At the end of our 12 week experiment, we found that mice transplanted with the *Dgcr8 fl/fl* cells had fewer and smaller lung metastases (**Fig. 6.9E–6.9F**). These data argue that *Dgcr8* is not a haploinsufficient tumor suppressor, as in the case for *Dicer* in KRas-induced lung cancer (Kumar et al. 2009), but that it is a contributing oncogene required for the growth of PI3K/Akt driven cancers. However, further experiments

to formally demonstrate this will be required. Nonetheless, our results suggest a critical role for miRNAs in governing essential aspects of normal mammary gland development and breast cancer.

DISCUSSION

We have used a cKO of *Dgcr8* to study the global role of miRNAs in normal MECs and in primary breast tumor cells. Loss of *Dgcr8* in fully-differentiated MECs impaired branching morphogenesis and resulted in a cell identity crisis and loss of differentiation. *Dgcr8*-null MECs expressed higher levels of mesenchymal markers, characteristic of a more progenitor-like phenotype, and were unable to reconstitute a cleared mammary fat pad. By using the MMTV-Cre line to specifically delete *Dgcr8* in the mouse mammary gland, we found that *Dgcr8* was required for pubertal ductal invasion, but that the mosaicism of the MMTV-Cre allowed non-recombined cells to dominate in the final ductal structure. Our work suggests that this may be due to defects in cell proliferation in *Dgcr8*-null cells, making them unable to contribute to the epithelial network.

Surprisingly, in a preliminary cohort of mice, we found that *Dgcr8*-null females were able to nurse their pups normally (data not shown), suggesting that loss of *Dgcr8* does not affect terminal differentiation into milk-producing alveolar cells. This is not expected, given the important role of miRNAs in almost every aspect of differentiation. We used both MMTV-Cre model as well as the inducible WAP-rtTA-Cre model to assess the consequences on alveolar differentiation. We found that regardless of the Cre line used, the females were able to nurse their pups, despite successful expression of the LSL-YFP reporter marking Cre recombinase activity. Further molecular and structural characterization of the glands will be required to determine whether recombination was successful in

enough mammary cells to exhibit a phenotype, and if so, whether total loss of miRNAs truly has no effect.

Furthermore, because global miRNA reduction has been observed in various cancer types (Lu et al. 2005), we investigated the effects of global miRNA down-regulation in breast cancer progression and metastasis. Contrary to the tumor-suppressive effects of miRNA biogenesis components observed in other cancer types (Merritt et al. 2008; Kumar et al. 2009; Martello et al. 2010), we found that loss of *Dgcr8*, and hence all miRNAs, is not well-tolerated in the MMTV-PyMT breast cancer model, even in late carcinoma cells. PyMT cells isolated from late carcinomas but lacking *Dgcr8* had defects in differentiation and growth, and expressed higher levels of stem cell and EMT markers, such as *Bmi1*, *Nanog*, *Zeb1*, *Zeb2*, *Slug*, and *Vimentin*. Using a 3D Matrigel culture model, we found that *Dgcr8* was required for branching morphogenesis in PyMT organoids. These results suggest that despite expressing an oncogene that drives proliferation (in this case, PyMT), loss of *Dgcr8*, even in late stage tumor cells, hits an Achilles heel and arrests tumor growth. It further suggests that activation of growth pathways induced by PyMT (e.g., the PI3K/Akt pathway (Meili et al. 1998)) are either dependent on sustained miRNA activity to signal and promote growth, or at the very least, are non-dominant to complete miRNA loss.

Interestingly, when we transplanted late carcinoma PyMT cells without *Dgcr8* into the mammary fat pad, we found that the *Dgcr8*-null cells were out-competed by the *Dgcr8*-sufficient cells. Although we used FACS to obtain

relatively pure populations of recombined cells, the few cells that were not recombined grew out to contribute more than 75% of the tumor. Consistent with this observation, there was a lag in tumor growth in the *Dgcr8 fl/fl* tumors compared to the controls. Since the PyMT spontaneously metastasizes to the lungs, we also examined metastases in a preliminary cohort of mice. We found that mice transplanted with the *Dgcr8 fl/fl* cells had fewer metastases, even after accounting for tumor size. These data argue that *Dgcr8* is not a haploinsufficient tumor suppressor, as in the case for *Dicer* in KRas-induced lung cancer (Kumar et al. 2009), but that it is a contributing oncogene required for the growth of PI3K/Akt driven cancers. However, further experiments to formally demonstrate this will be required. Nonetheless, our results suggest a critical role for miRNAs in governing essential aspects of normal mammary gland development and breast cancer.

Finally, future work aimed at uncovering the roles of individual miRNAs in branching morphogenesis and tumor formation in normal MECs and PyMT tumor cells will be important. Because *Dgcr8* deletion results in a total deletion of miRNAs, the background is essentially to zero, so that the effects of individual miRNAs can be studied. This is important because many miRNAs are members of larger families (e.g., the let-7 family, the miR-29 family, the miR-200 family). Thus, deleting individual miRNAs can be problematic because of redundancy and possible compensation by other family members. Indeed, many miRNA knockout mice are viable and show relatively mild phenotypes. For example, targeted deletion of miR-208, a cardiac-specific miRNA, does not appear to have

any effect on normal heart function. It is not until mice lacking miR-208 are stressed that a role for miR-208 in cardiac hypertrophy and remodeling is observed (van Rooij et al. 2007). In addition, knockouts of miR-1-2, another miRNA specifically enriched in muscle, had no obvious skeletal muscle phenotype and only minor defects in heart morphogenesis and electrical conduction (Zhao et al. 2007). This may be attributed in part to a compensatory effect by miR-1-1, another miR-1 family member that is expressed at a different genetic locus. These subtle phenotypes observed in miRNA knockout studies suggest that miRNAs share functional redundancy or have subtle effects on cellular processes, acting more as rheostats to fine-tune protein output (Baek et al. 2008). Therefore, investigating the role of individual miRNAs in the absence of other family members provides a much cleaner interpretation. Although this strategy has its own caveats, it has been successfully used to uncover novel roles for miRNAs in embryonic stem cells, thymocytes and cancer cells (Wang et al. 2008c; Steiner et al. 2011; Peric et al. 2012). We believe that performing a put-back screen in primary *Dgcr8*-null MECs subjected to the 3D Matrigel branching assay will reveal novel miRNAs important for branching morphogenesis, mammary development and breast cancer progression.

ACKNOWLEDGEMENTS

We thank members of the Werb laboratory for discussions and thank Elena Atamaniuc, Ying Yu, and Helen Capili for technical assistance. We thank Laurie Littlepage with help with the expression array data sets. We thank Robert Blelloch for the *Dgcr8* floxed mice, Tara Rambaldo and Michael Kissner for flow cytometer assistance, and Charina Choi for discussions and encouragement. This research was supported by funds from the National Cancer Institute (R01 CA129523 to Z.W.), a Department of Defense Predoctoral Fellowship (W81XWH-10-1-0168 to J.C.) and the UCSF Medical Scientist Training Program (J.C.).

EXPERIMENTAL PROCEDURES

Mammary Epithelial Cell Isolation and Adenovirus Transduction

Mammary glands from adult virgin females or tumors from MMTV-PyMT mice were digested in a collagenase/trypsin solution. Organoids were collected by brief centrifugation and digested with trypsin to dissociate them into single cells. Cells were plated on ultralow attachment plates (Corning) and transduced with adenoCre-GFP virus (Vector Biolabs). Cells were incubated for 2 – 3 days in media containing 10% FBS and insulin, and then processed for sorting or embedded into Matrigel.

Flow Cytometry

To sort primary mouse mammary epithelial cells, cells were trypsinized for 20 minutes in a 1:1 mixture of trypsin:D-PBS. Cells were then collected in serum-containing media and filtered through a 70 μ m filter. To fractionate luminal and basal cells, cells were stained for CD49f and CD24, as well as for CD31, CD45 and Ter119 (all from eBioscience). Prior to sorting, cells were filtered again through a 40 μ m filter. Cell sorting was performed on a FACS Aria II (Becton Dickinson), and analyzed using FlowJo software or FACSDiva software (BD Biosciences).

3D Cell culture

To embed MECs in Matrigel, MECs were first aggregated overnight on ultralow attachment plates (Corning) and then embedded into growth factor-reduced

Matrigel (BD Biosciences) the next day. Cells were grown for the indicated number of days in serum-free media, supplemented with insulin-transferrin-selenium (Invitrogen) and 2.5 nM FGF2 (Sigma).

Animal Studies

All animal experiments were performed at UCSF, and reviewed and approved by the UCSF IACUC. Mice were housed under pathogen-free conditions in the UCSF barrier facility. Nude mice were purchased from Simonsen Laboratories. *Dgcr8* floxed mice have been previously described (Wang et al. 2007), and LSL-YFP and LSL-RFP reporter mice have been previously described. For mammary reconstitution assays, 1×10^4 freshly sorted mammary epithelial cells (MECs) were injected into a cleared mammary fat pad of nude mice in 1:1 DMEM:Matrigel (BD Biosciences). Successful clearing was confirmed by Carmine staining. Cells were allowed to grow in vivo for at least 5 weeks.

Dgcr8 floxed were also crossed onto MMTV-PyMT/B6 mice. Tumors were harvested from donor mice when tumors measured at least 15 mm in one dimension. For experimental metastasis experiments, age-matched female nude mice were injected i.v. (via tail vein) with 1×10^5 freshly sorted cells in PBS. For primary tumors and spontaneous metastasis assays, age-matched female nude mice were injected with the 1×10^5 freshly sorted cells in 1:1 DMEM:Matrigel (BD Biosciences) into the fourth mammary fat pad without clearing. Tumor measurements were made using a caliper at least once per week, and volumes were calculated using the following formula: $V = 0.52 \times (\text{length})^2 \times \text{width}$.

Lentiviral and Retroviral Production

The miRZip29 knockdown vector (System Biosciences) was used according to manufacturer's instructions. Viral production was carried out using calcium phosphate-mediated transfection of HEK 293T, and was concentrated by ultracentrifugation. Primary MECs were transduced with lenti-virus and then selected by FACS.

Quantitative real-time PCR (qPCR)

cDNA was synthesized using the Superscript III RT First Strand Kit (Invitrogen). Real-time qPCR was performed using FastStart Universal SYBR Green master mix (Roche) in an Eppendorf Mastercycler realplex machine. Ct values were normalized to actin and GAPDH, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. For quantification of miRNA expression, Taqman probes were used according to manufacturer's protocol (Applied Biosystems). Ct values were normalized to snoRNA202 and snoRNA234. Primer sequences were found using the Harvard Primer Bank.

Immunostaining

Tissues were fixed in 4% PFA overnight, paraffin processed or embedded into OCT for frozen sections, and sectioned. Heat-mediated antigen retrieval was performed using sodium citrate (Invitrogen). The TSA Amplification Kit (Perkin

Elmer) was used to according to manufacturer instructions. Primary antibodies were incubated overnight, and secondary antibodies were incubated for 1 hour. The following antibodies were used: DGCR8 (Proteintech), Keratin-8 (TROMA-1, University of Iowa, Developmental Studies Hybridoma Bank), Keratin-14 (Covance), E-cadherin (BD Biosciences), phospho-histone H3 and cleaved caspase-3 (Cell Signaling), GFP and GATA3 (Abcam), biotinylated anti-rat (Jackson), and biotinylated anti-rabbit (Dako). Fluorescent secondary antibodies were from Molecular Probes (Invitrogen). Images were taken using a spinning disc confocal microscope or a C1si Nikon upright confocal microscope (UCSF Biological Imaging Development Center) or a spinning disc confocal microscope. Image analysis was performed using Nikon or ImageJ software.

Western blotting

Cells were lysed in RIPA buffer plus protease inhibitors (Roche) or directly in Lamelli Buffer with DTT. Protein concentration was measured using the BCA Protein Assay Kit (Thermo Scientific). Lysates were subjected to SDS-PAGE, transferred to PVDF membranes, blocked in 5% BSA, incubated with primary antibody overnight and visualized using ECL Detection Reagents (Pierce). Exposures were acquired using a LAS-4000 Imager (Fuji). Primary antibodies to actin (Santa Cruz), DGCR8, (Proteintech), Ecadherin (BD Biosciences), phospho-histone H3 (Cell Signaling) and pan-cytokeratin (Sigma) were used at manufacturer's recommended dilutions.

Statistical analysis

Statistical analysis was performed using Prism 4 software (GraphPad Software, Inc.). All data are presented as mean \pm SEM, unless otherwise stated. When two groups were compared, the two-tailed Student's t-test was used, unless otherwise stated. When three or more groups were compared, the one-way analysis of variance (ANOVA) test was used, followed by Tukey's test to determine significance between groups. We considered $p < 0.05$ as significant.

FIGURE LEGENDS

Fig. 6.1: DGCR8 is expressed in both luminal and basal cells in the mammary gland.

(A) Immunohistochemistry for DGCR8 in the virgin, adult mouse mammary gland shows expression in both the luminal and basal cells (scale bar = 100 μ m).

(B) Luminal and basal cells from the virgin, adult mouse mammary gland were sorted by FACS using CD49f and CD24. Successful fractionation was confirmed by analyzing for luminal (*Gata3*, *Krt8*, *Krt18*) and basal (*Krt14*, *p63*) markers by quantitative PCR (qPCR).

(C) Luminal and basal fractions were analyzed for *Dgcr8* mRNA by qPCR.

Fig. 6.2: *Dgcr8*-null mammary epithelial cells fail to maintain a differentiated state and cannot undergo branching morphogenesis.

(A) Mammary epithelial cells (MECs) from *Dgcr8* *+/+* or *fl/+* (Control) and *Dgcr8* *fl/fl* mice were infected with Adeno-Cre on low adhesion plates and sorted by FACS. A representative FACS plot shows that approximately 75% of the MECs were successfully transduced.

(B–C) The sorted MECs were grown on low-adhesion plates after sorting for an additional 1 – 2 days. These MECs were then collected and analyzed for *Dgcr8* mRNA levels by qPCR (B) and DGCR8 protein levels by western blot (C).

(D) The sorted MECs were analyzed for the expression of various microRNAs by Taqman qPCR after *Dgcr8* deletion to confirm impaired microRNA biogenesis.

(E) Following overnight re-aggregation on low-adhesion plates, the sorted MECs

were embedded into three-dimensional (3D) Matrigel and supplemented with serum-free media containing FGF2 to stimulate branching morphogenesis. These organoids were allowed to grow for 7–10 days. Representative phase-contrast images are shown (scale bar = 200 μm (top) and = 100 μm (bottom)).

(F) After 7–10 days in 3D Matrigel culture, the organoids were embedded into OCT, sectioned and stained for Keratin-8 (K8, red) and Keratin-14 (K14, green). Representative confocal microscope images are shown. The K8/K14 double-positive cells (arrows) were only observed in the *Dgcr8*-null organoids (scale bar = 20 μm).

(G) After 7–10 days in 3D Matrigel culture, the organoids were collected directly in Qiazol and the RNA was extracted. Expression of epithelial and mesenchymal genes was analyzed by qPCR.

Fig. 6.3: *Dgcr8*-null mammary epithelial cells fail to fully reconstitute the mammary gland.

(A–D) MECs isolated from *Dgcr8* *+/+*, *fl/+* and *fl/fl*; *LSL-YFP* mice transduced with Adeno-Cre were injected into cleared mammary fat pads and allowed to grow for 6 weeks. Representative fluorescence whole-mount microscopy images of mammary outgrowths are shown (scale bar = 1 mm).

Fig. 6.4: MMTV-Cre-mediated deletion of *Dgcr8* results in a subtle delay in ductal outgrowth during puberty.

(A–B) Whole-mount carmine-stained images from representative 5 week old (A)

and 10 week old (B) MMTV-Cre; *Dgcr8* fl/+ or fl/fl virgin mice. Arrow denotes the length of ductal invasion past the lymph node (scale bar = 1mm).

(C) MECs from MMTV-Cre; *Dgcr8* fl/+ or fl/fl; LSL-YFP mice were isolated and stained with CD49f, CD24 and lineage markers (CD45, CD31, Ter119). The average percentage of YFP-positive cells in the luminal (CD49f^{low}CD24^{hi}) population is depicted.

(D) YFP-positive and YFP-negative MECs from MMTV-Cre; *Dgcr8* fl/+ and fl/fl; LSL-YFP mice were isolated by FACS, re-aggregated overnight on low-adhesion plates, embedded into 3D Matrigel and grown in serum-free media supplemented with FGF2. Representative phase-contrast images of the organoids grown in culture for 7–10 days (scale bar = 100 μ m).

(E) Representative confocal microscopy images of mammary ducts from 5 week old MMTV-Cre; *Dgcr8* fl/+ and fl/fl; LSL-YFP mice stained with K8 (green), K14 (red), and E-cadherin (blue) (scale bar = 25 μ m). Arrows indicate K14-positive basal within the luminal layer of the epithelium.

(F) MECs from *Dgcr8* +/+, fl/+ and fl/fl mice expressing either the LSL-YFP or LSL-RFP reporters were isolated. MECs were mixed 1:1 as indicated, co-injected into cleared mammary fat pads and allowed to grow for 6 weeks. Representative fluorescence whole-mount images of mammary outgrowths are shown (scale bar = 1mm).

Fig. 6.5: Knockdown of miR-29b delays branching morphogenesis in vitro and but has only subtle effects on mammary reconstitution in vivo.

(A–B) MECs from 7 week old virgin female mice were isolated and transduced with miRZip-Control or miRZip-29b lentivirus and purified by FACS. MECs were re-aggregated overnight on low-adhesion plates, embedded into 3D Matrigel and grown in serum-free media supplemented with FGF2. Representative phase-contrast images of the organoids grown in culture for 3 days (A) and 7 days (B) (scale bar = 100 μ m).

(C–D) miRZip-Control or miRZip-29b-transduced MECs were purified by FACS, injected into cleared mammary fat pads on contralateral sides and allowed to grow for 6 weeks. Representative whole-mount fluorescence (C) and carmine-stained (D) images of mammary outgrowths are shown (scale bar = 1mm).

(E) RNA was extracted from miRZip-Control or miRZip-29b-transduced MECs grown in 3D Matrigel by direct lysis. Expression of various stem cell and mesenchymal markers were analyzed by qPCR (n=3 samples per group).

Fig. 6.6: Expression of DGCR8 is heterogeneous in various luminal and basal mouse breast cancers and low in poor-prognostic basal tumors.

(A–D) Primary tumors from two luminal models (MMTV-PyMT (A) and MMTV-Neu (B)) and two basal models (MMTV-Wnt (C) and C3(1)-Tag (D)) were stained for DGCR8 protein. Representative images are shown.

(E) Relative expression of *Dgcr8* mRNA in cell lines generated from MMTV-PyMT, MMTV-Neu and C3(1)-Tag tumors, as measured by qPCR.

(F) Relative expression of luminal (*Krt8*) and basal (*Krt14*, *p63*) markers in the cell lines generated from MMTV-PyMT, MMTV-Neu and C3(1)-Tag tumors to

validate the breast cancer subtype, as measured by qPCR (n=3 samples per group).

Fig. 6.7: *DGCR8* expression in human breast cancer is generally independent of subtype and is not robustly associated with prognosis.

(A) Relative *DGCR8* expression in Basal A, Basal B and Luminal breast cancer cell lines. Data adapted from Neve et al. 2006.

(B) Relative *DGCR8* expression in human Basal, ErbB2, Luminal A, Luminal B and Normal-like breast cancers. Data adapted from Chin et al. 2006.

(C) Relative *DGCR8* expression in human estrogen-receptor (ER)-positive and negative breast cancers. Data adapted from Chin et al. 2006.

(D–E) Kaplan-Meier survival plots of human breast cancer patients based on *DGCR8* expression, segregated into 3 groups (high/medium/low, (D)) or 2 groups (high/low, (E)). Data adapted from Chin et al. 2006.

Fig. 6.8: *Dgcr8*-null MMTV-PyMT cells fail to undergo branching morphogenesis and cannot maintain a differentiated state.

(A) PyMT mammary epithelial cells (PyMT-MECs) from *Dgcr8* *+/+* or *fl/+* (Control) and *Dgcr8* *fl/fl* mice were infected with Adeno-Cre, purified by FACS and embedded into 3D Matrigel. Organoids were grown in serum-free media containing FGF2 for 7 days. Representative phase-contrast images are shown (scale bar = 100 μ m).

(B) RNA was extracted from PyMT-MECs grown in 3D Matrigel and analyzed for

Dgcr8 mRNA levels by qPCR (n=3 samples per group).

(C–D) *Dgcr8 fl/+* and *fl/fl*; PyMT-MECs transduced with Adeno-Cre and grown in 3D Matrigel were stained for K8 (red) and K14 (green) (C) or E-cadherin (D). Representative confocal microscopy images are shown (scale bar = 25 μ m).

(E–F) Relative expression of EMT markers (E) and stem cell markers (F) of PyMT-Control and PyMT-*Dgcr8 fl/fl* MECs transduced with Adeno-Cre and grown in 3D Matrigel for 7 days (n=3 samples per group).

Fig. 6.9: Deletion of *Dgcr8* in PyMT tumor cells is not tolerated and decreases lung metastasis.

(A) Orthotopic primary tumor growth in mice injected with PyMT; *Dgcr8 +/+*, *fl/+* and *fl/fl*; LSL-YFP FACS-purified cells. There is a 2–3 week lag in growth of the PyMT; *Dgcr8 fl/fl* tumors compared to the Control PyMT tumors.

(B) Representative whole-mount fluorescence microscopy images of primary tumors, which shows that the bulk of the PyMT; *Dgcr8 fl/fl*; LSL-YFP tumors are not YFP-positive (scale bar = 1mm).

(C) Relative *Dgcr8* mRNA expression of tumor samples from Control and *Dgcr8 fl/fl* PyMT tumors (n=3 samples per group).

(D) Representative H&E images of primary tumors from Control and *Dgcr8 fl/fl* PyMT tumors.

(E) The number of lung metastases from mice bearing PyMT; *Dgcr8 +/+*, *fl/+* and *fl/fl*; LSL-YFP tumors were plotted.

(F) Representative H&E images of spontaneous lung metastases arising in mice

injected with PyMT-*Dgcr8* *+/+* (Control) and PyMT-*Dgcr8* *fl/fl* cells to form primary tumors.

FIGURES

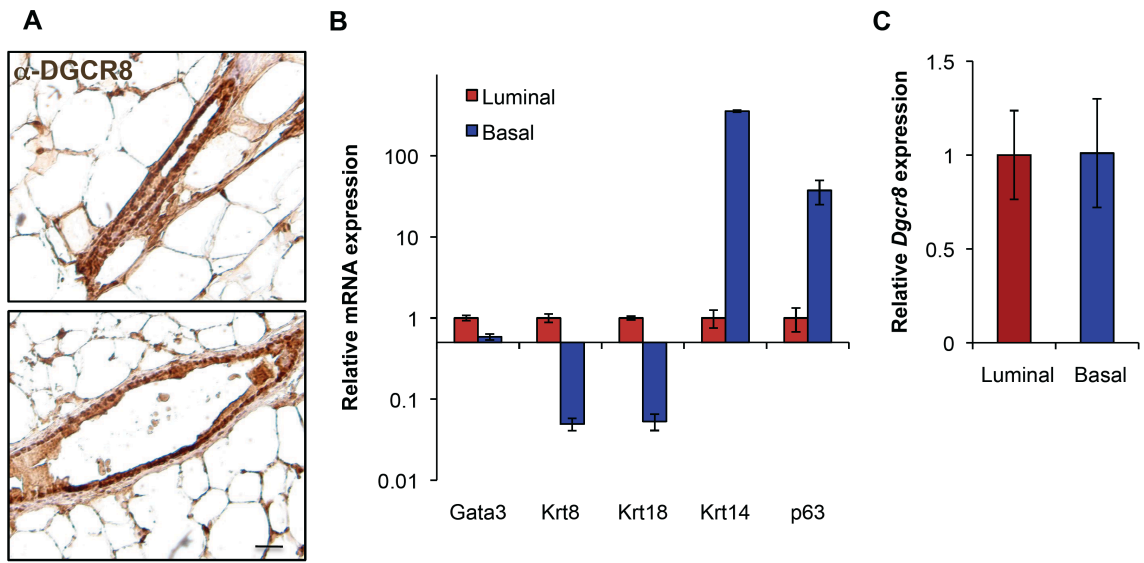


Figure 6.1

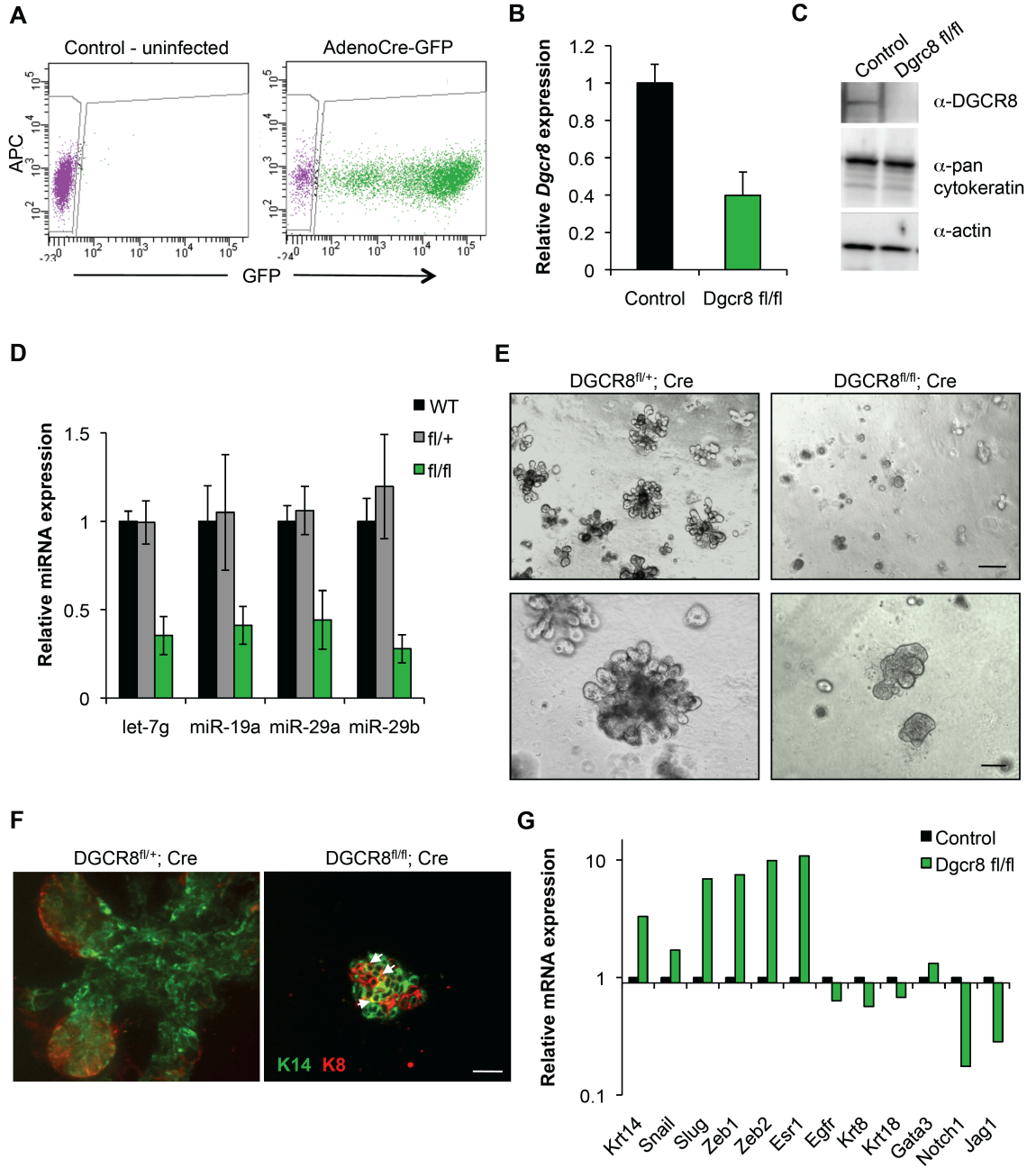


Figure 6.2

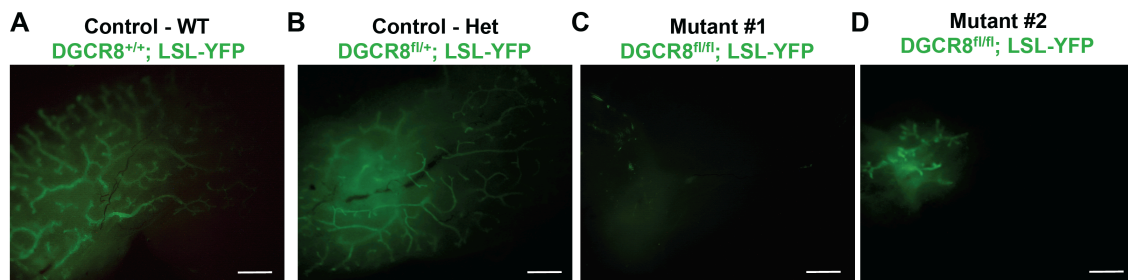


Figure 6.3

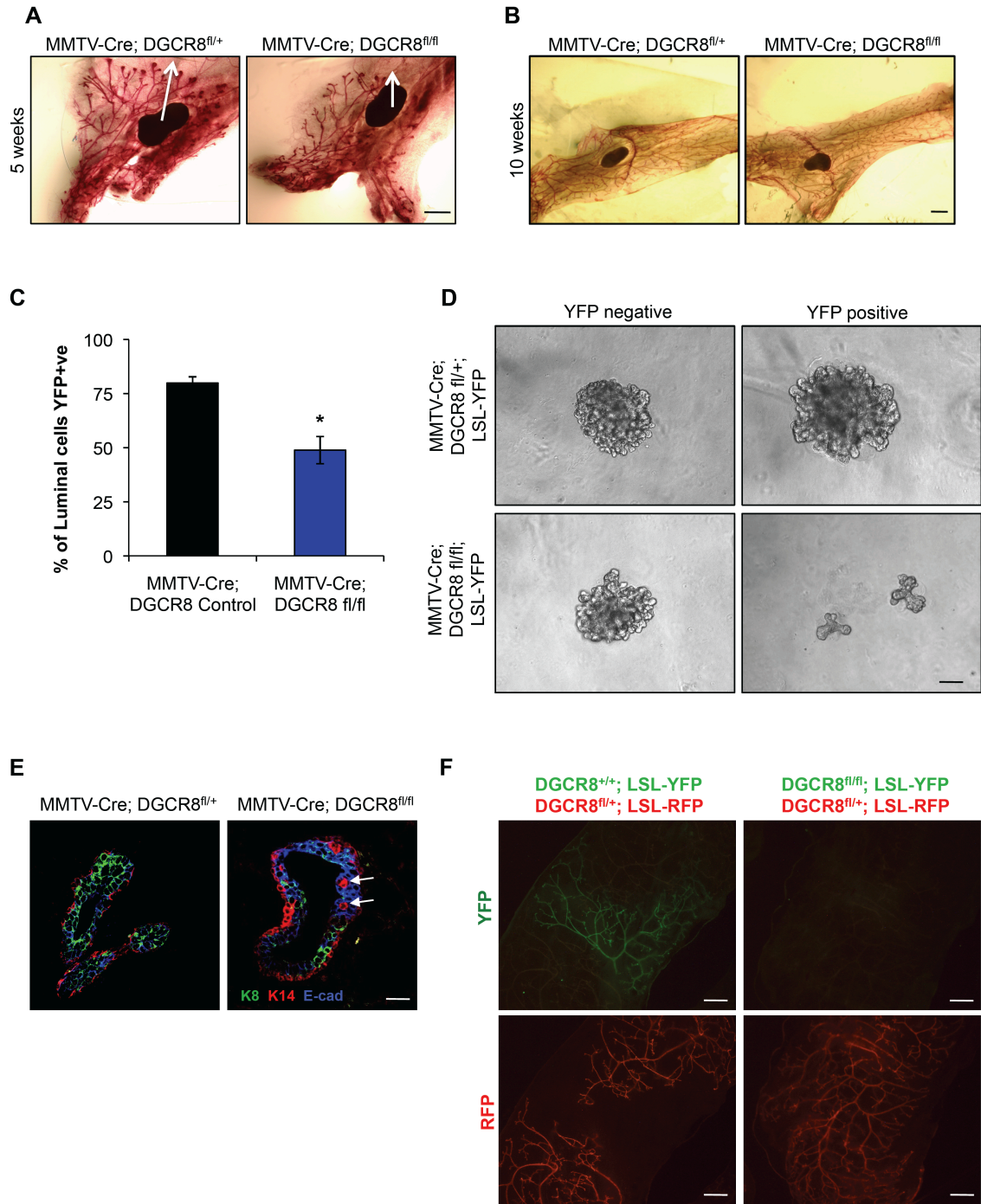


Figure 6.4

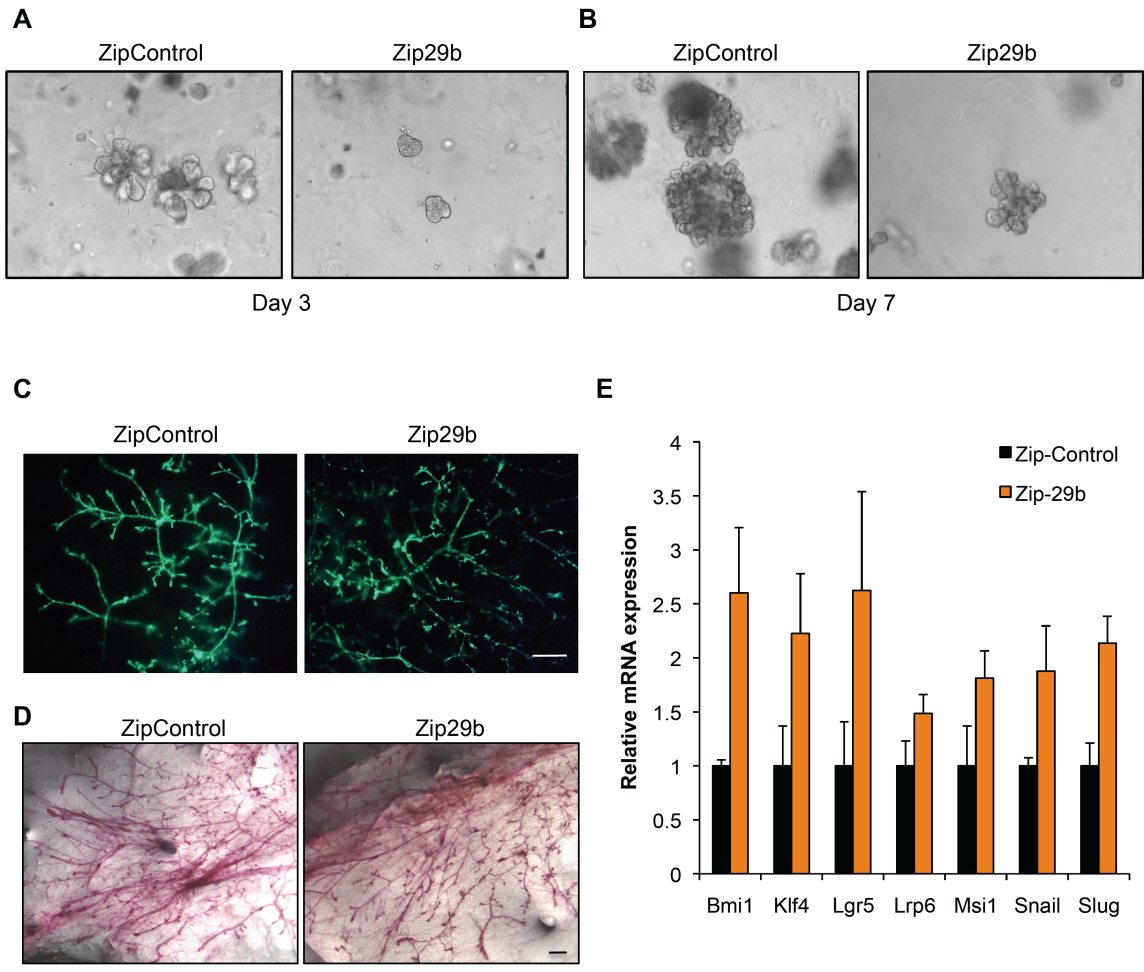


Figure 6.5

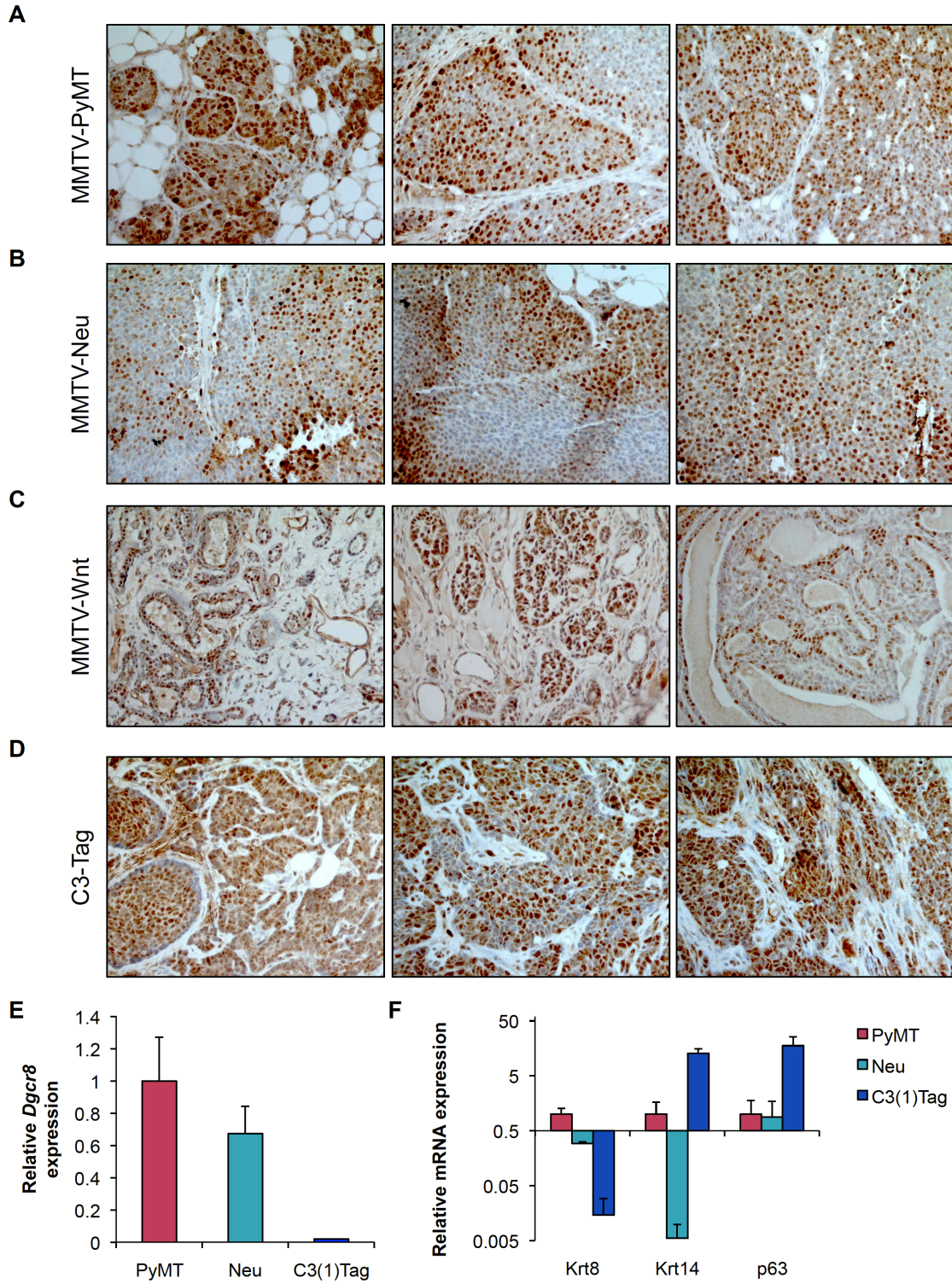


Figure 6.6

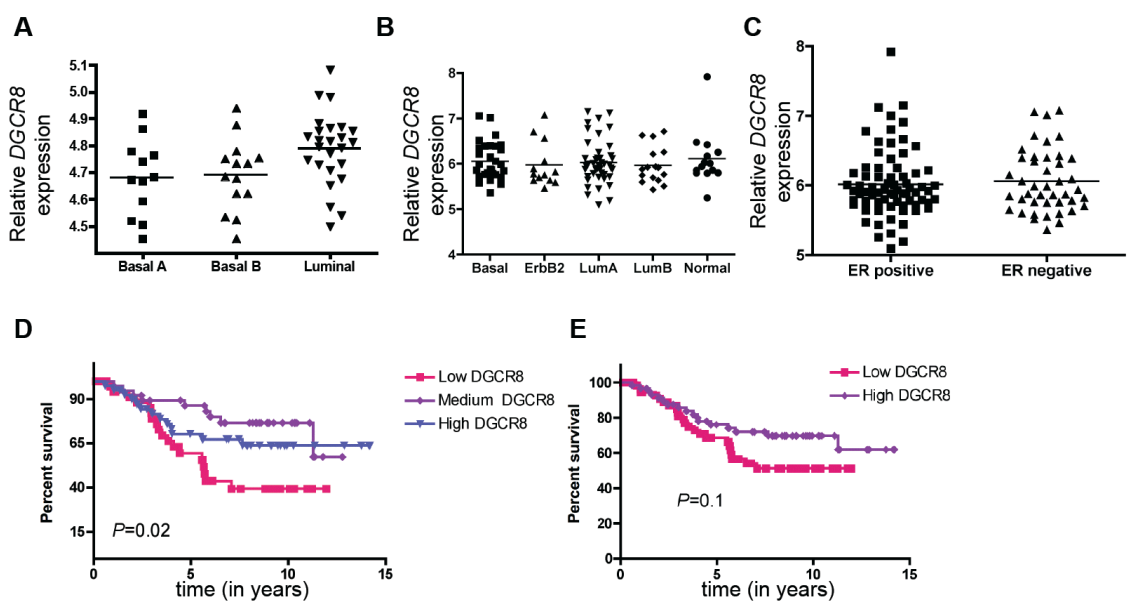


Figure 6.7

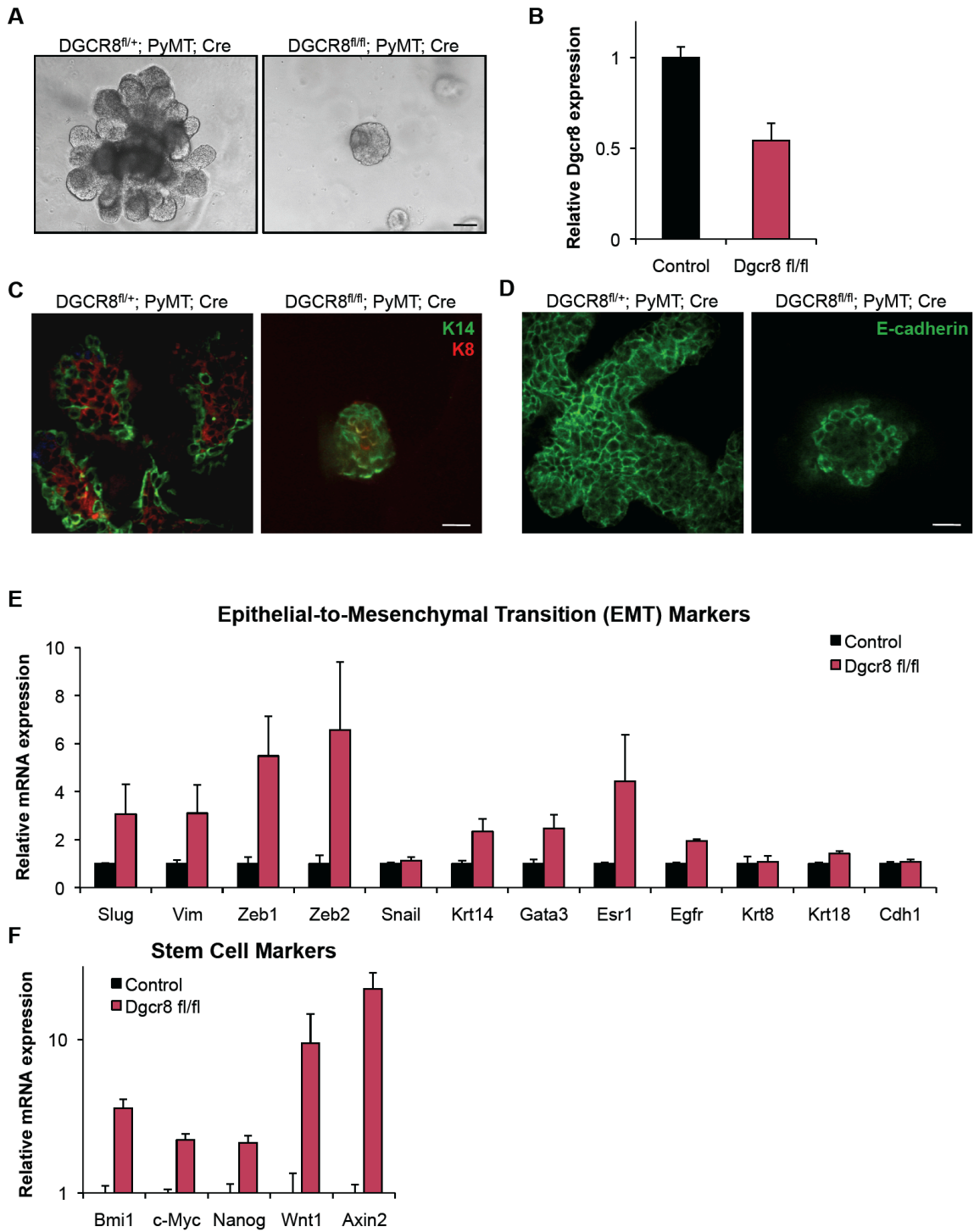


Figure 6.8

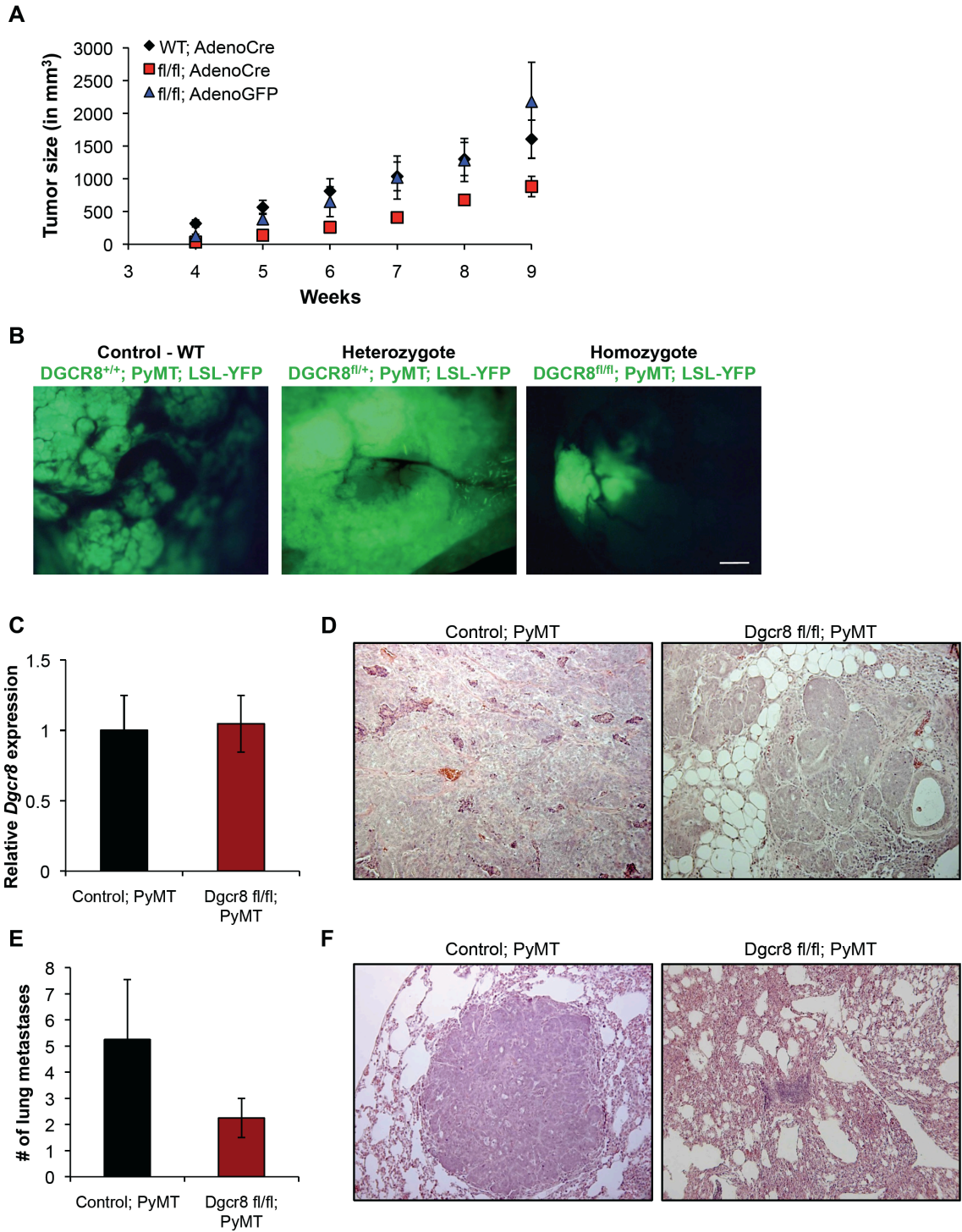


Figure 6.9

Chapter 7: Tumor-derived Systemic Factors and the Early Metastatic Niche

Source: The following chapter contains unpublished data that is being prepared as a manuscript.

Contributions: Vicki Plaks, Zena Werb and I conceived the idea for this project. Vicki Plaks and I performed most of the experiments together. Amy-Jo Casbon assisted with flow cytometry analysis. In addition, I received help from several outstanding undergraduate students: Joanne Dai, Jennifer Tai, Eline van Kappel, Kelly Kersten, Jeffrey Lin and Karren Wang. Zena Werb supervised the project.

Tumor-derived Systemic Factors and the Early Metastatic Niche

Jonathan Chou^{1,2,3}, Vicki Plaks^{1,3}, Amy-Jo Casbon¹ and Zena Werb^{1,2,4}

¹Department of Anatomy, and ²Biomedical Sciences Program, University of California, San Francisco CA 94143-0452. ³These authors contributed equally to this work.

⁴Corresponding author:

Zena Werb, Ph.D.

Department of Anatomy, HSW 1323

University of California

San Francisco, CA 94143-0452 USA

Tel.: (415) 476-4622

Fax: (415) 476-4565

Email zena.werb@ucsf.edu

Keywords: metastatic niche; myeloid-derived suppressor cells; polyoma middle T (PyMT); tumor microenvironment

ABSTRACT

It is now appreciated that the microenvironment is a critical contributor to metastasis. Although circulating tumor cells can travel throughout the body, clinical evidence shows that certain types of tumors preferentially seed metastases to certain organs, highlighting that the “soil” in which the cancer “seeds” land plays an important role. Recent work has demonstrated that tumor cells actively remodel and prepare distant organs for metastasis, which involves recruitment of myeloid cells and other bone-marrow derived cells and extracellular matrix remodeling. Using the MMTV-PyMT mouse model of breast cancer, we investigate the early lung metastatic niche. We demonstrate that even at early stages of hyperplasia and adenoma, circulating tumor cells can more readily colonize the lungs of MMTV-PyMT mice compared to their wildtype littermates. We identified several factors expressed at higher levels in early tumor/adenoma cells, including *S100a8*, *S100a9*, *Saa3*, and *Spp1* that have previously been implicated in myeloid cell recruitment. These molecular changes are associated with an increase in CD11b+Gr1+ cells specifically in the lungs. Future experiments will be focused on blocking these systemic factors in MMTV-PyMT mice to determine whether the metastatic niche can be altered to inhibit metastasis.

INTRODUCTION

Of the millions of cancer cells that enter circulation, very few will successfully engraft, survive and proliferate at secondary sites, suggesting that the metastatic process is inefficient. Although breast cancer can spread to any distant site in the body, common sites of metastases include bone, liver, lungs and lymph nodes. In estrogen receptor (ER) and progesterone receptor (PR)-positive breast cancer, metastatic disease can occur many years after diagnosis, recurring most frequently in the bone. Despite this known clinical pattern, the molecular mechanisms underlying organ tropism are still unclear, but probably relate to a complex crosstalk between the tumor cells themselves and the microenvironment of the metastatic site. These observations highlight the metastatic process depends on microenvironmental factors, and imply that certain environments are more hospitable than others. The “soil and seed” hypothesis, first proposed by Paget, posits that tumor cells (the “seeds”) flourish only in appropriate environments conducive to that particular tumor cells’ growth (the “soil”) (Paget 1889; Fidler 2003). An extension of this hypothesis is that tumor cells secrete factors that nourish and prepare the soil for metastasis prior to arrival to facilitate colonization and growth. This idea forms the basis of the early or pre-metastatic niche.

In stem cell biology, the “niche” refers to the specialized microenvironment that supports stem cell maintenance and controls the balance between proliferation and quiescence. Many examples of tissue-specific stem cell niches have been described (Lin 2002; Losick et al. 2011; Hsu and Fuchs 2012). In

cancer, cancer stem cells also reside in specific niches (Sneddon and Werb 2007; Lathia et al. 2011), and may even create niches that disrupt the niche of the normal resident stem cells (Colmone et al. 2008). Recently, work by Lyden and colleagues has demonstrated that during metastatic progression, special niches composed of bone-marrow derived cell (BMDC) clusters accumulate in organs that are future sites of metastasis (Kaplan et al. 2005). Using Lewis lung carcinoma (LLC) and B16 melanoma cells, they demonstrated that these BMDC clusters were characterized by their expression of VEGFR1, CD11b, and c-kit, and interacted with stromal cells to create a microenvironment permissive for the subsequent engraftment of tumor cells. These BMDCs accumulate at areas of increased fibronectin, and secrete proteases such as MMP9, which remodel the microenvironment and accelerate the accumulation of more tumor cells (Hiratsuka et al. 2002; Hiratsuka et al. 2006). In addition, the BMDCs up-regulate integrins and chemokines, such as SDF-1, that promote tumor cell recruitment, attachment and growth. Blocking VEGFR1+ BMDC cellular cluster formation by monoclonal antibodies prevents metastasis, further demonstrating the importance of these cells and the niche.

The initiating signals from tumor cells that signal BMDC activation and recruitment are not entirely clear, but several secreted factors have been proposed to participate in metastatic niche formation. These include VEGF, placental growth factor (PIGF) and fibroblast growth factor (FGF), although the absolute requirement of these factors has not been formally established. Other factors implicated are S100A8 and S100A9, which are up-regulated by immature

myeloid and endothelial cells and further recruit more CD11b⁺ myeloid cells. This ultimately activates NF- κ B signaling via toll-like receptor 4 (TLR4) signaling to create an inflammatory-like state that accelerates the migration of primary tumor cells to metastatic sites (Hiratsuka et al. 2008). In addition, lysyl oxidase (LOX), which cross-links collagen, facilitates CD11b⁺ myeloid cell recruitment, leading to increased MMP2 production and adherence of metastatic tumor cells (Erler et al. 2006). Inhibition of LOX activity prevents metastatic growth. Recently, Karin and colleagues conducted a screen for secreted macrophage activating factors and found that in LLC-conditioned media, versican activates macrophages via TLR2 and TLR6 (Kim et al. 2009). This increases myeloid cell-derived TNF α , which induces vascular permeability and enhances LLC metastatic growth. These studies emphasize the contributions of immune cells and inflammation to the metastatic process (Hanahan and Coussens 2012), and highlight the systemic crosstalk between tumor-derived factors, the immune system and distant organs.

Importantly, myeloid cells expressing CD11b and Gr1 (CD11b⁺Gr1⁺) have been referred to as myeloid-derived suppressor cells (MDSCs) because of their ability to inhibit T cell activation and inhibit the differentiation and maturation of antigen-presenting cells (APCs) from their precursors. There is considerable debate in the field whether these cells are neutrophils, since they express high levels of Ly6G, a neutrophil marker, or if they represent a distinct class of immune cells. In any case, these CD11b⁺Gr1⁺ cells secrete reactive oxygen species (ROS), characteristic of an inflammatory state, which facilitates tumor cell infiltration and adherence at the metastatic site, as well as arginase and nitric

oxide synthase (iNOS). They also express high levels of various MMPs, including MMP2 and MMP9, which increases the bioavailability of VEGF within tumors (Kaplan et al. 2005; Murdoch et al. 2008; Psaila and Lyden 2009). Interestingly, CD11b+Gr1+ MDSCs increase physiologically during pregnancy and exert an inhibitory effect on NK cells, which regulates metastatic permissiveness during pregnancy (Mauti et al. 2011). A comparative gene expression analysis of pregnant mouse lungs and liver and of premetastatic niches in organs of tumor-bearing mice revealed several overlapping genes, including *S100a8* and *S100a9*, demonstrating that increases in pro-inflammatory molecules and MDSCs are not unique to malignant conditions.

One of the major drawbacks of the models used thus far to investigate the pre-metastatic niche is that investigators have relied on transplanting late-stage and extremely aggressive tumor cell lines (e.g., B16, LLC) into mice. Unfortunately, these models do not accurately reflect the biological events and transitions during natural tumor progression. Thus, it is unclear whether the factors that have been identified play any role during early tumor progression, i.e. at the hyperplasia or adenoma stages when a pre-metastatic niche is actually developing. This is critical from a clinical perspective because tumors that are identified early might already be secreting factors that predispose the patient to metastasis many years later. Indeed, tumor cell dissemination is now thought to occur early during breast cancer progression (Husemann et al. 2008), contrary to the notion that metastatic disease only occurs late in advanced carcinomas.

Thus, identifying these early changes may yield novel therapeutic strategies to block the initiation of metastasis and formation of an early niche.

In this chapter, I describe an on-going, collaborative project to ask whether an early metastatic exists in an autochthonous model of breast cancer, the mammary tumor virus (MMTV) promoter-driven polyoma middle T antigen (PyMT) model. This model mimics progressive stages of human luminal breast cancer, and develops lung metastases naturally (Guy et al. 1992; Lin et al. 2003), giving us an opportunity to address whether an early metastatic niche forms in the lungs of these mice. (I refer to this niche as an “early” metastatic niche as opposed to a “pre” -metastatic niche because it is impossible to completely rule-out the presence of single-cell micro-metastases. However, we have never detected lung metastases at these early stages, despite previous reports that tumor cells could be found in the bone marrow (Husemann et al. 2008)). In addition, the expression of biomarkers in PyMT-induced tumors is consistent with those associated with poor outcome in humans, including loss of estrogen and progesterone receptors and the persistent expression of ErbB2/Neu and cyclinD1. Also, the MMTV-PyMT mouse is fully immunocompetent, and PyMT tumors have increased leukocyte infiltration associated with malignant transition. Thus, this model of breast cancer is clinically relevant and mimics many aspects of the human disease.

Therefore, we asked three major questions: 1) Does a metastatic niche exist in the lungs of MMTV-PyMT mice? 2) If so, when does the niche initiate and what molecular and cellular changes occur at this time? 3) What tumor-derived

factors contribute to metastatic niche formation, and does blockade of these factors inhibit metastasis? Here, I describe our preliminary findings, which suggest that the lung metastatic niche is set-up during the hyperplasia/early adenoma stage in MMTV-PyMT mice. This is associated with increased CD11b+Gr1+ cells in the lungs, which accumulate predominantly around metastatic foci. In addition, the tumor-derived cytokines interleukin-17 (IL-17), S100A8/A9, OPN, and VEGF may play an active role in the niche. In the later stages of tumor development, other factors such as GCSF are highly up-regulated, which may be important in maintaining a state of chronic inflammation to further drive tumor progression and metastasis.

RESULTS AND DISCUSSION

A metastatic lung niche is established in MMTV-PyMT mice in the hyperplasia/early adenoma stage of tumor development

To investigate whether a metastatic lung niche exists in the MMTV-PyMT mice (PyMT), we compared the lung metastatic efficiency of PyMT mice with their wildtype (WT) littermates. We used two independently-derived PyMT cell lines from a primary PyMT tumor, one generated in our laboratory (referred to as PyMT-C) and another generated by Halpern and colleagues (Halpern et al. 2006) (referred to as PyMT-VO), which forms osteolytic bone metastases after intratibial injection. These cells were labeled with green fluorescence protein (GFP) and luciferase so that metastatic growth could be monitored over time, and so that our cells could be distinguished from endogenous spontaneous metastases.

To probe the metastatic permissiveness of the lung, we injected intravenously (i.v.) the PyMT-C and PyMT-VO cells. After i.v. injection, MMTV-PyMT mice had more lung metastases compared to their WT littermates, as measured by bioluminescent imaging and direct histological examination of the lungs (**Fig. 7.1A–7.1B**). We injected circulating tumor cells (CTCs) into the mice at 4, 6, and 8 weeks of age, allowed the cells to grow for 3 – 4 weeks, and observed that for all ages tested, CTCs more readily colonized the lungs of MMTV-PyMT mice. Importantly, at 4 and 6 weeks of age, the PyMT mammary glands did not have any palpable tumors, indicating that tumors were in the early stages. Indeed, six-week old PyMT mice displayed mostly hyperplastic glands,

with some areas of early adenoma formation (**Fig. 7.1C**). These results suggest that the lung environment is primed for metastatic colonization and growth at the early stages of tumor formation.

Early adenoma tumors express higher levels of pro-inflammatory mediators

To identify tumor-derived factors responsible for forming the metastatic niche, we isolated tumor cells from the MMTV-PyMT mice at hyperplasia/early adenoma and late carcinoma stages, and compared their gene expression profiles to wildtype epithelium. We analyzed several candidate genes by quantitative PCR (qPCR), and found that at the hyperplasia/early adenoma stage, tumor cells expressed higher levels of pro-inflammatory mediators (e.g. *Il1a*, *Il1b*, *Il17*, *Spp1*, *S100a8*, *S100a9* and *Saa3*), integrins (*Itga4*, *Itga6* and *Itgb1*), myeloid-promoting growth factors (*Csf1* and *Csf3*), extracellular matrix components and remodeling enzymes (*Fn1*, *Lox*, *Loxl2*, *Loxl4*, *Postn*, *Tnc*), matrix metalloproteinases (*Mmp2*, *Mmp7* and *Mmp9*), and angiogenic factors (*Plgf*, *Pdgfc*, *Tgfb1*, *Tgfb3*, *VegfA*, *Vegfb*) (**Fig. 7.2A–7.2F**). Many of these factors have previously been implicated in the metastatic niche, either directly promoting tumor growth, recruiting accessory cells, or by remodeling the microenvironment to make it more favorable for colonization. Interestingly, despite a recent report describing the S1P1-STAT3-axis in establishing the metastatic niche (Deng et al. 2012), we found that PyMT tumors expressed lower levels of *S1pr1* compared to control cells (data not shown). In addition, hyperplasia/early adenoma cells did not

express higher levels of *Vegfc*, an important regulator of lymphangiogenesis, or versican, which was previously demonstrated to regulate macrophage recruitment (Kim et al. 2009).

We were intrigued by several factors identified, which affect the inflammatory state and BMDC recruitment and function. The S100A8/A9 proteins are often up-regulated in cancer, and function as chemoattractants to recruit myeloid cells. These proteins are expressed by both tumor and stromal cells, and can be induced by pro-inflammatory cytokines such as IL-1 and TNF α , as well as by VEGF and TGF β , many which were up-regulated in early PyMT cells. S100A8/A9 proteins also function in an autocrine manner to activate NF-kB signaling in MDSCs. This feed-forward loop maintains the suppressive functions of MDSCs within the tumor microenvironment. In addition, S100A8/A9 proteins facilitate tumor cell homing to metastatic sites and increase tumor cell motility by activating pseudopodia in a p38-MAPK dependent manner. Furthermore, secreted SAA3 proteins induce expression of S100A8/A9 (Gebhardt et al. 2006; Hiratsuka et al. 2006; Hiratsuka et al. 2008; Salama et al. 2008; Lukanidin and Sleeman 2012). Interestingly, *Saa3* was also expressed at higher levels in early PyMT cells. These results point to the SAA3-S100A8/A9 axis as a likely candidate for further investigation in the MMTV-PyMT model. Recently, the S100A8/A9 proteins were shown to promote chemoresistance by activating NF-kB signaling, thus enhancing tumor cell survival (Acharyya et al. 2012).

In addition, osteopontin (*Opn*, also known as *Spp1*) belongs to the small integrin-binding glycoprotein family, and is over-expressed in many tumors. OPN

plays multiple roles during metastasis, and is a sensitive and specific marker in predicting disease progression in many cancers. OPN signals through the $\alpha v \beta 3$ integrin and CD44 glycoprotein, and activates the PI3K/Akt, EGFR-ERK-MAPK, and NF- κ B pathways, promoting growth and angiogenesis, inhibiting apoptosis and activating proteolysis via MMPs. Furthermore, OPN activates BMDC recruitment and promotes the outgrowth of distant otherwise indolent metastases (McAllister et al. 2008). Interestingly, GATA3, a transcription factor with anti-metastatic properties, inhibits *OPN* expression (**Chapter 4**), suggesting that GATA3 expression in primary tumors might modulate the metastatic niche at distant sites. Taken together, our results demonstrate that at the early stages of tumor development, a cohort of genes involved in forming the metastatic niche are highly expressed in MMTV-PyMT. The factors include pro-inflammatory, angiogenic, ECM remodeling, and ECM components, some of which are enter systemic circulation, and contribute to the increased metastatic permissiveness in MMTV-PyMT mice.

Serum from MMTV-PyMT also demonstrate a pro-inflammatory state

To identify systemically secreted factors, we analyzed serum collected from MMTV-PyMT and wildtype mice using a multiplex chemokine/cytokine array to profile 32 proteins and identify proteins released into systemic circulation from the primary tumor. Sera from young (5 – 6 week old) and old (10 – 12 week old) mice were analyzed. Most of the cytokines that were evaluated were unchanged. However, young PyMT mice had increased levels of IL-17 and VEGF, while old

PyMT mice had a dramatic increase in GCSF (**Fig. 7.3A**). Because increased IL-17 was detected early in PyMT tumor cells, and was a candidate identified by both serum analysis and real-time qPCR, we reasoned that IL-17 could play a role in forming the niche.

To test this possibility, we asked whether adding exogenous IL-17 directly to the lungs by intra-tracheal (i.t.) injection would increase metastasis. The probing tumor cells were inoculated into the circulation via tail vein into mice injected i.t. with either IL-17 or PBS (Control). We found that after a two-week period, mice treated with IL-17 had an increased metastatic burden in the lungs (**Fig. 7.3B**). These data support our hypothesis that IL-17, a pro-inflammatory mediator, mediates the enhanced metastatic lung niche in MMTV-PyMT mice.

Early lung metastatic niche formation is accompanied by an increase of inflammatory CD11b+Gr1+ myeloid cells

Because IL-17 and a number of other pro-inflammatory molecules were up-regulated in early adenoma cells, we next sought to determine whether there was an accumulation of inflammatory neutrophils or monocytes at metastatic sites. The CD11b+Gr1+ population of inflammatory myeloid cells participates in angiogenesis, T cell suppression and matrix degradation (Psaila and Lyden 2009; Yan et al. 2010a), features that are important for successful metastasis. Therefore, CD11b+Gr1+ cells have been widely implicated in forming the metastatic niche. Two markers can further subdivide the Gr1+ population: Ly6C,

which marks monocytes, and Ly6G, which marks granulocytes. For the majority of our studies, we do not differentiate between these two populations.

We evaluated whether the lungs of MMTV-PyMT mice had increased CD11b+Gr1+ cells compared to WT mice. Using flow cytometry, we found that this population was increased in early (6 – 7 weeks old) MMTV-PyMT mice compared to littermate controls (**Fig. 7.3C**), coinciding with the hyperplasia/early adenoma stage and increased expression of genes such as *Il17*, *Il1b*, *S100a8*, *S100a9*, and *Spp1*. We did not find an increase in the CD11b+Gr1+ population in the blood, spleen, or mammary glands in mice of this age (data not shown), suggesting that this accumulation is specific to the lung environment, which is the native site for future metastases.

To more comprehensively analyze the myeloid cells present in the lungs, we have isolated CD11b+Gr1+ cells, as well as CD11c+F4/80+ macrophages, from the lungs of MMTV-PyMT and WT mice (**Fig. 7.3D**). Future experiments are aimed at analyzing gene expression of these myeloid cell populations by Illumina microarrays and qPCR. We expect to find that genes such as arginase and inducible nitric oxide synthase (iNOS) (Murdoch et al. 2008), as well as novel genes that might play a role in immune regulation, extracellular matrix remodeling, and growth or apoptotic signaling, will be differentially expressed in cells isolated from MMTV-PyMT mice.

CD11b+Gr1+ myeloid cells accumulate around metastatic cancer cells in the lung

To assess the distribution of CD11b+Gr1+ in the lungs of MMTV-PyMT and WT mice after i.v. injection of PyMT-VO cells, we stained lung sections using Gr1+ antibody. Whereas WT mice had sparsely distributed Gr1+ cells in the lung, MMTV-PyMT had concentrated numbers of Gr1+ cells surrounding the PyMT-VO metastases, which were fluorescently labeled with GFP (**Fig. 7.4A–7.4B**). These cells were present mainly at the margins of the metastases, but also within the interior of the foci (**Fig. 7.4B**). In the WT mice injected with PyMT-VO cells, we did not find continuous clusters of CD11b+Gr1+ cells (**Fig. 7.4C–7.4D**). This suggests that in the absence of a primary tumor, CD11b+Gr1+ cells do not accumulate in the lungs, even after i.v. injection of tumor cells. Because the injected cells are labeled with GFP, we micro-dissected regions of the lung that contained GFP-positive metastatic foci and regions lacking GFP-positive tumor cells and analyzed the percentage of CD11b+Gr1+ cells by flow cytometry. We found increased CD11b+Gr1+ cells around GFP-positive regions compared to GFP-negative areas (**Fig. 7.4E**), indicating that CD11b+Gr1+ cells preferentially cluster around lung metastases. These Gr1+ cells also stained positive for MMP9 (**Fig. 7.4F**).

Does the lung metastatic niche in MMTV-PyMT support metastasis of PyMT-independent cells?

We were concerned that our observation of increased metastasis in MMTV-PyMT mice was due to the use of PyMT-driven cancer cells as our probe of the lung environment. Although PyMT is not expressed on the cell surface, cross-

presentation of endogenous antigens could hypothetically lead to immune tolerance in MMTV-PyMT and not WT mice. Therefore, the difference in lung colonization in MMTV-PyMT and WT mice could be due to an immune related response specifically to foreign PyMT antigen in WT mice, which would mount an immune attack to destroy these foreign cells.

One argument against this possibility is that PyMT-VO and PyMT-C cells grow orthotopic tumors and form bone metastases (PyMT-VO) very readily in WT FVB/n mice, indicating that immunocompetent WT mice tolerate these tumor cells and do not completely reject them. However, to formally exclude a PyMT-specific immune response, we generated several non-PyMT FVB/n cell lines and tested them in our system (see **Appendix D**). This included cell lines derived from the MMTV-Her2/Neu breast cancer model (NeuC1), the C3(1)-Tag breast cancer model (C3T-R1T), as well as a cell line derived from a spontaneous mammary adenocarcinoma (MaP0008) (Huang et al. 2008). In addition, we used two non-breast cancer FVB/n tumor lines, one derived from a spontaneous lung adenomacarcinoma (LaP0297) and from a Myc and Ras driven liver tumor (LT2MR). We transduced these cells with GFP and luciferase so that we could monitor metastatic growth using the IVIS bioluminescent. We are currently titrating the dosage of cells and determining whether they are competent to grow in the lungs. Preliminary results with LT2MR cells injected i.v. into female mice indicate that these cells are rapidly cleared from both the PyMT and WT mice. Despite initial seeding into the lung, these cells fail to form macroscopic metastases, and are most likely cleared by the immune system. One possible

explanation is that the LT2MR cells originated from male mice, and we are injecting into female mice. Future work is aimed to optimize the system for the other cell lines to test in our model. These results will allow us to determine whether the lung metastatic niche in MMTV-PyMT mice supports the colonization and growth of non-PyMT cells.

Does adaptive immunity play a role in the early metastatic niche?

The adaptive immune system has also been shown to play an important role in tumor development and metastasis. For example, B lymphocytes are required for establishing chronic inflammatory states that promote de novo carcinogenesis in a K14-HPV16 mouse model of skin cancer (de Visser et al. 2005; Andreu et al. 2010). In breast cancer, Coussens and colleagues have shown that CD4⁺ T cells promote metastasis by regulating macrophages (DeNardo et al. 2009). In addition, Karin and colleagues have shown that Tregs promote metastasis by supplying RANKL to breast tumors (Tan et al. 2011). Taken together, these results indicate that the adaptive immune system plays a critical role regulating metastasis and the tumor microenvironment.

We were interested in whether the adaptive immune system plays a role in the early metastatic niche, and whether B and T cells promote formation of the lung niche, or whether B and T cells might be important in rejection of the probing i.v. injected cells. To test this, we have begun to cross the MMTV-PyMT onto JH^{-/-} mice (to delete B cells) and CD4^{-/-}CD8^{-/-} mice (to delete T cells) (de Visser et al. 2005). If B or T cells are important in promoting the early metastatic niche, we

expect that the increase in lung metastases in the PyMT mice will be diminished once the cells are deleted. If B or T cells are important for immune rejection, we expect that the mice will have increased lung metastases. Any outcome from these experiments will be interesting and will allow us to elucidate the role of the adaptive immune system in early metastatic niche formation in this model of breast cancer.

CONCLUSION

We have demonstrated that an early metastatic niche exists in the lung in a transgenic mouse model of breast cancer. Although previous studies have shown that this niche exists and is established prior to tumor cell arrival, our model sheds new insight into the timing of this process because the model recapitulates many features of human breast cancer, including the natural progression from hyperplasia to adenoma and then to carcinoma. Therefore, we find it more informative than those experiments published in the literature, which often use late-stage, highly aggressive cell lines to make claims about the “pre”-metastatic niche. In addition, we have identified several important candidates that we will pursue, including MMPs, the S100A8/A9 family of proteins, LOX and LOXL-family members, various interleukins (e.g., IL-17), and OPN. We plan to block each these factors in the MMTV-PyMT mice, starting with IL-17 and MMP2, MMP9 and MMP7, and asking whether perturbation of these molecules diminishes the metastatic permissiveness of the lung. This work may help form the basis of novel therapeutic approaches to target the metastatic niche early during tumor progression, which may block or delay the time to metastasis in human patients and improve survival outcomes for this currently untreatable disease.

ACKNOWLEDGEMENTS

We thank members of the Werb Lab for many helpful discussions and input into the project. We thank Elena Atamaniuc and Ying Yu for animal assistance, Ying Yu for help with bioluminescent imaging, Marja Lohela for maintaining the MMTV-PyMT colony and Helen Capili for histology assistance. We also thank the following students, who have contributed to various aspects of the project: Joanne Dai, Jennifer Tai, Eline van Kappel, and Kelly Kersten. We thank the UCSF Biological Imaging Development Center for microscope use and microscopy advice. Finally, we thank Sylvain Provot and Jay Kim for establishing the PyMT-C cell line, and Conor Lynch (Vanderbilt University) for providing the PyMT-VO cells.

EXPERIMENTAL PROCEDURES

Animal Studies

All animal experiments were performed at UCSF, and reviewed and approved by the UCSF IACUC. Mice were housed under pathogen-free conditions in the UCSF barrier facility. MMTV-PyMT mice on FVB/n background, originally from Jackson Laboratories, were bred in-house. To probe the metastatic permissiveness of the lungs, 6 week old female PyMT mice and their control littermates were injected i.v. (via tail vein) with 1×10^5 cells (PyMT-VO or PyMT-C cells) in D-PBS (without Mg^{2+} or Ca^{2+}). Bioluminescence imaging was performed weekly using an IVIS Spectrum (Caliper Life Science). Image radiance values were normalized using the Living Image software. Mice were sacrificed 4 weeks post injection, and tissues collected for processing.

For the IL-17 experiments, wildtype 6 week old female FVB/n mice were first pre-conditioned with intra-tracheal (i.t.) injection of 1 μ g of recombinant IL-17A (R&D) or PBS. Mice were then i.v. injected with 1×10^5 tumor cells, and i.t. injected with IL-17A or PBS twice a week for two weeks. Lung metastases were quantified by counting nodules and by fluorescence intensity.

Cell culture

The PyMT-C cell line was generated by isolating a late stage tumor from the MMTV-PyMT/FVB mouse, dissociating the cells in collagenase. The PyMT-C cells were cultured in DMEM/F12 media supplemented with 5% FBS, insulin and hydrocortisone. The PyMT-VO line was provided by Conor Lynch, and cultured in

DMEM supplemented with 10% FBS. Cells were transduced with retroviruses carrying GFP and luciferase, and were sorted using FACS or selected in puromycin for at least 5 days.

Serum cytokine analysis and ELISA

Sera from MMTV-PyMT mice and their wildtype littermates were collected from young (6 – 8 weeks old) and old (10 – 13 weeks old) mice. Samples were allowed to clot, spun, and the serum fraction was then removed. Samples were analyzed in duplicate by Eve Technologies using a multiplex bead platform. n = 4–5 samples were obtained per condition.

Flow Cytometry

To sort CD11b+Gr1+ and CD11c+F4/80+ cells from the lungs, lungs from MMTV-PyMT and their wildtype littermates were digested with collagenase. Lungs were further digested with trypsin to dissociate into single cells. The cells were blocked with rat IgG, and then stained with antibodies against CD11b, CD11c, Gr1 and F4/80 (all from eBioscience). Analysis and cell sorting was performed on FACS Aria II (Becton Dickinson), and analyzed using FlowJo software or FACSDiva software (BD Biosciences).

Quantitative real-time PCR (qPCR)

Total RNA was isolated from primary tumor cells using the miRNeasy Mini Kit (Qiagen). cDNA was synthesized using the Superscript III RT First Strand Kit

(Invitrogen). qPCR was performed using FastStart Universal SYBR Green master mix (Roche) in an Eppendorf Mastercycler realplex machine. Ct values were normalized to actin and GAPDH, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. Primer sequences for qPCR were found using the Harvard Primer Bank, and are detailed in Supplementary Table 1.

Immunostaining

Tissues were fixed in 4% PFA overnight, paraffin processed or embedded into OCT for frozen sections, and then sectioned. Antigen retrieval was performed using citrate heat-mediated retrieval or using proteinase K. The TSA Amplification Kit (Perkin Elmer) was used according to manufacturer instructions. Primary antibodies were incubated overnight, and secondary antibodies were incubated for 1 hour. The following antibodies were used: Gr1 (BD Biosciences), MMP9 (Abcam), GFP (Abcam), F4/80 (Invitrogen), biotinylated anti-rat (Jackson), and biotinylated anti-rabbit (Dako). Fluorescent secondary antibodies were from Molecular Probes (Invitrogen). Images were taken on a C1Si confocal microscope (Nikon) or spinning-disc confocal microscope (Yokogawa). Image analysis was performed using Imaris and ImageJ software.

Statistical analysis

Statistical analysis was performed using Prism 4 software (GraphPad Software, Inc.). All data are presented as mean \pm SEM, unless otherwise stated. When two groups were compared, the two-tailed Student's t-test was used, unless

otherwise stated. When three or more groups were compared, the one-way analysis of variance (ANOVA) test was used, followed by Tukey's test to determine significance between groups. We considered $p < 0.05$ as significant.

FIGURE LEGENDS

Fig. 7.1: During early stages of tumor development, the lungs of MMTV-PyMT mice are more permissive to circulating tumor cells colonization and growth.

(A–B) Littermate controls or MMTV-PyMT 6-week old mice were injected with luminescent PyMT-VO cells and monitored for metastases by bioluminescent imaging three weeks after intravenous (i.v.) injection. MMTV-PyMT mice show a significant increase in the metastatic luminescent signal. Lungs metastases were confirmed by histological examination (H&E). n = 10 mice per group.

(C) Mammary glands from control littermates or MMTV-PyMT mice were examined histologically to determine the degree of tumor development. At 6-weeks old, MMTV-PyMT mammary glands were mostly hyperplasias and early adenomas. Scale bars = 200 μm (B) and 100 μm (C).

Fig. 7.2: Gene expression changes associated with increased inflammation, extracellular matrix (ECM) adhesion and remodeling, matrix metalloproteases, and angiogenesis are manifest early in primary adenomas and sustained during tumor progression.

(A–F) Primary mammary epithelial cells were isolated from littermate control, young (6 – 8 week old) or old (10 – 13 week old) MMTV-PyMT mice. Gene expression was profiled by quantitative PCR (qPCR). Inflammatory molecules (A), integrins involved in adhesion and differentiation (B), chemokines and myeloid-cell promoting growth factors (C), extracellular matrix (ECM) enzymes

and components (D), matrix metalloproteases (E), and angiogenic growth factors and signaling molecules (F) were assessed. n = 4-8 samples per group.

Fig. 7.3: A pro-inflammatory state in the lungs of MMTV-PyMT mice is characterized by systemic changes in the serum and increased CD11b+Gr1+ cells.

(A) Sera from young and old MMTV-PyMT mice and littermate controls were analyzed using a multi-plex platform. Most changes are characterized later in tumor development when primary growths are at the carcinoma stage.

(B) Wildtype FVB/n mice pre-conditioned by intra-tracheal (i.t.) injection of recombinant IL-17A had a higher lung metastatic burden after i.v. injection of tumor cells. n = 3 mice in each group.

(C) Lungs from MMTV-PyMT mice or littermate controls were analyzed for CD11b+Gr1+ cells by flow cytometry. Representative FACS plots are shown on the right. n = 6-8 mice in each group.

Fig. 7.4: Gr1+ myeloid cells accumulate around metastases only in the lungs of MMTV-PyMT mice, and express MMP9.

(A–D) Confocal microscopy of MMTV-PyMT (A–B) or littermate control (C–D) lungs stained with GFP (green) to mark the injected PyMT-VO cells and Gr1 (red). Gr1+ cells accumulate and surround the metastatic foci, and are absent in the littermate control mice.

(E) Lungs from MMTV-PyMT mice i.v. injected with PyMT-VO cells labeled with GFP were dissected into GFP-positive and GFP-negative regions (either containing or lacking metastases, respectively) using a fluorescence dissection scope. These regions were digested into single cells and analyzed for CD11b and Gr1 by flow cytometry.

(F) Confocal microscopy of MMTV-PyMT lungs stained with GFP (green), MMP9 (red) and Gr1 (blue) shows colocalization of Gr1⁺ cells with MMP9 (denoted by arrowheads). Scale bar in A and C = 100 μm ; in B, D and F = 25 μm .

FIGURES

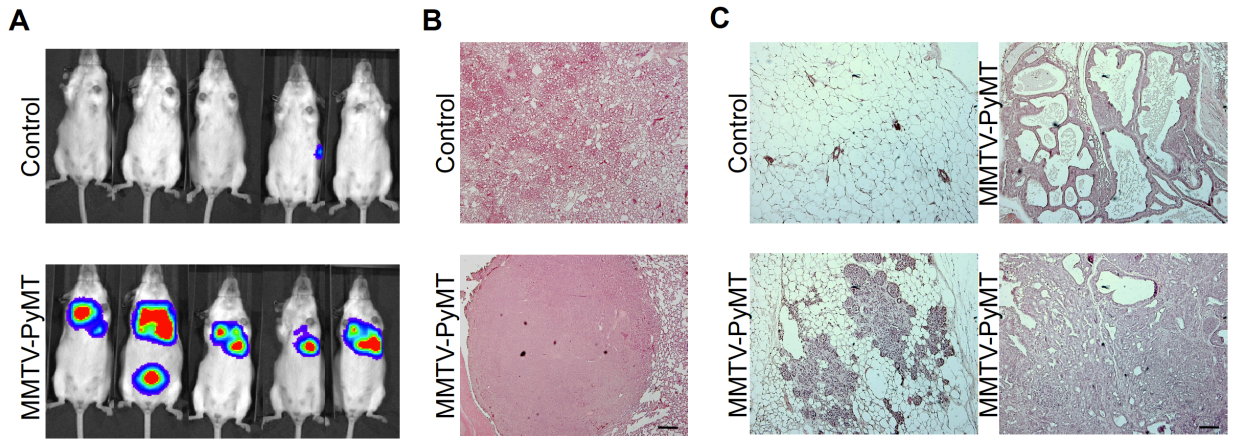


Figure 7.1

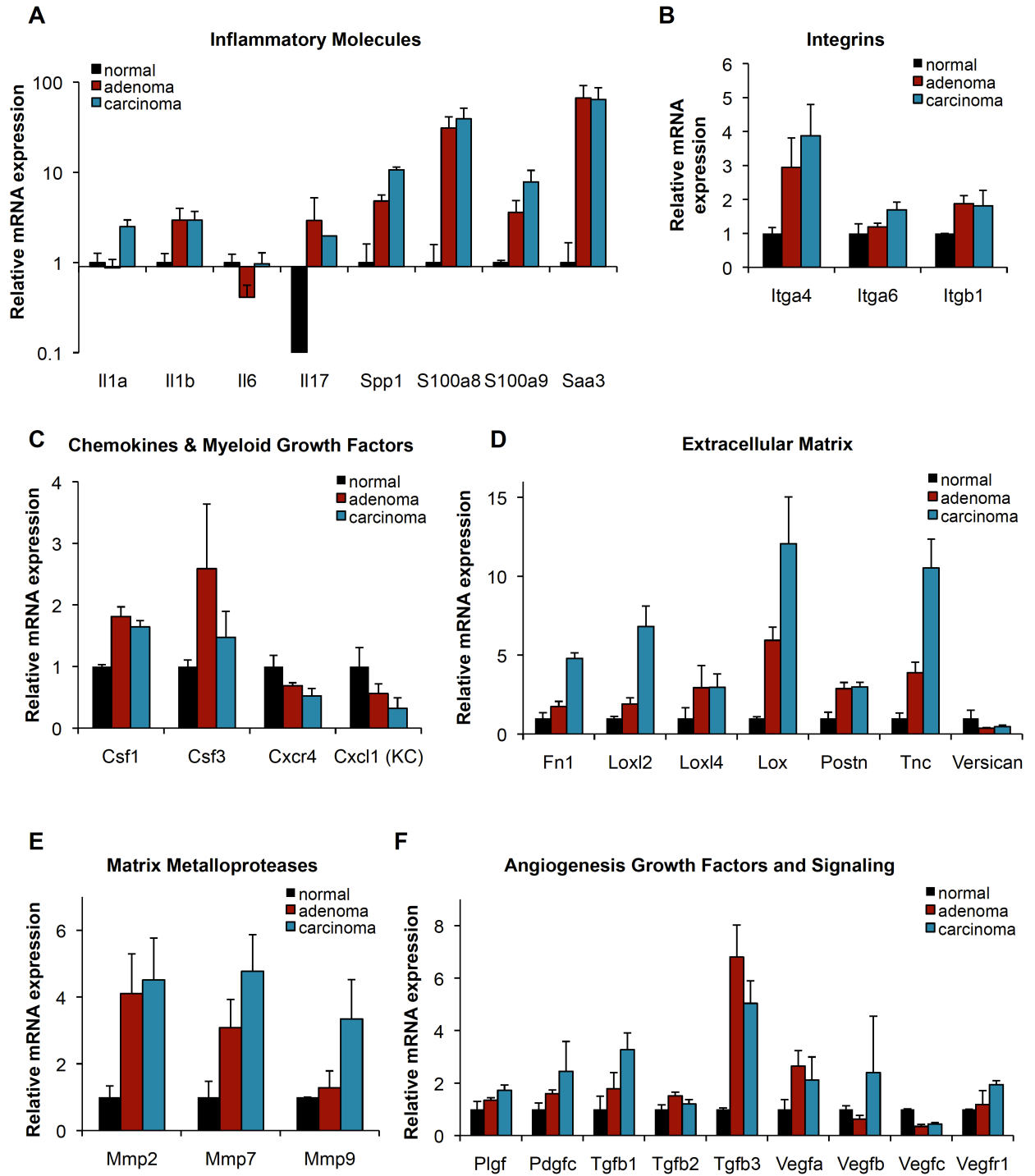


Figure 7.2

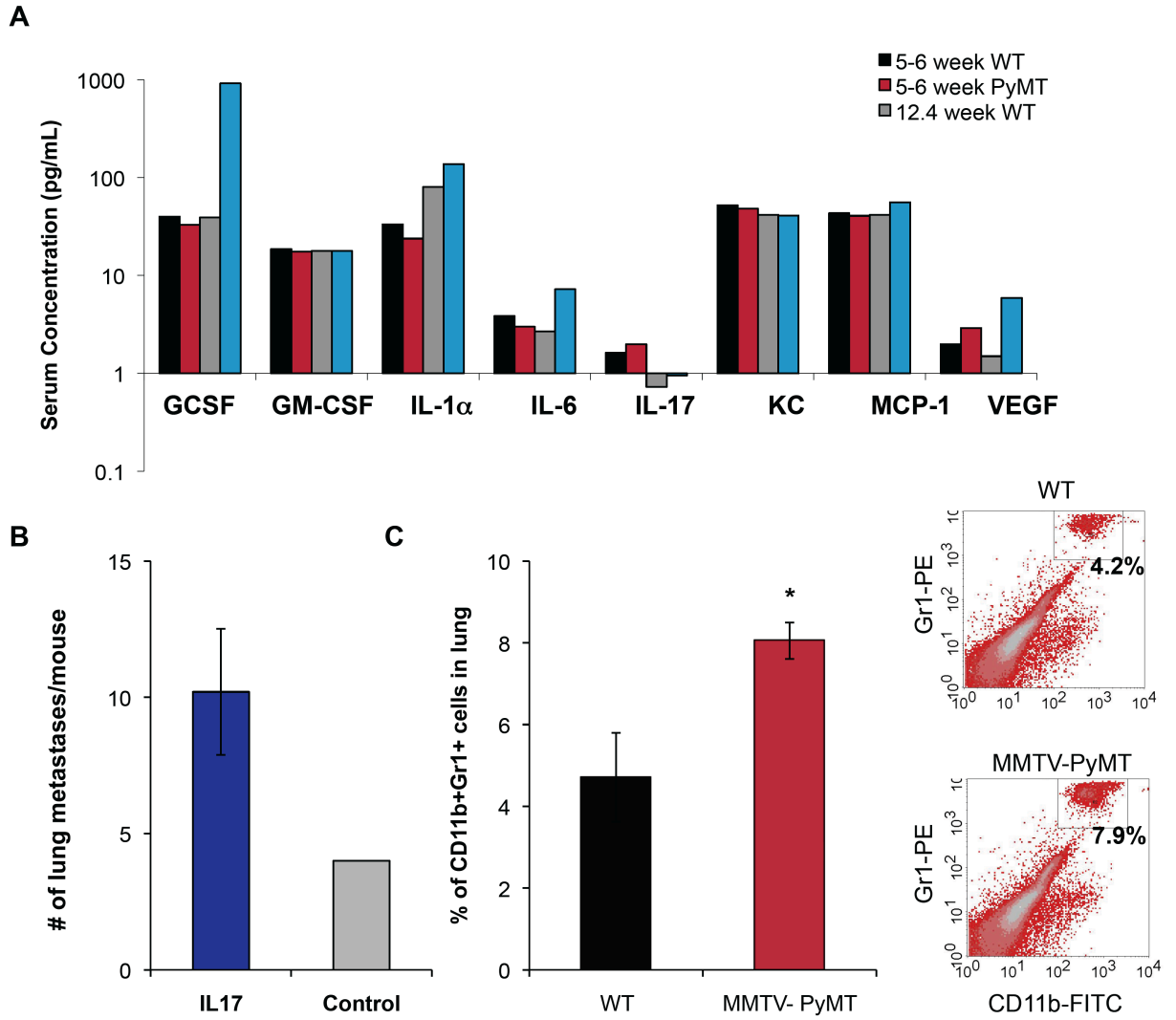


Figure 7.3

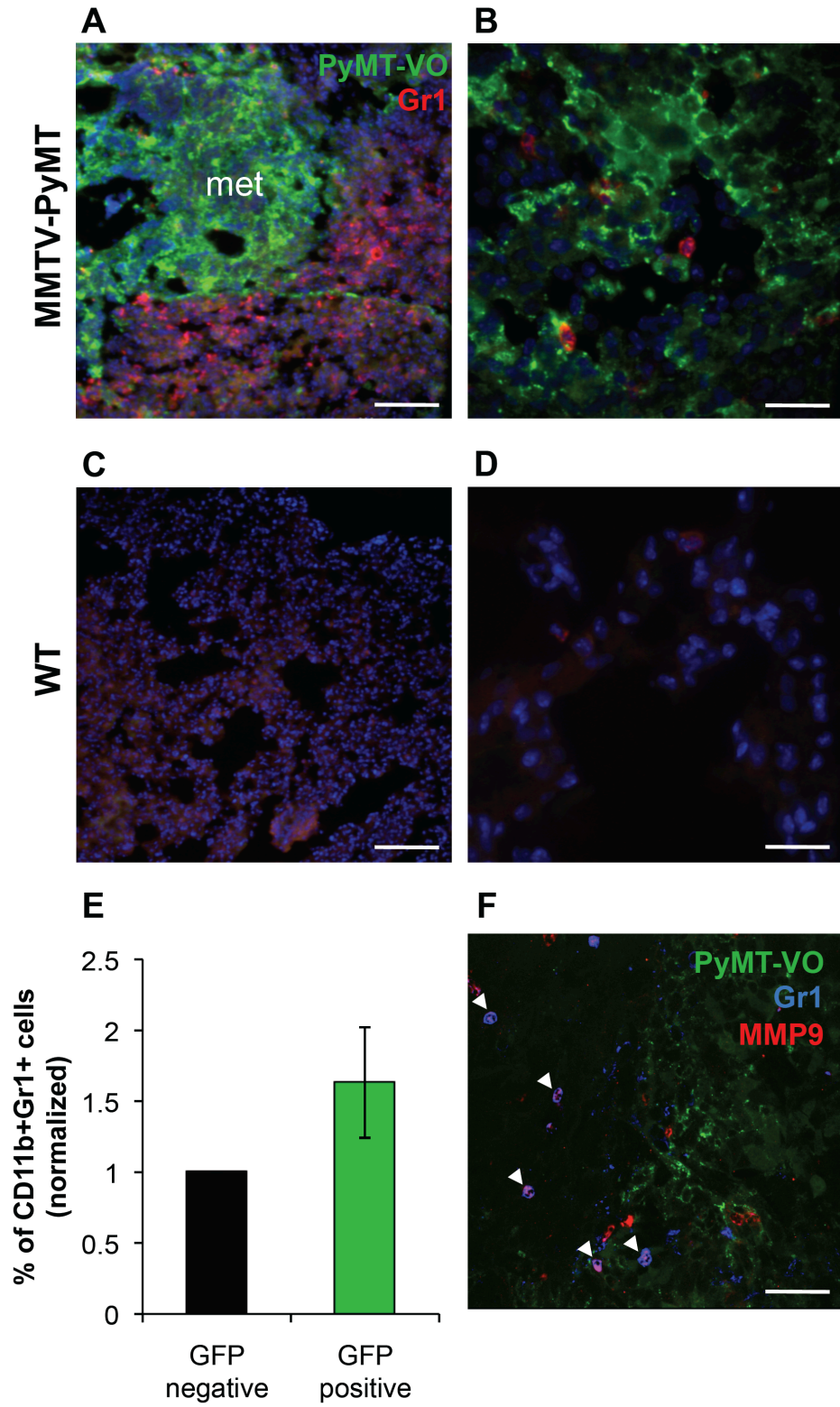


Figure 7.4

Chapter 8: Conclusions

Source: This chapter summarizes the content of my dissertation and contains outstanding questions and experiments that follow the work contained in this dissertation, as well as thoughts and perspectives of where the field of cancer metastasis research might head in the next decade.

Contributions: This is an original perspective written by me, with helpful discussions with members of the Werb Laboratory.

Chapter 8: Conclusions

Breast cancer remains the most common malignancy that affects women today. Most of the deaths from breast cancer are attributable to metastatic disease; unfortunately, whereas surgical resection and adjuvant therapy can cure well-confined primary tumors, metastatic disease is largely incurable because of its systemic nature and resistance to current therapies. Indeed, the median survival of breast cancer patients with distant metastases is between one and two years, and only a quarter of patients will survive beyond five years (Jones 2008). These grim statistics underscore the need to better understand cancer metastasis and the cellular and molecular mechanisms that underlie this process, so that novel therapeutic approaches to prevent and treat metastases can be brought into the clinic.

Tumors as organs: the importance of the tumor microenvironment

In 2000, Hanahan and Weinberg published a seminal review, *The Hallmarks of Cancer*, which outlined six common challenges that cancer cells must overcome (Hanahan and Weinberg 2000). They provided a conceptual framework for thinking about and understanding the complexities of this disease. Much of the focus on cancer research up until that point had been on the tumor cell itself and the intrinsic biochemical pathways and signal transduction networks that endow a cell with the properties required to become a bona fide cancer. Work done to understand oncogenes and tumor suppressors, whose respective activation or loss contribute to transformation and/or uncontrollable growth provided us with a

wealth of information, including important signaling pathways that cancer cells use to promote their growth and survival. This has subsequently led to the identification of molecular targets and the development of novel drugs, including HER2/ERBB2 (Herceptin), the EGF receptor (Cetuximab, Gefitinib) BCR-ABL (Gleevec) and others. Despite these successes, there remain many oncogenic pathways that remain un-druggable (e.g., Myc, Ras, β -catenin). Furthermore, the effects of these drugs are often transient as the cancer cells develop resistance, and in patients with more advanced disease these drugs have not significantly improved patient survival rates. In addition, current treatments have not eliminated the progression to metastatic disease.

Over the past decade, work done in numerous laboratories, including our own lab, has uncovered the role of the tumor microenvironment as a central mediator of tumor progression and metastasis (Joyce and Pollard 2009). We have begun to understand how tumor cells co-opt stromal cells to promote malignant growth; how tumor cells interact with immune infiltrates; how tumor cells secrete and regulate proteases and other enzymes to degrade or remodel the extracellular matrix (ECM); and how tumor cells regulate angiogenesis (Bergers et al. 2000; Orimo et al. 2005; DeNardo et al. 2009; Levental et al. 2009; Andreu et al. 2010; Kessenbrock et al. 2010; Engelhardt et al. 2011; Nakasone et al. 2011; Tan et al. 2011). Furthermore, the microenvironment at distant metastatic sites is now recognized as an important contributor to tumor progression and is only beginning to be understood. This metastatic niche consists of extracellular matrix components (such as periostin, tenascin and lysyl

oxidase), as well as various inflammatory cells, including a population of myeloid cells characterized by expression of CD11b and Gr1 (Kaplan et al. 2005; Erler et al. 2009; Psaila and Lyden 2009; Malanchi et al. 2011; Oskarsson et al. 2011). These studies all highlight the reciprocal interactions between the tumor and the various cell types present within the tumor and metastatic microenvironments, changing the way we think about tumors (Egeblad et al. 2010).

The importance of the tumor microenvironment has been further emphasized in the updated version of Hanahan and Weinberg's review (Hanahan and Weinberg 2011). Indeed, although this complexity makes the study of tumors more daunting, it also means that new therapeutic opportunities to target the tumor microenvironment will emerge. The challenge is to identify the combination of targets and cell types to hit that will be the Achilles' heel for each patient's specific cancer. Therefore, understanding the mechanisms that regulate the tumor and metastatic microenvironments and thinking about the diverse cell populations that comprise a tumor will be critical to successfully combating metastasis.

Factors that promote and suppress tumor invasion and metastasis

Although the classical hallmark of tumor invasion and metastasis was described over a decade ago, it is surprising how little was known then. While it was clear that alterations in cell adhesion was a prerequisite to invasion so that cells could detach from each other, little was known about how adhesion genes such as E-cadherin (*CDH1*) were down-regulated. Loss of E-cadherin was frequently

observed in invasive carcinomas, which suggested that E-cadherin functioned as a tumor suppressor, but the mechanism underlying its down-regulation and its connection to differentiation and the epithelial-to-mesenchymal transition (EMT), were largely unknown. Genes that promote and suppress metastasis were not yet identified (Smith and Theodorescu 2009), and non-coding RNAs such as microRNAs or lincRNAs (large intervening non-coding RNAs) that mediate tumor invasion and metastasis were completely unknown (Calin and Croce 2006; Ma et al. 2007; Gupta et al. 2010; Valastyan and Weinberg 2011). I anticipate that there will continue to be significant progress made into identifying and understanding metastasis promoting and inhibiting factors (including classical genes, as well as microRNAs, lincRNAs and epigenetic modifiers) in the next decade.

Links between tumor differentiation and cancer stem cells

While pathologists have long appreciated the connection between tumor differentiation and tumor aggressiveness, a molecular understanding of this phenomenon was not well understood. There is now increasing evidence that loss of differentiation (e.g., through an EMT) generates cells with stem-like properties, giving the cells self-renewal capacities and motility (Mani et al. 2008). Moreover, many of the EMT-inducing transcription factors (e.g., Slug) also control stem-like properties in both normal and cancer cells, often in collaboration with factors such as Sox9 (Guo et al. 2012). Because un-differentiated tumors are generally more aggressive and less favorable, this suggests that reverting EMT by enhancing differentiation pathways, (i.e. by stimulating a pro-

differentiation factor such as GATA3), might suppress metastasis. In addition, promoting differentiation might also lead to greater sensitivity to therapy. Indeed, cancer stem cells are generally quite resistant to conventional therapy, which may explain why tumor shrinkage is often followed by tumor re-growth. Efforts to selectively eliminate cancer stem cells have yielded some promising results (Gupta et al. 2009), in part by inducing differentiation. This suggests that promoting differentiation might be a complementary strategy to reducing the aggressiveness of cancers.

Lessons from development and their extension to cancer

The mammary gland is a unique organ in that most of its development occurs during puberty and pregnancy (Khokha and Werb 2010). In response to hormones during puberty, the mammary gland undergoes a rapid period of proliferation and invasion into the mammary fat pad. This invasion ends when the gland reaches the distal limits of the mammary fat pad, which is mediated in part by TGF β signaling. Then during pregnancy, extensive mammary epithelial proliferation occurs, resulting in a greater than ~100-fold increase in cell number. This is accompanied by further cell differentiation and polarization into secretory, milk-producing units in a relatively short gestation period of ~20 days in the mouse. After lactation, the mammary gland involutes, which involves rapid regression of the differentiated gland and activation of cell death programs to eliminate most of the secretory epithelium within a week of weaning. Furthermore, the ECM undergoes dramatic changes during this process.

Eventually, the mammary gland returns to a state that resembles the pre-gestation state, devoid of secretory alveoli. Studying these physiologic stages of proliferation, invasion, differentiation and regression during normal development thus gives us insight into mechanisms that might be hijacked and utilized during tumor progression. Indeed, cancers can often be viewed as development gone awry. Therefore, understanding the genes and microRNAs responsible for proliferation, invasion, differentiation and regression during normal development is critical to understanding mechanisms utilized in cancer.

The plasticity of the mammary epithelium and its distinct developmental stages (e.g., proliferation and involution in response to hormonal cues) are also intimately tied to breast cancer risk in humans. Epidemiological studies have revealed, for example, that parity and breast feeding are considered protective against breast cancer when compared with nulliparous women and those who have never breast-fed (Akbari et al. 2010). In addition, a higher number of menstrual cycles, either due to early menarche or late menopause, increases the risk of developing breast cancer (Veronesi et al. 2005). The involuting mammary gland after pregnancy has also been linked to an increased risk for more aggressive cancer and metastasis. Those women diagnosed within five years of pregnancy have worse survival outcomes and greater risk of developing metastatic disease (Schedin 2006). These studies underscore the need to better understand the links between normal development and cancer.

Putting it all together – summary of thesis

My graduate work has been focused on the aforementioned themes: understanding the tumor microenvironment; elucidating factors that promote and suppress metastasis; investigating the link between tumor differentiation and tumor aggressiveness, including the role of cancer stem cells; and studying normal development as a platform to learn more about cancer biology.

In **Chapter 2 and 3** of this thesis, I have illustrated the powerful pro-differentiation effects of the GATA3 transcription factor on breast cancer progression and metastasis. I have shown that miR-29b is an important node downstream of GATA3 that promotes luminal differentiation and down-regulates many pro-metastatic components of the tumor microenvironment, including angiogenic factors, metalloproteinases, integrins and remodelers of the ECM. Loss of miR-29b, even in GATA3-expressing cells, increases metastasis and results in a de-differentiated, more mesenchymal phenotype. Re-expressing these pro-metastatic regulators in miR-29b-expressing cells restores metastasis. Taken together, our results demonstrate that a GATA3-miR-29b axis is critical to regulating tumor differentiation, the microenvironment and metastasis.

In addition to regulating lung metastasis, I have shown in **Chapter 4** that GATA3 also inhibits bone metastasis, likely by suppressing osteopontin (*Opn* or *Spp1*) expression. Interestingly, knockdown of osteopontin suppresses both lung and bone metastasis in experimental metastasis models. Even though more work remains to be done to conclusively demonstrate the importance of GATA3-

mediated regulation of osteopontin in bone metastasis, our results are encouraging and illustrate the general anti-metastatic role of GATA3.

Although we believe that GATA3 regulation of miR-29b and osteopontin is important, we acknowledge that these two factors are not the only downstream targets of GATA3. Indeed, GATA3 expression in breast cancer cells leads to widespread gene expression changes, as shown by both our group and other groups (Dydenborg et al. 2009; Asselin-Labat et al. 2011; Chu et al. 2012). More work to uncover novel GATA3 targets in breast cancer cells using non-biased, genome-wide techniques will be important to fully understand how this pleiotropic transcription factor works. Indeed, we have described in **Chapter 5** preliminary results from a collaborative ChIP-Seq study with Dr. Peggy Farnham's group to identify new GATA3 binding sites in mammary epithelial cells. We anticipate that follow-up studies will allow us to discover new mediators of luminal cell differentiation and metastasis suppression.

Together, this work highlights the need to expand our concept of how traditional "tumor suppressors" might function. Candidate factors must include both classical genes and also non-coding RNAs, and assays should investigate non-traditional effects on tumors such as differentiation instead of focusing solely on size as the primary endpoint. The effects of tumor differentiation on the mesenchymal and progenitor/stem cell populations will also be important to understand, both from a developmental biology standpoint, as well as from a therapeutic perspective.

In **Chapter 6**, I have demonstrated the importance of microRNAs in normal mammary development by using a conditional knockout of *Dgcr8*. Global down-regulation of microRNAs leads to a crisis in normal differentiation and suppresses the normal organization of luminal and basal cells, leading to an inability to successfully undergo branching morphogenesis in 3D Matrigel culture. Loss of *Dgcr8* also leads to up-regulation of a stem/progenitor-like gene expression program. Although miRNA biogenesis components have been described as tumor suppressors in other cancer types (Kumar et al. 2009; Ravi et al. 2012), we found using a murine breast cancer model that loss of *Dgcr8* inhibits cell proliferation and tumor growth. This work suggests that total loss of microRNAs can severely cripple tumors. In contrast to other studies (Martello et al. 2010), our work warns against restoration of miRNA biogenesis as a general strategy to inhibit tumor growth and metastasis. Further work to restore specific miRNAs in *Dgcr8*-null cells will elucidate the specific roles of individual miRNAs.

In **Chapter 7**, I have described our investigations into the early metastatic niche in the MMTV-PyMT breast cancer mouse model. This model recapitulates many aspects of human breast cancer, and the epithelial cells naturally progress from hyperplasia to adenoma to carcinoma and lung metastasis. Thus, the MMTV-PyMT serves as an excellent model to investigate how and when the metastatic niche develops. We have found that early during tumor progression (when the glands exhibit hyperplasias and adenomas), the MMTV-PyMT lungs are much more hospitable for colonization and metastatic growth compared to the lungs of wildtype littermates. Using a candidate qPCR screening approach,

we sought to identify factors that might mediate this effect and found an increase in several growth factors, cytokines, metalloproteinases and ECM molecules previously implicated in the pre-metastatic niche (Psaila and Lyden 2009). Our results yielded a significant number of pro-inflammatory molecules, including *S100a8*, *S100a9*, *Saa3*, *Il1b*, *Il17*, *Tgfb* and others. The role of inflammation has been widely appreciated in promoting cancer, especially the contributions of myeloid-derived cells (Coussens and Werb 2002; Murdoch et al. 2008). Consistent with this, we have found that MMTV-PyMT mice have an increased number of CD11b+Gr1+ cells in their lungs that accumulate around lung metastases. These cells express MMP9 and may contribute to metastatic niche formation by remodeling the ECM (Yan et al. 2010a). In addition, we have conducted a cytokine screen to search for systemic factors in the serum of MMTV-PyMT mice. This has revealed high levels of the pro-inflammatory cytokine GCSF, which has previously been implicated in promoting metastasis (Kowanetz et al. 2010). Taken together, our study demonstrates the importance of the lung microenvironment and suggests that tumors may secrete factors that prepare distant sites for metastasis early during tumor progression. Although still preliminary, this work suggests that prophylactic treatment of metastatic disease, even in patients with early stage cancer, might be beneficial.

Finally, I have included four additional chapters in the Appendix. In **Appendix A** and **Appendix B**, I have described work to understand two novel metastasis-promoting transcription factors, *Zeppo1/Znf703* and *Znf217*, which are both associated with reduced survival and often amplified in various cancer

types. We have found that Zpo1 regulates cell adhesion proteins such as E-cadherin and p120-catenin to promote migration, invasion and metastasis (Slorach et al. 2011), whereas ZNF217 promotes a de-differentiated, progenitor-like phenotype, metastasis and chemoresistance (Littlepage et al. 2012).

In **Appendix C**, I have described a collaborative project with my wife, Charina Choi, to develop next-generation tetrapod quantum dots as tools to measure cellular forces in biology. We have illustrated that these nanocrystal tools can be used to study beating cardiomyocytes, and anticipate that future work to incorporate these tetrapod quantum dots into more physiologic systems such as collagen and Matrigel will allow studies to be done in three-dimensional geometries. In **Appendix D**, I have described the development and characterization of some new FVB/n breast cancer cell lines. I hope that these lines will be useful for future investigations.

Perspectives and future directions

As I leave graduate school, I am excited to return back to medical school and finish my medical training, but I also realize the many exciting questions that remain unanswered. I have outlined several questions that stem directly from my graduate work that I hope will be addressed in the near future.

Uncovering new targets of GATA3

What other genes and miRNAs does GATA3 regulate? One of the exciting aspects of GATA3 is that it binds a specific DNA motif, but how it recognizes its

targets in different cell types is largely unknown. Presumably, GATA3 regulates different targets by interacting with other transcriptional and epigenetic regulators. Evidence for the idea that master transcription factors act collaboratively and in cell-type specific manners was elegantly shown for GATA1 and Smad3 (Mullen et al. 2011; Trompouki et al. 2011). This pair of papers shows how cell lineage regulators like GATA1 can direct the BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. Uncovering these targets and protein interaction partners will be exciting areas of biology that will deepen our fundamental understanding of cell differentiation and provide new targets for therapeutics.

Future work will determine the broad repertoire of GATA3 targets, which will undoubtedly be involved in diverse functions such as promoting differentiation, regulating collective cell behaviors like branching morphogenesis, modulating the immune response or reorganizing the cytoskeleton and affecting cell motility. The ChIP-Seq experiments in MDA231 cells have already begun to address this. From the preliminary target list, a number of genes come up whose functions are yet unknown in mammary development, metastasis or tumor progression. For example, a group of semaphorin family members are bound; these molecules play roles in neuronal axon guidance, but little is known about its role in mammary gland development (Harburg and Hinck 2011). Indeed, semaphorins have been implicated in angiogenesis and other aspects of tumor biology (Neufeld and Kessler 2008). In addition, other targets include the dual specificity phosphatases (DUSP), which regulate signaling flux through the

mitogen-activated protein kinase (MAPK) pathway and have been implicated in tumor progression (Bermudez et al. 2010). The TGF β pathways also emerges prominently, which will be interesting to study given its importance in differentiation, EMT and maintenance of stem-like properties (Scheel et al. 2011).

In addition, it will be interesting to cross compare targets identified from different cell lines. This will reveal both the common and the unique sequences to which GATA3 is bound and shed insight into how GATA3 collaborates with other factors. Lastly, identification of GATA3 targets in primary human patient samples will give insight as to whether GATA3-bound targets in vivo are similar to targets identified in MDA231 cells.

Potential therapeutic applications of the miR-29 family

One significant finding of our work is that miR-29b is a central regulator of microenvironmental genes. For example, we have shown that miR-29b regulates collagen remodeling by down-regulating not only collagen itself, but also Lox and multiple members of the lysyl oxidase family of enzymes (Loxl2 and Loxl4). These enzymes play important roles in cancer and other fibrotic diseases, and are actively being pursued as drug targets (Barker et al. 2012). Thus far, strategies to inhibit lysyl oxidase, including using small molecules like β -aminopropionitrile, antibodies or copper chelators have had limited success, possibly because of the redundancy within this family. miR-29b delivery would be a novel way to promote degradation and inhibit translation of *Lox*, *Loxl2* and *Loxl4* mRNAs, thus targeting a family of lysyl oxidase enzymes with a

mechanism of action independent of current strategies. In addition, we have shown that miR-29b inhibits other key pro-metastatic regulators of the tumor microenvironment such as *Angptl4*, *Mmp2*, *Mmp9*, *Vegf* and *Pdgfs*. These observations point to the central importance of miR-29b in regulating the microenvironment and its potential to be used as a therapeutic. Whether miR-29b inhibits metastasis in other cancer types and whether it can be effectively delivered to tumors in animal models await further investigation.

An anti-metastatic or anti-tumorigenic function of miR-29b in fibrotic cancers such as hepatocellular carcinoma

Extensive fibrosis is a hallmark of many cancer types, including hepatocellular carcinoma (HCC). Interestingly, miR-29b levels are decreased in HCC models. It will be interesting to determine if miR-29b inhibits tumorigenesis and metastasis in HCC, and whether miR-29b might be used to inhibit fibrosis in organs such as the liver or the lungs (Cushing et al. 2010; Roderburg et al. 2010). Moreover, it will be interesting to determine whether loss of miR-29b in these tumors increases ECM deposition and fibrosis. Because miR-29b plays an important role in cardiac fibrosis after myocardial infarction (van Rooij et al. 2008) and fibrosis in other model systems, we anticipate that the proposed anti-metastatic role of miR-29b will be confirmed in an HCC model. Moreover, miRNAs can be successfully delivered to the liver (Kota et al. 2009), suggesting that this strategy could be feasibly translated into a therapeutic.

Does miR-29b have a metastasis-suppressive role when over-expressed in fibroblasts?

Many of these tumor microenvironmental genes are highly expressed in cancer-associated fibroblasts (CAFs), which play important pro-angiogenic and pro-fibrotic roles (Pietras et al. 2008; Franco et al. 2009). It would be interesting to determine whether miR-29b over-expression in fibroblasts either alone or in combination with miR-29b over-expression in tumor epithelial cells suppresses metastasis.

How does miR-29b regulate differentiation and EMT?

Given our data on miR-29b's effects of differentiation, however, we were hoping to find that EMT-inducing genes previously described such as *Zeb1*, *Zeb2*, *Twist*, *Snail*, *Slug*, *Gooseoid*, *Foxc2*, *Cdh2* (N-cadherin) would also have miR-29b binding sites in their 3' UTRs. This would have been complementary to studies with the miR-200 family (which targets *Zeb1*). Unfortunately, we did not find this to be the case.

We did find that genes such as *Tgfb2* and *Tgfb3* do have miR-29b binding sites, which might partially explain the effects on EMT. miR-29b decreases *Tgfb2* and *Tgfb3*, while loss of miR-29b increases basal phospho-Smad3 levels. This suggests that miR-29b inhibits TGF β signaling, a key regulator of EMT. However, other genes involved in differentiation might play an important role. For example, there are three miR-29 binding sites in *LRP6*. LRP6 is a co-receptor for Frizzled, which when stimulated by Wnt, activates the canonical Wnt signaling pathway.

As previously reported, Wnt signaling is involved in EMT, and mammary stem cell maintenance and differentiation (DiMeo et al. 2009; Zeng and Nusse 2010). Weinberg and colleagues recently demonstrated that Wnt signaling promotes EMT in breast cancer cells (Scheel et al. 2011). Whether regulation of Wnt signaling is important for miR-29b's effects on EMT/differentiation would be interesting to study.

In addition, previous papers have described other miR-29 targets that might be important in EMT/de-differentiation, including KLF4, DNMT3A and HDAC4 (Fabbri et al. 2007; Winbanks et al. 2011; Fu et al. 2012). Also, we have found that all three *TET* family members (*TET1*, *TET2*, and *TET3*) have multiple miR-29b binding sites in their 3' UTRs, and we have preliminary evidence that they are all down-regulated with miR-29b over-expression, consistent with these being bona fide miR-29b targets. These TET proteins have previously been implicated in the maintenance of hematopoietic stem cells by regulating 5' methyl-cytosine modifications (Ito et al. 2010; Cimmino et al. 2011; Ficz et al. 2011; Koh et al. 2011). However, their roles in EMT and differentiation in breast cancer and normal mammary cells have not been investigated.

What microRNAs mediate branching morphogenesis?

We have shown that miRNAs play an important role in branching morphogenesis. It would be interesting to determine the role of individual miRNAs by adding back single or combinations of miRNAs in *Dgcr8*-null mammary epithelial cells in a system without any noise or redundancy. This

would allow us to identify miRNAs that control different aspects of cell biology such as proliferation, differentiation and polarity. Important candidates that one might start with include those most highly expressed in the mammary gland. We have previously conducted miRNA profiling of mammary epithelial cells, which demonstrates high expression of the miR-200 family, the let-7 family, and the miR-30 family.

What tumor-derived factors affect the distant metastatic niche and what cell types are involved?

Many studies have demonstrated the importance of the tumor microenvironment and the metastatic niche. The identification of tumor-derived factors that prepare the distant soil for colonization, whether they are cytokines, growth factors, exosomes or other molecules, will be important in how we think about and treat metastatic disease. Inhibiting these tumor-derived factors will be important to demonstrate their contributions to metastasis progression and their potential clinical use. More work done in immunocompetent models will be necessary, given the contribution of the immune system to cancer progression. In addition, future studies should avoid using advanced late-stage cell lines to define the “pre-metastatic” niche, and instead utilize more clinically relevant models, especially those that recapitulate the early stages of tumor development.

Appendix A: Zeppo1 is a novel metastasis promoter that represses *E-cadherin* expression and regulates p120-catenin isoform expression and localization

Source: The following appendix was published as a manuscript from: Slorach E, Chou J, and Werb Z. (2011). Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. *Genes Dev* 25, 471-484.

Contributions: For this project, I generated the *Zpo1* shRNA knockdown cell lines, performed the real-time PCR experiments, generated the time-lapse movies of the cell aggregates cultured in 3D matrigel, performed western blotting and co-IP experiments, helped with the in vivo assays and analyzed tumors. I helped write excerpts of the manuscript and participated in the editorial process. Euan Slorach conceived the idea of the project and performed most of the experiments described in the manuscript. Zena Werb initiated and supervised the project.

Zeppo1 is a novel metastasis promoter that represses *E-cadherin* expression and regulates p120-catenin isoform expression and localization

Euan M. Slorach¹, Jonathan Chou^{1,2} and Zena Werb^{1,2,3}

¹Department of Anatomy, and ²Biomedical Sciences Program, University of California, San Francisco CA 94143-0452

³Corresponding author:

Zena Werb, Ph.D.

Department of Anatomy, HSW 1323

University of California

San Francisco, CA 94143-0452 USA

Tel.: (415) 476-4622

Fax: (415) 476-4565

Email zena.werb@ucsf.edu

Keywords: Znf703; breast cancer; metastasis; p120-catenin; E-cadherin; Wnt

ABSTRACT

Amplification of 8p11-12 in human breast cancers is associated with increased proliferation, tumor grade and reduced metastasis-free patient survival. We have identified *Zeppo1* (*FLJ14299/ZNF703*) within this amplicon, as a regulator of cell adhesion, migration and proliferation in mammary epithelial cells. Overexpression of *Zeppo1* reduces cell-cell adhesion and stimulates migration and proliferation. Knockdown of *Zeppo1* induces adhesion and lumen formation. *Zeppo1* regulates transcription, complexing with Groucho and repressing *E-cadherin* expression and Wnt and TGF β reporter expression. *Zeppo1* promotes expression of metastasis-associated p120-catenin isoform 1, and alters p120-catenin localization upon cell contact with the extracellular matrix. Significantly, *Zeppo1* overexpression in a mouse breast cancer model increases lung metastases, while reducing *Zeppo1* expression reduces both tumor size and the number of lung metastases. These results indicate that *Zeppo1* is a key regulator of breast cancer progression.

INTRODUCTION

Chromosomal rearrangements of the short arm of chromosome 8 are relatively common in epithelial cancers (Birnbaum et al. 2003). These rearrangements are often complicated (Pole et al. 2006), and involve loss, amplification and inversion of chromosomal DNA. Amplification of 8p11-12, in particular, is seen frequently in both familial (Melchor et al. 2007) and sporadic (Garcia et al. 2005; Prentice et al. 2005) breast cancers and may occur in as many as 24% of sporadic cases. Amplification in breast tumors is associated with a high proliferative index, a high tumor grade and reduced metastasis-free survival (Gelsi-Boyer et al. 2005), suggesting that this region contains a tumor/metastasis promoting gene or genes. A minimal 8p12 amplicon and the genes that lie within it have been identified (Garcia et al. 2005).

Tumor metastasis is the major cause of poor prognosis in breast cancer. It is a multi-stage process involving changes in tumor cell migration and invasiveness. These properties in turn, depend upon alterations in cell-cell and cell-extracellular matrix (ECM) adhesion that are highly dynamic and regulated processes (Hynes 2002; Gumbiner 2005). E-cadherin is the primary adhesive molecule in epithelial cell-cell adhesion. It interacts with members of the catenin family of proteins to form a protein complex at sites of cell-cell contact known as adherens junctions (AJs). p120-catenin (p120) binds to the juxtamembrane domain of the E-cadherin cytoplasmic tail and stabilizes E-cadherin membrane localization (Ishiyama et al.). β -catenin binds to the cytoplasmic tail of E-cadherin and recruits α -catenin to facilitate the connection between E-cadherin and the

actin cytoskeleton (Nelson 2008). Cadherins and catenins also interact with receptor tyrosine kinases such as the epidermal growth factor receptor (EGFR) at the cell membrane (Qian et al. 2004). This allows AJs to integrate and coordinate growth factor signaling with cell adhesion, migration and gene transcription.

The Wnt and TGF β signaling pathways integrate cell signaling with changes in cell adhesion. The Wnt pathway in particular, is strongly implicated in this process, because it requires β -catenin to mediate extracellular signals from the cell membrane to the nucleus (Clevers 2006). Activation of the canonical Wnt pathway leads to nuclear accumulation of β -catenin and the formation of a protein complex between β -catenin and members of the Lef/Tcf family of transcription factors. This protein complex can activate or repress gene transcription depending on context (MacDonald et al. 2009). Significantly, in mammals, *E-cadherin* expression is repressed through a combination of Wnt and TGF β signaling (Jamora et al. 2003; Medici et al. 2006; Nawshad et al. 2007) in epithelial-to-mesenchymal transition (EMT).

In this study, we sought to identify the gene that accounts for the heightened metastatic properties of the 8p11-12 amplicon. We cloned and characterized *Zeppo1* (*Zpo1*), the mouse ortholog of *ZNF703/FLJ14299*, a gene that lies within the amplicon and is frequently overexpressed in breast cancers.

RESULTS

Identification of Mouse Zeppo1 and Zeppo2

Within the 8p12 amplicon, the gene *FLJ14299/ZNF703* has been repeatedly identified as being amplified and significantly overexpressed in sporadic and familial breast cancers (Garcia et al. 2005; Adelaide et al. 2007; Melchor et al. 2007). Using the *ZNF703* human sequence we identified a mouse *Znf703* ortholog and another closely related family member (*Znf503*) from the NCBI database. We screened a BAC library to clone the mouse genes in their genomic context. Both genes consisted of two exons separated by a short intron. The exons for each gene were amplified by PCR and ligated to generate full length cDNAs.

The mouse and human genes are 96% identical at the amino acid level. Homology searches identified the genes as orthologs of the *Drosophila* zinc finger gene *elbow* (Dorfman et al. 2002). We therefore named *Znf703* and *Znf503* as *Zeppo1* (Zinc finger elbow related proline domain protein 1) and *Zeppo2*, respectively. *Zeppo1* (*Zpo1*) and *Zeppo2* (*Zpo2*) are also orthologs of the zebrafish *Nlz1* (Runko and Sagerstrom 2003) and *Nlz2* (Runko and Sagerstrom 2004) genes, respectively. *Zpo1* and *Zpo2* proteins are 54% identical. Comparison of the protein sequences from *Drosophila*, zebrafish, mouse and human identified several conserved domains, including an amino-terminal Sp-domain, a C₂H₂ zinc finger, and a carboxy-terminal proline-tyrosine rich domain (**Figure A.1A**). In this study we focus on understanding *Zpo1*.

Zpo1 is Expressed in the Mammary Epithelium

We first asked if *Zpo1* is expressed in the normal mouse mammary gland. *Zpo1* was expressed in the developing mammary placodes from E11.5 onwards (**Figure A.1B**). Interestingly, *Zpo1* was also strongly expressed in the developing intestinal epithelium (**Figure A.1B**), a tissue in which the 8p11-12 amplification is also associated with tumor formation (Nakao et al. 2004; Pole et al. 2006). We also found *Zpo1* in the adult mouse and human mammary epithelium (**Figure A.1C** and data not shown).

Using a rabbit polyclonal antibody directed against a peptide in the C-terminal region of the *Zpo1* protein, we observed that endogenous *Zpo1* protein was localized predominantly in the nucleus in non-tumorigenic EpH4.9 mouse mammary epithelial cells (MECs), despite the lack of an identifiable nuclear localization signal, and was also present in the cytoplasm (**Figure A.1D**). We obtained a similar result using an anti-V5 tag antibody in cells infected with a carboxy-terminal V5-tagged *Zpo1* (*Zpo1*-V5) lentiviral expression construct (**Figure A.1D**).

Zpo1 Functions as a Transcriptional Repressor

We then generated a plasmid expressing full-length *Zpo1* fused to the GAL4 DNA binding domain (*Zpo1*-GAL4) and determined its effect on the expression of a GAL4-luciferase reporter construct. Increasing levels of *Zpo1*-GAL4 repressed transcription of the reporter gene (**Figure A.1E**). Transcriptional repression is frequently achieved through recruitment of histone deacetylase (HDAC) to gene

promoters (Yang and Seto 2008). Trichostatin A (TSA), a specific inhibitor of HDAC function, partially inhibited Zpo1-mediated repression (**Figure A.1F**).

Drosophila elbow and zebrafish Nlz1 proteins interact *in vitro* with the transcriptional co-repressor Groucho (Dorfman et al. 2002; Runko and Sagerstrom 2003). There are at least 5 members of the Groucho-related gene (Grg) family in mice (Gasperowicz and Otto 2005). We observed that *Grg4* was expressed in the mammary epithelium at the same developmental time-points as *Zpo1* (**Figure A.1G**). Using embryos from *Grg4* gene-trap mice expressing β -geo under the control of the endogenous *Grg4* promoter, we detected expression in mammary placodes from E11.5 (**Figure A.1H**). Immunoprecipitation of lysates from cells expressing V5-tagged Zpo1 and FLAG-tagged Grg4 with a V5 antibody followed by western blotting with an anti-FLAG antibody, demonstrated co-immunoprecipitation of Grg4 with Zpo1, but not with V5-tagged lacZ (**Figure A.1I**). This interaction suggests that Grg4 may affect Zpo1-mediated transcriptional repression. To test this directly we co-expressed increasing amounts of Grg4 with Zpo1 in luciferase transcription assays. Indeed, increasing levels of Grg4 led to enhanced repression of the luciferase reporter (**Figure A.1J**).

Overexpression of Zpo1 in Mammary Epithelial Cells Reduces Cell-cell Adhesion and Increases Cell Invasion

Amplification of 8p12 is associated with metastasis, a process that involves changes in cell adhesion and migration. Therefore, we generated a polyclonal

EpH4.9 cell line overexpressing Zpo1 (**Supplemental Figure A.S1**) at levels comparable to those seen in human tumor samples (Garcia et al., 2005; Haverty et al., 2008). We then assessed the effect of *Zpo1* overexpression on EpH4.9 cell migration using a wound-healing assay. *Zpo1* overexpressing cells not only migrated a greater distance than control cells (EpH4.9-pEiZ) infected with empty vector (**Figure A.2A**), but also migrated individually with reduced cell-cell adhesion at the wound edge (**Supplemental Movies A.1 and A.2**). We found that cells overexpressing *Zpo1* were significantly more invasive through a cell culture insert coated with Matrigel than control cells (**Figure A.2B**).

Control and *Zpo1* overexpressing cells showed similar levels of E-cadherin and α -, β - and p120-catenin protein, the major components of AJ complexes (**Figure A.2C**). In 2D on tissue culture plastic, all four AJ proteins were membrane localized (**Supplemental Figure A.S2**). When we repeated these experiments with cells cultured in 3D within Matrigel, we observed that EpH4.9-Zpo1 cells formed numerous large filopodia/invadopodia-like structures at the periphery of the aggregates and migrated out from the body of the aggregates to form long chains of single cells (**Figure A.2D**). The cells within the chains showed minimal or a complete absence of cell-cell adhesion. Time-lapse imaging revealed the rapid formation and retraction of membrane protrusions (**Supplemental Movies A.3 and A.4**). In contrast, EpH4.9-pEiZ cells formed tightly adherent aggregates (**Figure A.2D**).

AJ proteins in cells within the main body of both EpH4.9-pEiZ and EpH4.9-Zpo1 aggregates in 3D were localized along the cell membrane (**Figure A.2E**). In

migratory cells of EpH4.9-Zpo1 aggregates, AJ proteins were present at significantly reduced levels, and demonstrated cytoplasmic localization. Similarly, actin stained diffusely throughout the cytoplasm in these cells (**Figure A.2E**), unlike the cortical localization in cells within the aggregate bodies. These results suggest that *Zpo1* overexpression inhibits or disrupts AJ formation at sites of cell interactions with the microenvironment, consequently reducing cell-cell adhesion. As a control we knocked down *Zpo1* expression in EpH4.9-pEiZ and EpH4.9-Zpo1 cells using a *Zpo1* specific shRNA. Knockdown of *Zpo1* expression was confirmed by qPCR and Western blot (**Figure A.2F**). Reduced *Zpo1* expression in both EpH4.9-pEiZ and EpH4.9-Zpo1 cells in 3D culture induced the formation of tightly adherent aggregates containing small lumens and an absence of migratory cell chains (**Figure A.2G** and **A.2H**). This confirms that loss of cell adhesion and increased migration was due to *Zpo1* overexpression. These results demonstrate that *Zpo1* levels directly affect not only epithelial cell-cell adhesion, but also cell polarization and multicellular organization.

Overexpression of Zpo1 in Mammary Epithelial Cells Increases Cell Proliferation in 3D Culture

Increased *ZPO1* expression in human breast cancers is associated with an increased proliferative index. In 2D culture there was no difference in cell proliferation between *Zpo1* overexpressing and control cells (**Figure A.3A**). In contrast, aggregates from *Zpo1* overexpressing cells cultured in 3D were significantly larger compared to control cells (**Figure A.2D**). When we assessed

the number of cells in the aggregates that were positive for phosphorylated histone H3, we observed that about 1% of EpH4.9-pEiZ cells were positive, while over 3% of EpH4.9-Zpo1 cells were positive ($p < 0.001$; **Figure A.3B**). This confirms that cell proliferation was increased in *Zpo1* overexpressing cells in 3D.

Zpo1 Represses *E-cadherin* Gene Expression

The reduction in cell adhesion caused by *Zpo1* overexpression in 3D culture suggested that *Zpo1* may regulate the expression of genes involved in cell-cell adhesion. *Zpo1* overexpression had no effect on α -, β - and *p120-catenin* gene expression levels in EpH4.9 cells, but reduced *E-cadherin* mRNA by approximately 45% (**Figure A.3C**). To determine if *Zpo1* repressed transcription, we cloned a 570 bp fragment of the human *E-cadherin* promoter immediately upstream of the ATG transcription start site, 5' to a luciferase reporter gene. Transfection of the reporter construct into 3T3 cells resulted in significant luciferase activity that was reduced by co-transfection of a *Zpo1* expression plasmid (**Figure A.3D**).

E-cadherin gene expression is reported to be negatively regulated by the Wnt pathway (Jamora et al. 2003; Yook et al. 2005). Surprisingly, we found that co-transfection of a β -*catenin* expression plasmid significantly increased expression from the *E-cadherin* reporter construct (**Figure A.3E**). However, co-transfection of *Zpo1* with the β -*catenin* expression plasmid abrogated transcriptional activation (**Figure A.3E**). We then repeated these experiments using the TOPFLASH reporter plasmid that contains multiple copies of a Lef/Tcf

binding site upstream of a luciferase reporter gene. Increased β -catenin activity significantly increased luciferase activity as expected; however, co-transfection of *Zpo1* again reduced β -catenin mediated transcriptional activation (**Figure A.3F**).

Since both Wnt and TGF β signaling negatively regulate *E-cadherin* expression, we used a luciferase reporter plasmid containing Smad binding elements to determine if *Zpo1* also regulates TGF β -responsive transcription. Co-expression of *Zpo1* was sufficient to repress transcription from this reporter by ~50% in both the absence and presence of exogenous TGF β (**Figure A.3G**).

Zpo1 Overexpression Induces an Epithelial-to-mesenchymal Transition

Zpo1 overexpressing cells at the invasive edge of 3D aggregates clearly differed from those in the center, as evidenced by reduced E-cadherin expression, remodeling of the cytoskeleton and increased cell migration. These are hallmarks of EMT and a more progenitor or stem cell-like phenotype (Mani et al. 2008). To validate this observation we evaluated other EMT markers. *Zpo1* overexpressing cells within the body of the aggregates were negative for vimentin, a marker of mesenchymal cells, as were control cells; however, *Zpo1* overexpressing cells at the interface with the microenvironment were clearly vimentin positive (**Figure A.4A**). All EpH4.9-*Zpo1* cells were positive for Snail and N-cadherin, while control cells were negative. Interestingly, N-cadherin staining was mainly plasma membrane localized in the body of the *Zpo1* overexpressing aggregates, but in the cytoplasm in peripheral cells (**Figure A.4A**). Both control and *Zpo1* overexpressing cells were positive for cytokeratin (**Figure A.4A**).

Zpo1 Overexpression Promotes a Switch to a Pro-migratory p120-catenin Isoform in 3D Culture

To gain mechanistic insight into how Zpo1 mediates the differences between 2D and 3D culture conditions, we next investigated another pro-migratory molecule. p120 exists as multiple isoforms due to alternative splicing (Mo and Reynolds 1996; Keirsebilck et al. 1998) and 4 alternative transcriptional start sites. The most commonly expressed isoforms are 1 and 3. Using an antibody recognizing all p120 isoforms, we observed that the banding pattern from EpH4.9-Zpo1 cells differed from controls (**Figure A.4B**). An antibody that detects isoform 1 but not isoform 3, showed that EpH4.9-Zpo1 cells in 3D culture predominantly express isoform 1, while controls predominantly express isoform 3 (**Figure A.4C**). In contrast, both isoforms 1 and 3 were expressed in 2D culture (**Figure A.4C**), with isoform 3 as the predominant isoform in both control and *Zpo1* overexpressing cells.

If the relative levels of p120 isoforms affect the adhesive and/or migratory phenotype of MECs, then increasing the level of p120 isoform 3 in EpH4.9-Zpo1 cells should reduce the invasive phenotype in 3D culture. Accordingly, we developed stable cell lines from EpH4.9-Zpo1 cells carrying either a p120 isoform 3 expression construct or the empty expression vector as a control. Cell aggregates generated from EpH4.9-Zpo1 cells overexpressing isoform 3 and cultured in 3D Matrigel demonstrated reduced cell migration compared to control cells (**Figure A.4D**), consistent with our hypothesis.

Zpo1 Regulates ECM-dependent Intracellular p120 Isoform Distribution

A striking phenotype of the *Zpo1* overexpressing cells cultured in 3D was that redistribution of AJ proteins appeared to be restricted to cells on the periphery of the aggregates that are in contact with the ECM (**Figure A.2E**). This suggested that redistribution might be a result of cell-ECM interactions. To test this we cultured EpH4.9-pEiZ and EpH4.9-*Zpo1* cells on top of a thin 2D layer of Matrigel and determined AJ protein localization. EpH4.9-pEiZ cells formed compact epithelial colonies, whereas EpH4.9-*Zpo1* cells were more scattered with reduced cell-cell contact (**Figure A.4E**). All AJ proteins, with the exception of p120, were membrane localized in control and *Zpo1* overexpressing cells (data not shown). EpH4.9-*Zpo1* cells showed a predominantly nuclear localization of p120 with only weak cytoplasmic and membrane staining using a pan-isoform antibody (**Figure A.4E**); however, with the isoform 1/2-specific antibody, we saw protein only at the cell membrane and in the cytoplasm, but not in the nucleus (**Figure A.4E**). These data suggest that isoform 3 specifically localizes to the nucleus.

Based on our results, we propose that cell-ECM interactions mediate the effects of *Zpo1* overexpression. To test this hypothesis we used a function-inhibiting antibody directed against laminin 332 (also called laminin 5) to perturb cell-ECM interactions. Laminin 332 is not present in Matrigel, but is expressed endogenously by EpH4 cells (Maschler et al. 2005). Laminin 332 negatively regulates the number of actin-rich invadopodial protrusions (Liu et al.) that

facilitate ECM degradation and invasion (Albiges-Rizo et al. 2009). Addition of the anti-laminin 332 antibody to the growth medium of 3D cultures had no visible effect on control EpH4.9-pEiZ cells (**Figure A.5A**); however, in EpH4.9-Zpo1 cell cultures, migrating cells were more elongated, possessed longer membrane protrusions and invaded further into the Matrigel compared to cells grown in control medium (**Figure A.5A**). These results suggest that loss of cell-laminin 332 interactions promotes epithelial cell invasion in *Zpo1* overexpressing cells.

Rac1 Activity and Inhibition of RhoA is Required for the Zpo1 Overexpression Phenotype

p120 can promote cell migration via interactions with Rho GTPases, increasing Rac1 activity and inhibiting RhoA. Isoform 1 specifically promotes increased cell migration compared to other p120 isoforms due to its increased affinity for, and inhibition of, RhoA (Yanagisawa et al. 2008). Addition of Rac1 inhibitor significantly inhibited EpH4.9-Zpo1 cell migration and invasion, but had no obvious effect on control cells (**Figure A.5B**). We then inhibited ROCK, a downstream mediator of RhoA that becomes activated upon RhoA activity. Addition of a ROCK inhibitor enhanced migration and invasion of *Zpo1* overexpressing cells, but had no obvious effect on control cells (**Figure A.5B**).

Epithelial cell-ECM interactions and binding of p120 to RhoA (Castano et al. 2007), are both mediated through activity of the Src family of kinases (SFKs). Therefore, inhibition of SFK activity would be expected to alter the *Zpo1*-mediated increase in MEC migration and invasion. EpH4.9-pEiZ cells cultured in

3D in the presence of the SFK inhibitor SU6656 (**Figure A.5C**) or PP1 (data not shown) showed no obvious effect. However, EpH4.9-Zpo1 cells in the presence of SFK inhibitors had a rounder shape, formed significantly fewer membrane protrusions, and for the most part remained attached to the aggregate (**Figure A.5C**). Sectioning of aggregates from both EpH4.9-pEiZ and EpH4.9-Zpo1 cells showed the presence of small lumens (**Figure A.5D**).

Zpo1 Overexpression Promotes Tumor Metastasis.

To determine directly if *Zpo1* overexpression is tumor and/or metastasis promoting, we transplanted EpH4.9-pEiZ or EpH4.9-Zpo1 cells into the mammary fat pad of syngeneic mice. Parental EpH4.9 cells are non-tumorigenic; however, *Zpo1* overexpression alone was not sufficient to induce tumor formation, even 12 months after transplantation (data not shown).

We then asked if *Zpo1* overexpression affected tumor metastasis. The mammary epithelial cell line 4T1 is both tumorigenic and metastatic to the lung, liver, bone and brain, when transplanted into the mammary fat pad of immunocompetent mice (Aslakson and Miller 1992). We generated Zs-green-labeled polyclonal 4T1 cells, overexpressing *Zpo1* (4T1-Zpo1) and control 4T1 cells infected with empty vector (4T1-pEiZ), and transplanted them in to the mammary fat pads of syngeneic BALB/c mice. After 3 weeks, the weight of tumors from mice transplanted with *Zpo1* overexpressing cells was slightly increased compared to controls (**Figure A.6A**). Significantly, the number of metastatic lung nodules was dramatically increased in mice transplanted with

Zpo1 overexpressing cells compared to controls (**Figures A.6B** and **A.6C**). Furthermore, the 4T1-*Zpo1* primary tumors clearly showed a reduction in E-cadherin protein levels compared to control tumors (**Figure A.6D**). We next determined if reducing *Zpo1* expression in 4T1 cells would affect tumor progression. We generated polyclonal 4T1 cell lines stably expressing either a control scrambled shRNA or a *Zpo1* specific shRNA that reduced *Zpo1* expression by ~60% (data not shown), and repeated the transplant experiments. After 3 weeks, both the size of the primary tumors and the number of lung metastases were significantly reduced in *Zpo1* knockdown cells compared to controls (**Figure A.6E**). These results suggest that *Zpo1* is the driver gene for increased tumor metastases in patients with the 8p12 amplicon.

DISCUSSION

In this study we have identified *Zpo1/Znf703*, a critical gene in the 8p12 breast cancer-associated amplicon, as one of the earliest genes expressed during mammary gland development, and have elucidated its functions as a transcriptional repressor and as a regulator of cell motility and invasion. Together with the observation that *Zpo1* increases metastasis, we propose that *Zpo1* is a metastasis promoter that is responsible for poor breast cancer outcome.

Zeppo1 and Transcriptional Regulation

Zpo1 is a member of the NET (Noc/Nlz, Elbow, Tlp-1) family of proteins (Nakamura et al. 2004) that function in embryonic development in zebrafish (Hoyle et al. 2004; Brown et al. 2009), *Drosophila* (Dorfman et al. 2002) and *C. elegans* (Zhao et al. 2002). We show here that *Zpo1* functions as a transcriptional repressor. While some of *Zpo1*'s repressor function requires HDAC activity, it also demonstrates HDAC-independent repressor activity, possibly as a result of its interactions with other co-repressors such as Grg4. These results are consistent with studies showing that, while Nlz1 activity is associated with HDACs *in vitro* (Runko and Sagerstrom 2004), HDAC activity accounts for only part of the repression (Nakamura et al. 2008).

We have shown that increased *Zpo1* expression represses endogenous mouse *E-cadherin* gene transcription and inhibits transcription from a human *E-cadherin* promoter fragment. Interestingly, *Zpo1* also inhibited β -catenin-mediated transcriptional activation of both the human *E-cadherin* and the

TOPFLASH promoters. Since β -catenin competes with Grg proteins for binding to Lef/Tcf transcription factors (Daniels and Weis 2005), one possible mechanism of Zpo1 function may be to promote binding of Grg proteins to Lef/Tcf. An alternative possibility is that Zpo1 functions to convert β -catenin from a transcriptional activator into a repressor. β -catenin can function as part of a repression complex, either in conjunction with members of the Lef/Tcf family (Jamora et al. 2003; Theisen et al. 2007) or other transcription factors (Olson et al. 2006), and β -catenin mediated repression requires binding of co-repressors and recruitment of HDAC activity (Olson et al. 2006; Iguchi et al. 2007). Nonetheless, Zpo1-mediated repression of *E-cadherin* transcription is significant since E-cadherin is a tumor and metastasis suppressor, and reduced gene expression, protein levels or changes in protein localization are found in the majority of metastatic carcinomas including breast cancers (Cowin et al. 2005). Repression of *E-cadherin* gene expression is also regulated by transcription factors such as Snail1/2, ZEB1/2 and E12/47, and like Zpo1, they are frequently upregulated during tumorigenesis (Peinado et al. 2007).

Interestingly, while Zpo1 was able to repress *E-cadherin* gene transcription, E-cadherin protein levels were not reduced in 2D culture (**Figure A.2C**), and only minimally reduced in 3D culture. This is likely a consequence of the high stability and low turnover of E-cadherin protein under the conditions of stable cell-cell adhesion that we see in 2D. In contrast, E-cadherin is rapidly degraded or recycled upon loss of cell-cell contact (Daniel and Reynolds 1999); however, since this phenotype is restricted to cells on the outer edge of the 3D

aggregates, overall protein levels are not greatly affected.

Zpo1 and Epithelial-to-mesenchymal Transition

Zpo1 overexpression promoted EMT and the expression of the EMT markers vimentin, Slug and N-cadherin. Interestingly, Slug and N-cadherin expression was induced in all cells, while vimentin expression was restricted to migratory cells at the interface with the microenvironment. This suggests that expression of these genes is differentially regulated, and that cell-ECM interactions specifically regulate *Zpo1*-mediated vimentin expression.

Loss of E-cadherin in EMT not only reduces cell-cell adhesion, but also induces a loss of cell polarity and acquisition of a more progenitor-like state. EpH4.9 cells cultured in 3D Matrigel form poorly polarized, adherent cell aggregates. We showed that knockdown of endogenous *Zpo1* expression in EpH4.9 cell aggregates induced the formation of small lumens lined with cells rich in apical actin, suggesting polarization, while overexpression of *Zpo1* induced the cytoplasmic localization of AJ proteins. Interestingly, we obtained a similar phenotype when we treated cell aggregates with SFK inhibitors. Src activation has long been known as a tumor promoter (Frame 2002) and a regulator of cell polarity (Timpson et al. 2001; Yamana et al. 2006; Grande-Garcia et al. 2007; Wang et al. 2009). The inhibition of the *Zpo1* phenotype by SFK inhibitors and the formation of lumens by either *Zpo1* knockdown or SFK inhibition, suggests that *Zpo1* may function in part, through SFK activity.

p120 in Cell Migration and Cancer

p120 is a multi-functional protein that localizes to the cell membrane, cytoplasm and nucleus. At the cell membrane, p120 binds to E-cadherin, promoting clustering (Yap et al. 1998) and stabilization of E-cadherin by inhibiting protein turnover (Davis et al. 2003). In the cytoplasm, p120 interacts with members of the Rho family of GTPases to regulate the actin cytoskeleton and consequently several actin-dependent cellular processes including cell shape, adhesion, migration and polarity (Anastasiadis 2007). Nuclear p120 function is less well understood, but it can interact with the transcription factor Kaiso (Daniel and Reynolds 1999) to inhibit alleviate Kaiso-mediated transcriptional repression (Daniel et al. 2002). The combined regulation of both cell-cell adhesion and actin dynamics provides an important role for p120 in regulating the motile state of cells.

p120 may function as both a tumor suppressor and a metastasis promoter (Reynolds and Roczniak-Ferguson 2004), and its expression is altered in the majority of human cancers, either at the level of gene expression or protein localization (van Hengel and van Roy 2007). In one study, 88% of lobular human breast carcinomas had cytoplasmic localization of p120 with a corresponding absence of E-cadherin (Sarrio et al. 2004).

p120 exists as multiple isoforms due to alternative splicing of a single gene (Reynolds and Roczniak-Ferguson 2004). The most commonly expressed isoforms are isoforms 1 and 3. Isoform 1 is predominantly expressed in motile cells such as fibroblasts and in epithelial tumors, while isoform 3 is the

predominant isoform in sessile epithelial cells. p120 expression is an independent prognosticator of breast cancer survival, and isoform 1 expression predicts metastatic disease (Talvinen et al.). Isoform 1 expression and metastasis are also significantly correlated in both lung (Miao et al. 2009) and renal carcinoma (Yanagisawa et al. 2008).

Our studies show that EpH4.9 MECs preferentially express p120 isoform 3 when cultured on 2D plastic. Culturing EpH4.9 cells within a laminin-rich 3D matrix, however, further reduces isoform 1 expression. Overexpression of *Zpo1* in these cells had no effect on p120 isoform expression in normal 2D culture; however, when cultured within an ECM, increased *Zpo1* expression induced an isoform switch such that isoform 1 was preferentially expressed. Interestingly, transcriptional repression of E-cadherin in combination with isoform switching of p120 from isoform 3 to isoform 1 is seen in MDCK cells overexpressing the EMT-inducing transcriptional repressors Snail, Slug or E47 (Ohkubo and Ozawa 2004; Sarrío et al. 2004).

Zpo1 overexpression also promoted re-localization of p120 protein away from the cell membrane that was also dependent upon cell-ECM contact. In 2D Matrigel culture, *Zpo1* overexpression induced nuclear localization of isoform 3, while isoform 1 was restricted to the cell membrane and cytoplasm. In 3D culture, both isoforms were localized to the cell membrane in cells in the body of the aggregates, away from the ECM, but were localized to the cytoplasm in migratory cells in contact with the ECM. These results provide further evidence that *Zpo1* function is regulated by cell-ECM interactions.

A Model for Zpo1-mediated Increase in Tumor Metastasis

The metastatic process requires cells to alter their normal interactions with neighboring cells and the extracellular environment (Gupta and Massague 2006). Understanding how these processes are regulated is of major importance for the development of therapies that might interfere with metastasis formation. Based on the data presented in this paper we propose a model for Zpo1 function (**Figure A.7**) to explain how overexpression can lead to increased tumor metastasis. The 8p12 amplification and the overexpression of multiple genes within the amplicon results in increased *Zpo1* gene expression. Increased levels of Zpo1 protein repress *E-cadherin* expression, resulting in a decrease in cell-cell adhesion and re-localization of p120 from the cell membrane to the cytoplasm. Concomitantly, increased *Zpo1* expression induces a switch in p120 isoforms from isoform 3 to isoform 1 that leads to increased binding of p120 to RhoA, decreasing RhoA activity. *Zpo1* expression also increases cell proliferation through an as yet unknown mechanism. In combination, these events reduce tumor cell adhesion and cell polarity while increasing tumor cell migration and proliferation, thus promoting a higher tumor grade and an increase in tumor metastases.

Role of Other Genes Within the 8p11-12 Amplicon

Overexpression of *Zpo1* in non-tumorigenic EpH4.9 cells was insufficient to induce tumor formation, suggesting that *Zpo1* overexpression by itself may not

be a tumor-initiating event. It is possible that an unrelated genetic event is responsible for tumor formation and that amplification of 8p12 occurs later in the process. Alternatively, since 8p12 amplification affects the expression of multiple genes within the locus, altered expression of more than one gene may be responsible for tumor formation and metastatic progression. Several genes within this locus have been studied with regard to their oncogenic capabilities; however, none have been convincingly identified as classical oncogenes. One such gene, *RCP* (also known as *RAB11FIP1*) was recently identified as a candidate oncogene associated with breast cancer (Miao et al. 2009); however, overexpression of *RCP* in tumorigenic MECs has no effect on tumor cell growth, invasion or metastasis *in vivo*. *RCP* overexpression in MCF10A cells did result in soft-agar colony formation, reduced requirement for growth factors and altered cell migration. Interestingly, similar results were obtained when three other genes located at 8p11-12 (*LSM1*, *BAG4* and *C8orf4*) are overexpressed individually or in combination in MCF10A cells (Yang et al. 2006). It is therefore possible that amplification and overexpression of some combination of these genes is required for the full 8p11-12 amplification phenotype seen in human breast cancer. However, a study that profiled 1172 primary breast cancers also has identified ZNF703 as the sole 8p12 candidate responsible for poor prognosis in ER+ tumors (Holland et al. 2011; Sircoulomb et al. 2011).

In conclusion, we show that *Zpo1* overexpression plays an important role in regulating mammary epithelial cell adhesion, migration and polarity, and can significantly increase the number of mammary tumor metastases. Our data

strongly implicate *Zpo1* overexpression as a metastasis driver in breast carcinogenesis.

MATERIALS AND METHODS

RNA in situ hybridization of mouse tissues

Gene expression on paraffin sections of mouse embryos by RNA in situ hybridization was carried out as described previously (Lum et al. 2007). Wholemount embryo *in situ* hybridizations were performed on CD1 mouse embryos as previously described (Cross et al. 1995). *Zpo1* riboprobe was generated from a 635bp *BamHI* fragment from exon2. *Grg4* riboprobe was generated from a PCR fragment amplified using primers 5'-TCCAACCTCCACGCATCAAGG and 5'-TGCTATGAGGAGGAGTCCAGTTTTG.

Cell Culture

Mouse EpH4.9 cells are a subclone derived from EpH4 cells. EpH4.9 cells grown on tissue culture plastic were cultured in DME-H21 medium supplemented with 5% (v/v) fetal bovine serum (FBS), Insulin (5 µg/ml), 100 IU penicillin, and 100 µg/ml streptomycin. EpH4.9 cells to be cultured in 3D Matrigel were maintained in low adhesion plates overnight to form aggregates. The aggregates were separated from remaining single cells by gravity, resuspended in ice-cold Matrigel and plated in 25 µl volumes. The Matrigel/cell mix was allowed to solidify at 37°C for 1hour and cultured in serum-free DMEM:F12 medium supplemented with 1X ITS (Invitrogen), 50 ng/ml EGF (Invitrogen), 100 IU penicillin and 100 µg/ml streptomycin. SFK inhibitor SU6656 (Sigma-Aldrich) was added to EpH4.9 growth medium at 10 µM. Mouse 3T3 and 4T1 cells and human 293T cells were cultured in DME H-21 medium supplemented with 10% (v/v) FBS),

100 IU penicillin and 100 µg/ml streptomycin.

Cell Proliferation

Cell proliferation in 2D cultures was quantified using a colorimetric assay (CellTiter 96, Promega) to determine the number of viable cells, according to the manufacturer's instructions. Proliferation of EpH4.9 cells cultured in 3D for 24 hours was determined by counting the number of cells positive for phospho-histone H3. This was achieved by staining 50 µm frozen sections of cultured cells using a phospho-histone H3 antibody (Cell Signaling Technology) and taking serial stack images of individual aggregates by confocal microscopy. The total number of cells visualized by DAPI staining and the number of phospho-histone H3 positive cells in the 5th and 10th image of each stack was counted to determine the percent of proliferating cells.

Invasion Assay

Transwell cell migration assays were performed using the Millipore Cell Invasion Assay kit following manufacturer's instructions. Briefly, 3×10^5 cells were plated on to ECM coated cell culture inserts and incubated for 48 hours. Cells that had invaded through the ECM to the lower surface of the insert were stained and quantified by colorimetric analysis.

Wound Healing Assay

EpH4.9 cells were grown to confluence in a 24-well plate, and wounded using a

200 μ l sterile pipette tip. Movies of wound healing were generated using an inverted microscope over a 24-hour time period. The distance travelled by cells at the migrating edge after 15 hours, was calculated by measuring 3 separate points within each field of view, from at least 3 separate experiments.

Antibodies and Immunofluorescence

Cells cultured in 2D were fixed in 4% paraformaldehyde, incubated with primary antibody followed by an appropriate secondary antibody conjugated to Alexa Fluor 568 or 647 (Invitrogen). The actin cytoskeleton was stained with phalloidin (Invitrogen) and cell nuclei with DAPI (Invitrogen). Immunofluorescence images were obtained using confocal microscopy and Imaris imaging software (Bitplane Scientific Software). Primary antibodies used were specific for p120-catenin (all isoforms), V5 epitope and GAL4 DNA binding domain (Invitrogen); FLAG tag, α -catenin, p120-catenin (isoforms 1 and 2) and Vimentin (Sigma-Aldrich); N-cadherin, Snail and phospho-histone H3 (Cell Signaling Technology); E-cadherin (BD Biosciences), and β -catenin (Abcam).

Lentiviral Production and shRNA Knockdown

Lentiviral production was carried out using calcium phosphate-mediated transfection of HEK 293T cells as previously described (Welm et al. 2008). Expression Arrest GIPZ lentiviral shRNA constructs were purchased from Open Biosystems. Cells were transduced with lentivirus according to manufacturer's instructions and cells were selected in puromycin for at least 5 days.

Quantitative Real-time PCR

Total RNA was isolated from cells using the RNeasy Mini Kit (Qiagen). cDNA was synthesized from 1 µg of total RNA using the Superscript III Reverse Transcriptase First Strand synthesis kit (Invitrogen). Quantitative PCR was performed using FastStart Universal SYBR Green master mix (Roche Applied Science) in an Eppendorf Mastercycler realplex machine, and samples were run in quadruplicate. Ct values were normalized to actin, and relative expression was calculated using the 2^{-DDCt} method. Statistical analysis was performed using Student's t test and $p < 0.05$ was considered to be significant. All statistical values indicate means \pm standard deviation. Primer sequences are detailed below:

Primer sequences for qPCR

<u>Primer Name</u>	<u>Sequence (5' to 3')</u>
Actin-F	CACAGCTTCTTTGCAGCTCCTT
Actin-R	CGTCATCCATGGCGAACTG
GAPDH-F	AGGTCCGGTGTGAACGGATTTG
GAPDH-R	TGTAGACCATGTAGTTGAGGTCA
Ecadherin-F1	CCAACAGGGACAAAGAAACAAAGG
Ecadherin-R1	GATGACACGGCATGAGAATAGAGG
Ecadherin-F2	TCCAGGAACCTCCGTGATGAAG
Ecadherin-R2	GGCGATGGCAGCGTTGTAG
Zpo1-F	CTCCACCTCATCAAACCTCGTCTTC
Zpo1-R	GGCTTCCTGCTCTTTCTTGTCC
α -catenin-F	GCCAAGCAGATGTGCATGATC
α -catenin-R	CAGAGGTGTTTTTGTGAGTGGACCTT
β -catenin-F	GGGTGGCATAGAGGCTCTTGT
β -catenin-R	GCTCAGTGATGTCTTCCCTGTCA
p120 catenin-F	AGCTTGTGGAGAATTGTGTTTGC

p120 catenin-R	TGCCTGTGGGATTTACGAT
----------------	---------------------

Transcription Assays

3T3 cells were seeded at 5×10^4 cells/well in 48-well plates, and transfected with 400 ng of pFR-Luc luciferase reporter construct (Stratagene), 5ng pRL-TK (Promega) internal control plasmid and the appropriate test plasmids. Cell lysates were obtained 48 hours post-transfection and luciferase reporter activity was determined using the Promega Dual-Luciferase Reporter assay system following the manufacturer's protocol. Firefly luciferase activity was normalized to the Renilla luciferase internal control and expressed relative to control plasmid. The data presented are pooled from 3 replicate experiments in which each assay condition was performed in triplicate. The E-cadherin promoter luciferase reporter construct was generated by amplifying human genomic DNA using the primers 5'-TTGCTAGCAAAGAAAAAAAAAATTAGCCTGGCGTGG-3' and 5'-AAACTCGAGCGGGCTGGAGCGGGC-3'. Amplified DNA was digested with *NheI/XhoI* and cloned in to the pGL3-basic reporter plasmid (Promega). The β -catenin expression plasmid was a kind gift from Dr. Thaddeus Allen (UCSF, San Francisco, CA). Trichostatin A (Calbiochem) was added to 3T3 growth medium at various concentrations as described in the results section.

Tumorigenesis and Metastasis Assays

Cells were harvested, counted and resuspended in PBS. 1×10^6 cells were injected in to the fat pad of the inguinal (#4) mammary gland of syngeneic 5-week old BALB/c mice. Primary tumors were dissected, weighed and measured

after 3 weeks. The number of lung metastases were determined by counting the number of Zs-Green positive areas visible on the lung surface. All animal protocols were reviewed and approved by the UCSF Institutional Animal Care and Use Committee.

Plasmid Construction and Transfection

Zpo1-V5/His and GAL4-Zpo1 were generated by cloning the Zpo1 cDNA in to the pUB6-V5/His (Invitrogen) and pFA-CMV (Stratagene) plasmids respectively. lacZ-V5/His was purchased from Invitrogen. pRc-CMV and HA-tagged p120-1A and 3A expression plasmids were a kind gift from Dr. A. Reynolds (Vanderbilt University, Nashville, TN). Plasmids were transfected in to EpH4.9 or 3T3 cells using FuGene transfection reagent (Roche) following the manufacturer's protocol. pEiZ/HIV-ZsGreen was previously described (Welm et al. 2008). pEiZ-Zpo1-V5 was generated by subcloning Zpo1-V5 in to pEiZ.

Western Blotting and Co-immunoprecipitation

Cells were lysed in RIPA buffer plus protease inhibitors (Roche). To enable equal loading of protein for SDS-PAGE, protein concentration was measured directly from cells cultured in 2D (BCA Protein Assay Kit, Thermo Scientific). Equivalent amounts of cellular protein from 3D cultured cells were obtained by normalizing lysates to GAPDH protein levels using an anti-GAPDH antibody (Cell Signaling Technology). Lysates were subjected to SDS-PAGE and transferred to PVDF membrane. Blots were blocked in 5% skimmed milk and incubated with primary

antibody. Bands were visualized using ECL Detection Reagents (GE Healthcare). For co-immunoprecipitation, 3T3 cells were transfected with FLAG-Grg4 plus either LacZ-V5 or Zpo1-V5 expression plasmids. 48 hours later cells were lysed in co-ip buffer (Tris-Cl, pH8.0 (20 mM), NaCl (25 mM), MgCl₂ (1.5 mM), EGTA (1 mM), Triton X-100 (1%), glycerol (10%), DTT (1 mM) plus protease inhibitors). Anti-V5 agarose (Novus Biologicals) was added to the lysates and samples were incubated overnight at 4°C. Immunoprecipitates were washed and analyzed by Western blot using an anti-FLAG antibody (Sigma-Aldrich).

ACKNOWLEDGEMENTS

We thank Heidi Savage and Ying Yu for excellent technical assistance. We are grateful to Dr. Monique Aumailley, Center for Biochemistry, Cologne, Germany, for the gift of the rabbit anti-laminin332 antibody. This study was supported by funds from the UCSF Program for Breakthrough Biomedical Research, the National Cancer Institute (CA057621, CA129523 and CA129523S1), the Human Frontiers Science Program (RG 0051/1999-M) and a Stand up to Cancer-American Association for Cancer Research Dream Team Translational Cancer Research Grant (SU2C-AACR-DT0409). EMS was funded by fellowships from The Wellcome Trust and the California Breast Cancer Research Program. JC was funded by the Medical Scientist Training Program and a pre-doctoral fellowship from the California Breast Cancer Research Program.

FIGURE LEGENDS

Figure A.1. *Zpo1* is a transcriptional repressor expressed in mammary epithelium. (A) Schematic diagram of *Zpo1* protein showing domains conserved with *Drosophila* elbow, Sp - Sp box, Cys - cysteine rich domain, Zn - C₂H₂ zinc finger, P/T - proline and tyrosine rich domain. (B) Wholmount RNA *in situ* hybridization demonstrating *Zpo1* expression in the mammary placodes (black arrowheads) of E12.5 mouse embryo. *Zpo1* expression (yellow) in the developing intestinal epithelium (white arrowheads) by RNA *in situ* hybridization. (C) RT-PCR demonstrating human *ZPO1* expression (arrow) in a panel of normal (samples 1 - 12) and tumor (samples A - L) human breast epithelium. Negative water control (N). (D) Localization of endogenous *Zpo1* protein (green) using rabbit anti-*Zpo1* antibody in mouse EpH4.9 epithelial cells counterstained with DAPI (blue). Localization of V5-tagged *Zpo1* protein (red) in EpH4.9-pEiZ and EpH4.9-*Zpo1* cells using an anti-V5 antibody. *Zpo1*-V5 protein is detected predominantly in the nucleus. Scale bar, 50µm. (E) 3T3 cells were transfected with 400 ng of the GAL4 luciferase reporter construct pFR-Luc and increasing quantities of either a *Zpo1*-GAL4 expression construct or vector control and assayed for luciferase activity. Firefly luciferase activity was normalized to Renilla luciferase and activity from *Zpo1*-GAL4 transfected cells is expressed relative to control. Data represent the mean of three separate experiments with each experiment performed in triplicate. Error bars represent ±SEM. (F) 3T3 cells were transfected with 400 ng pFR-Luc and 20 ng *Zpo1*-GAL4 expression construct or vector control. Increasing concentrations of Trichostatin A (5-800 nM) were

added to cell cultures after 24 hours and luciferase activity was measured and normalized to Renilla after a further 24 hours. Error bars represent \pm SEM. (G) *Grg4* expression (yellow) in E13.5 mouse embryonic mammary epithelium (arrowhead) by RNA *in situ* hybridization. (H) LacZ staining of E12.5 *Grg4* gene trap mouse embryo demonstrating *Grg4* expression in the mammary placodes (arrowheads). (I) Co-immunoprecipitation of Zpo1 and Grg4 from 3T3 cells transfected with FLAG-Grg4 and Zpo1-V5 or LacZ-V5 expression plasmids. Lysates were immunoprecipitated with anti-FLAG antibody and blotted with anti-V5 antibody. (J) 3T3 cells were transfected with 400 ng pFR-Luc, 20 ng Zpo1-GAL4 and increasing quantities of a *Grg4* expression plasmid or a vector control. Luciferase activity was measured and normalized to Renilla after 48 hours. Activity from Grg4 transfected cells is expressed relative to vector control. Error bars represent \pm SEM.

Figure A.2. *Zpo1* overexpression promotes epithelial migration and invasion. (A) Migration of control (pEiZ) and *Zpo1* overexpressing (Zpo1) EpH4.9 cells in wound healing assays. Distance travelled is averaged from 6 separate assays per cell line and 3 separate measurements per assay; $p < 0.05$. (B) Quantification of Matrigel invasion of control (pEiZ) and *Zpo1* overexpressing (Zpo1) EpH4.9 cells. Experiments performed in triplicate; $p < 0.05$. (C) Lysates from EpH4.9-pEiZ and EpH4.9-Zpo1 cells analyzed by Western blot for adherens junction proteins. (D) Brightfield images of EpH4.9-pEiZ and EpH4.9-Zpo1 cells cultured in 3D Matrigel for 4 days. Scale bar, 500 μ m upper panels, 50 μ m lower panels.

(E) Immunofluorescent localization of E-cadherin, β -catenin, p120-catenin and α -catenin (green) in EpH4.9-pEiZ and EpH4.9-Zpo1 cells cultured in 3D Matrigel. F-actin localization is shown in red using Alexa Fluor 568 conjugated phalloidin. Cells counterstained with DAPI (blue). Arrows point to regions of reduced and/or re-localized protein expression. Scale bar, 50 μ m. (F) Quantification of *Zpo1* expression in EpH4.9 cells stably infected with control or *Zpo1* specific shRNA (C5) by Western blot and qPCR using GAPDH as a control. (G) Brightfield images of EpH4.9-pEiZ and EpH4.9-Zpo1 cells stably infected with scramble or *Zpo1* specific shRNA lentivirus and cultured in 3D Matrigel. Scale bar, 100 μ m. (H) EpH4.9-pEiZ and EpH4.9-Zpo1 cells stably expressing a *Zpo1* shRNA (C5) cultured in 3D Matrigel and stained with phalloidin demonstrating lumen formation (arrowhead). Scale bar, 50 μ m.

Figure A.3. *Zpo1* promotes increased proliferation and represses *E-cadherin* expression. (A) Proliferation of control (solid line) and *Zpo1* overexpressing (dotted line) EpH4.9 cells in 2D culture. Cells in 3 wells counted for each time-point. (B) Quantification of EpH4.9-pEiZ and EpH4.9-Zpo1 nuclei positive for phospho-histone H3 when cultured in 3D Matrigel for 48 hours. Error bars represent \pm SEM. (C) mRNA quantification by qPCR of AJ proteins from EpH4.9-pEiZ and EpH4.9-Zpo1 cells normalized to GAPDH expression. Error bars represent mean \pm SD. (D) 3T3 cells were transfected with 400 ng pEcad-Luc reporter plasmid and increasing quantities of *Zpo1*-V5 expression plasmid and assayed for luciferase activity. (E) 3T3 cells were transfected with pEcad-Luc and

the effect of β -catenin and *Zpo1* expression was determined. (F) 3T3 cells were transfected with TOPFLASH reporter and increasing quantities of Zpo1-V5. (G) 3T3 cells were transfected with a TGF β reporter plasmid and increasing quantities of Zpo1-V5, with or without exogenous TGF β .

Figure A.4. *Zpo1* overexpression promotes EMT and p120-catenin isoform expression and localization. (A) Immunofluorescent localization of vimentin, Snail, N-cadherin and cytokeratin (green) in EpH4.9-pEiZ and EpH4.9-Zpo1 cells in 3D Matrigel culture. Arrows in vimentin staining of EpH4.9-Zpo1 cells point to vimentin-positive migratory cells. Cells counterstained with DAPI (blue). Scale bar, 50 μ m. (B) Western blot of AJ proteins from 3D cultures of control and *Zpo1* overexpressing EpH4.9 cells after 6 days in culture. (C) Western blot showing p120-catenin isoform expression in 2D and 3D cultured EpH4.9-pEiZ and EpH4.9-Zpo1 cells using antibodies recognizing all p120-catenin isoforms (p120-all) or specifically isoforms 1 and 2 (p120-1/2). (D) Brightfield image of EpH4.9-Zpo1 cells stably transfected with control or p120-3A expression constructs and cultured in 3D Matrigel for 48 hours. Scale bar, 100 μ m. (E) Brightfield image of EpH4.9-pEiZ and EpH4.9-Zpo1 cells cultured on 2D Matrigel (left panels). Immunolocalization of all p120-catenin isoforms (p120-all) or isoforms 1/2 only (p120-1/2) demonstrates nuclear localization of p120 isoform 3 only in *Zpo1* overexpressing cells (red). Scale bar, 100 μ m.

Figure A.5. *Zpo1*-mediated changes in cell adhesion and migration are regulated

by the ECM and require Rho GTPase and Src family kinase activity. (A) Low magnification brightfield images of EpH4.9-pEiZ and EpH4.9-Zpo1 cells cultured for 5 days in the absence or presence of the anti-laminin 332 antibody, showing increased EpH4.9-Zpo1 cell migration in the presence of the antibody. Scale bar, 100 μ m. High magnification brightfield images of EpH4.9-Zpo1 cells cultured in the presence of the anti-laminin 332 antibody, showing elongated cells with long, membrane protrusions (arrows). Scale bar, 10 μ m. (B) Brightfield images of EpH4.9-pEiZ and EpH4.9-Zpo1 cells cultured in the presence of the ROCK inhibitor Y-27632 (5 μ M) and Rac1 inhibitor (100 μ M). Scale bar, 100 μ m. (C) Brightfield images of EpH4.9-pEiZ and EpH4.9-Zpo1 cells cultured in 3D Matrigel in the presence of the SFK inhibitor SU6656 (5 μ M). Scale bar, 100 μ m. (D) F-actin staining of EpH4.9-pEiZ and EpH4.9-Zpo1 cells cultured in 3D in the presence of 5 μ M SU6656 showing lumen formation (arrows). Scale bar, 100 μ m.

Figure A.6. *Zpo1* overexpression increases breast cancer metastasis. (A) Graph showing weight of primary tumors formed by 4T1-pEiZ and 4T1-Zpo1 cells, 3 weeks post-transplantation. (B) Graph showing number of lung surface metastases formed by 4T1-pEiZ and 4T1-Zpo1 cells 3 weeks post-transplantation. Error bars represent \pm SEM. (C) Representative hematoxylin-eosin stained section of lungs from mice transplanted with 4T1-pEiZ and 4T1-Zpo1 cells containing lung metastases (dotted lines). Scale bar, 200 μ m. (D) Western blot of E-cadherin protein from 3 primary tumor lysates isolated from 4T1-pEiZ and 4T1-Zpo1 tumors. Immunofluorescence staining of E-cadherin

protein in sections of primary tumors formed from 4T1-pEiZ and 4T1-Zpo1 cells. Scale bar, 50 μ m. (E) Graph showing weight of primary tumors formed by 4T1 cells expressing either a scrambled or *Zpo1* specific (C5) shRNA, 3 weeks post-transplantation. Representative images of lungs from mice transplanted with shRNA expressing 4T1 cells showing a reduction in lung surface metastases from *Zpo1* knockdown cells.

Figure A.7. Model of *Zpo1* overexpression and metastasis promotion. Amplification of chromosome 8p11-12 in a mammary tumor cell leads to overexpression and increased activity of *Zeppo1*. As a result, *E-cadherin* expression is reduced with a consequent reduction in tumor cell adhesion. Loss of E-cadherin at the cell membrane reduces membrane localized p120-catenin and increases the pool of cytoplasmic p120-catenin that is able to interact with Rho-GTPases. At the same time *Zeppo1* overexpression also promotes a switch in p120-catenin isoform expression from isoform 3 to isoform 1 that promotes increased cell migration by binding to and inhibiting RhoA while increasing Rac1 activity. *Zeppo1* overexpression also reduces tumor cell interactions with the ECM and increases tumor cell migration through unknown mechanisms. In combination, these effects lead to a more proliferative tumor with increased migratory and invasive behavior leading to increased tumor metastasis.

Supplemental Figure A.S1. Overexpression of Zpo1-V5 in EpH4.9 cells. Quantitative PCR of Zpo1 mRNA levels in EpH4.9-pEiZ and EpH4.9-Zpo1 cells normalized to actin expression. Error bars represent mean \pm SD.

Supplemental Figure A.S2. Immunofluorescent localization of adherens junction proteins in 2D cultured EpH4.9-Zpo1 cells. Immunofluorescent localization of E-cadherin, β -catenin, p120-catenin and α -catenin (green) in 2D cultured EpH4.9-pEiZ and EpH4.9-Zpo1 cells counterstained with DAPI (blue). Scale bar, 50 μ m.

Supplemental Movie A.1. EpH4.9-pEiZ cells in wound healing assay. EpH4.9-pEiZ cells grown to confluence in a 24-well plate and wounded with a pipette tip. Cells were imaged over a 49hr period with images taken every 5 minutes.

Supplemental Movie A.2. EpH4.9-Zpo1 cells in wound healing assay. EpH4.9-Zpo1 cells grown to confluence in a 24-well plate and wounded with a pipette tip. Cells were imaged over a 49hr period with images taken every 5 minutes.

Supplemental Movie A.3. Morphogenesis of 3D cultured EpH4.9-pEiZ cells. EpH4.9-pEiZ cells were aggregated overnight on low adhesion culture plates then embedded in Matrigel. Cells were cultured at 37 °C in 5% CO₂ and imaged for 5 days with images taken every 15 minutes.

Supplemental Movie A.4. Morphogenesis of 3D cultured EpH4.9-Zpo1 cells. EpH4.9-Zpo1 cells were aggregated overnight on low adhesion culture plates then embedded in Matrigel. Cells were cultured at 37 °C in 5% CO₂ and were imaged for 5 days with images taken every 15 minutes.

FIGURES

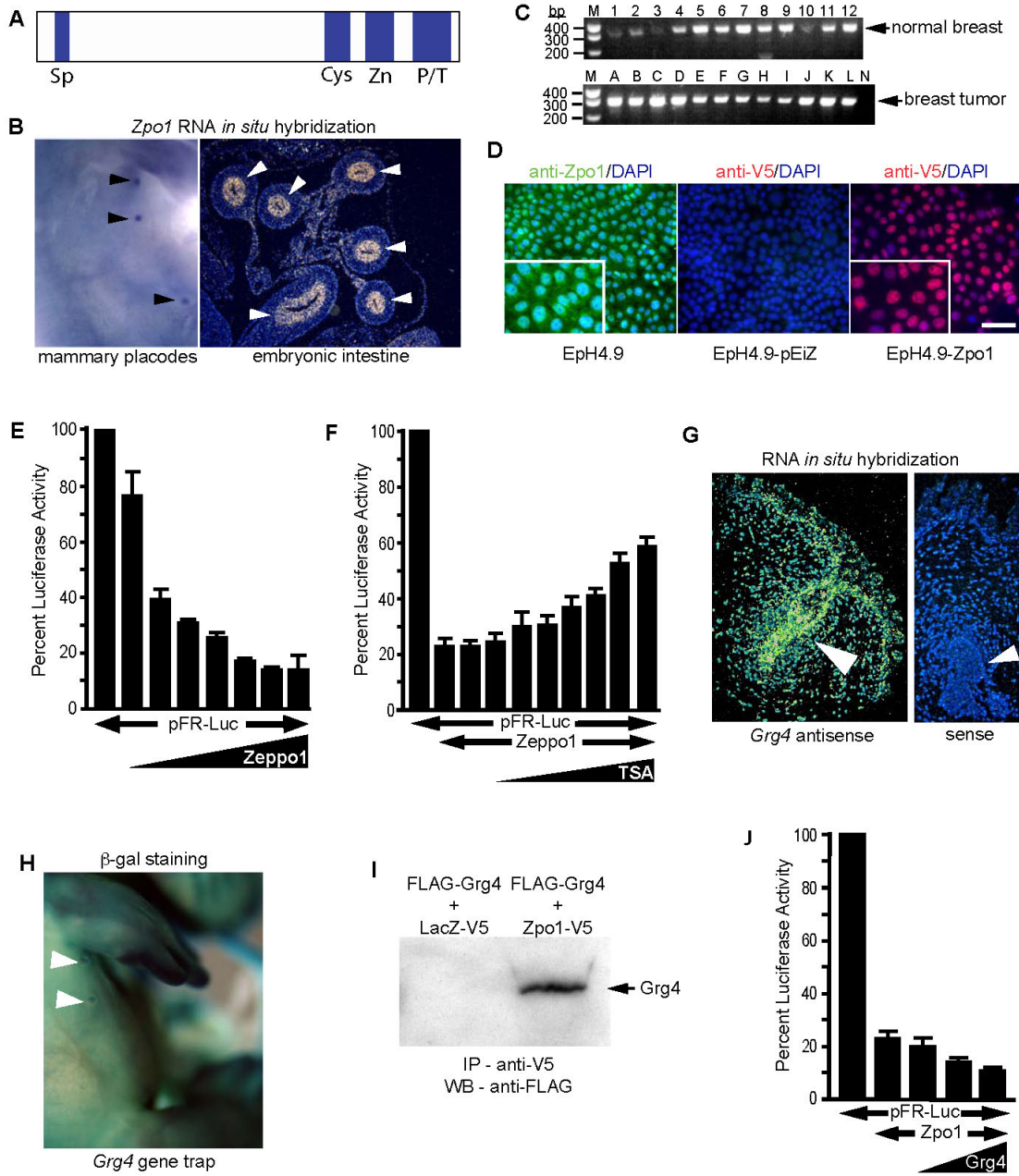


Figure A.1

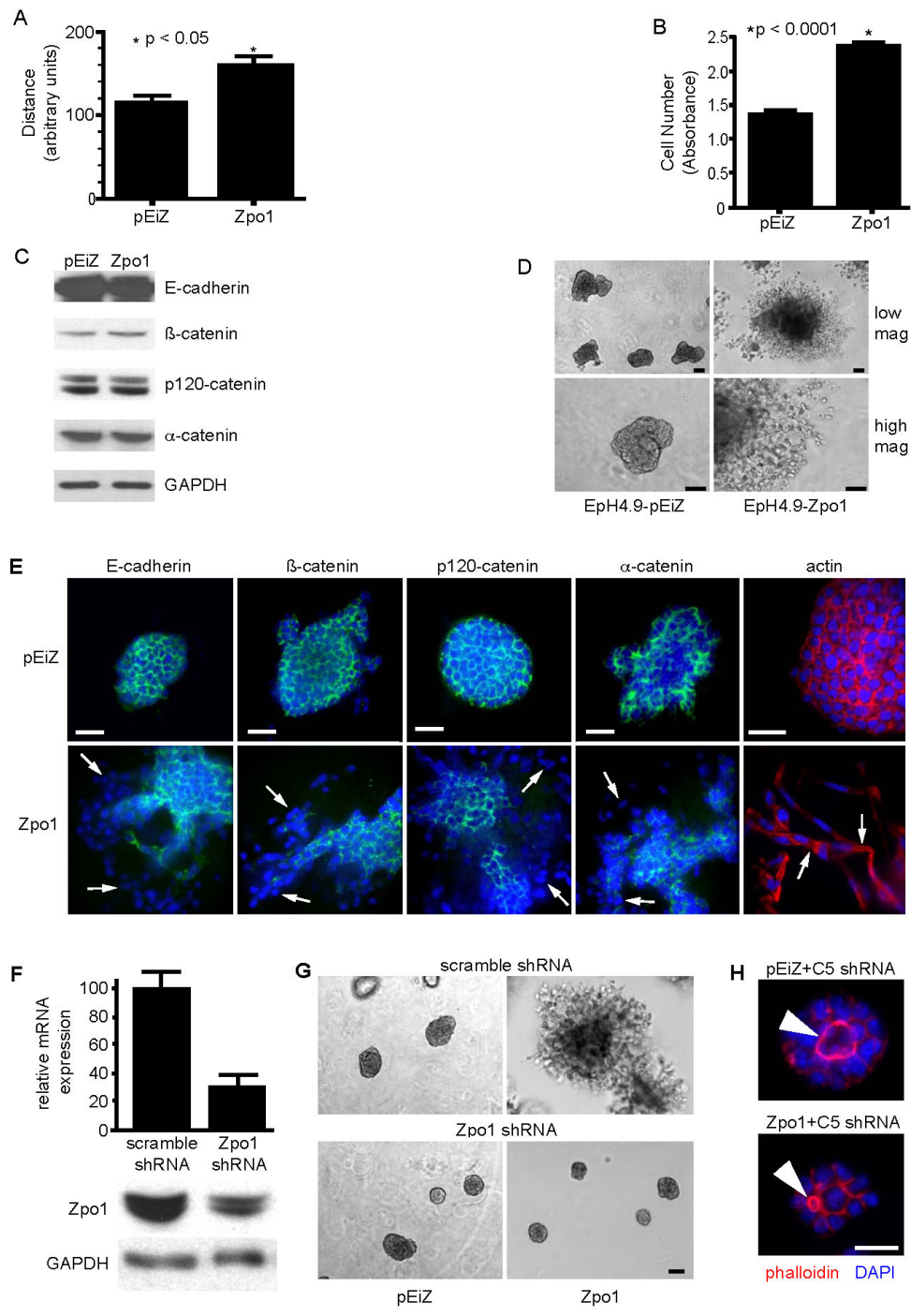


Figure A.2

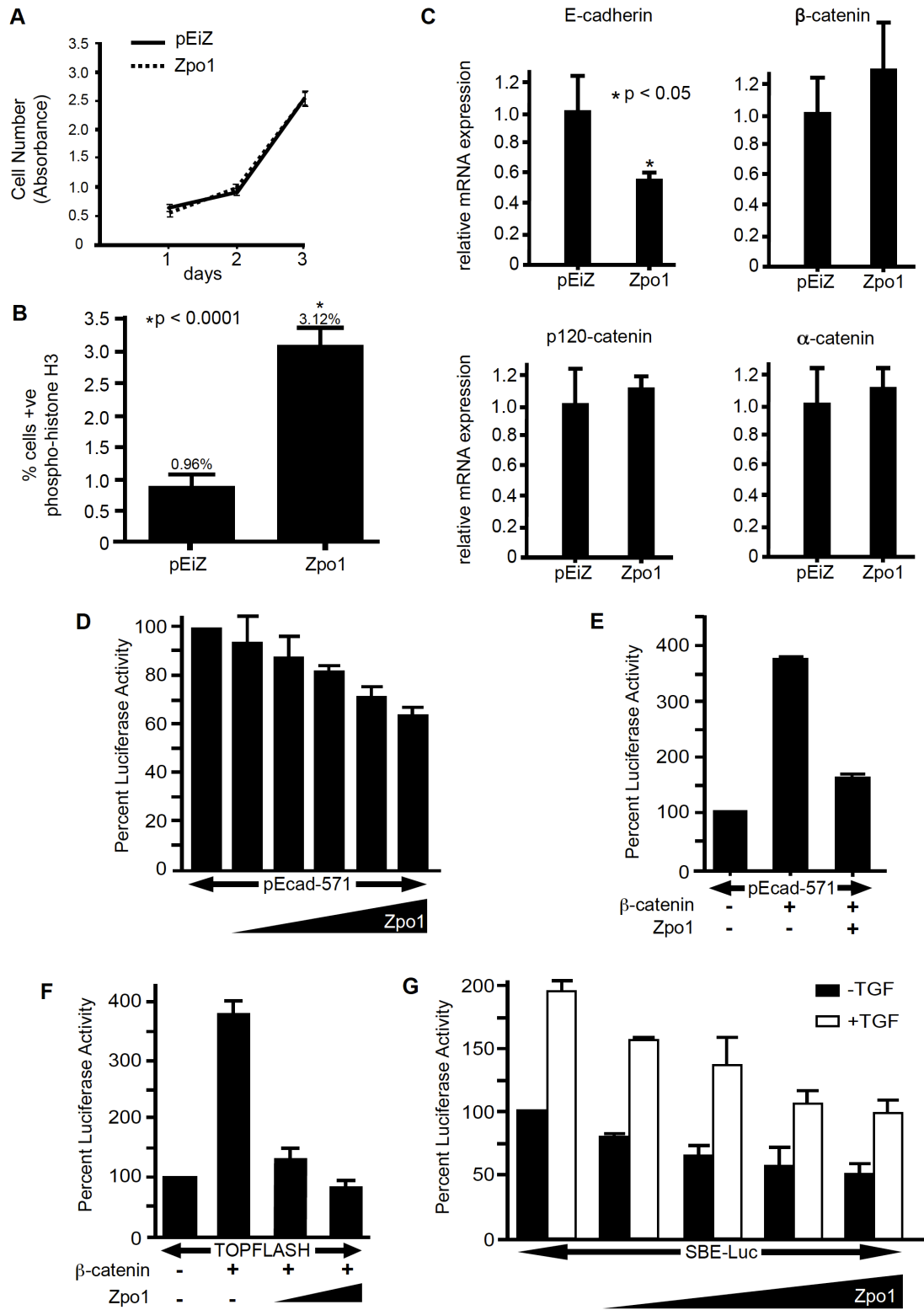


Figure A.3

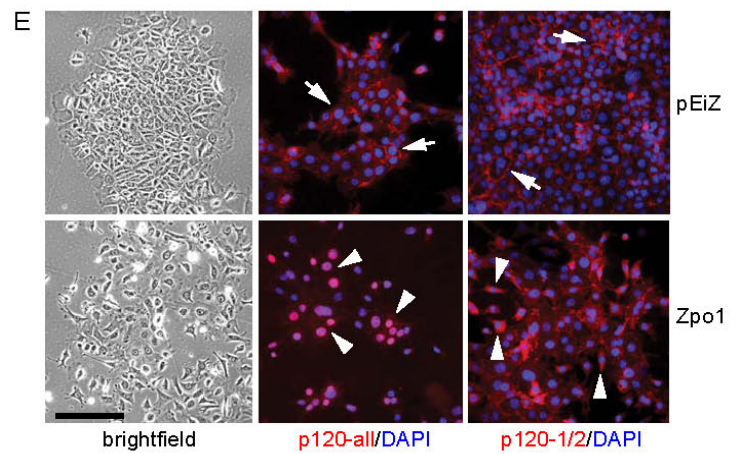
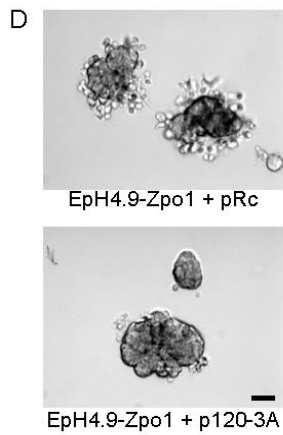
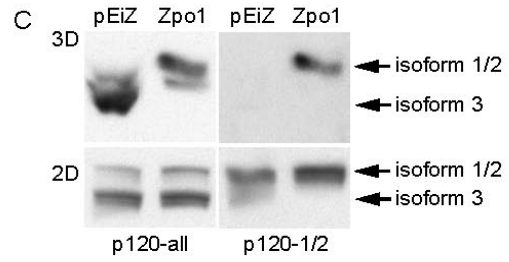
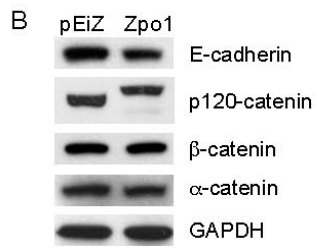
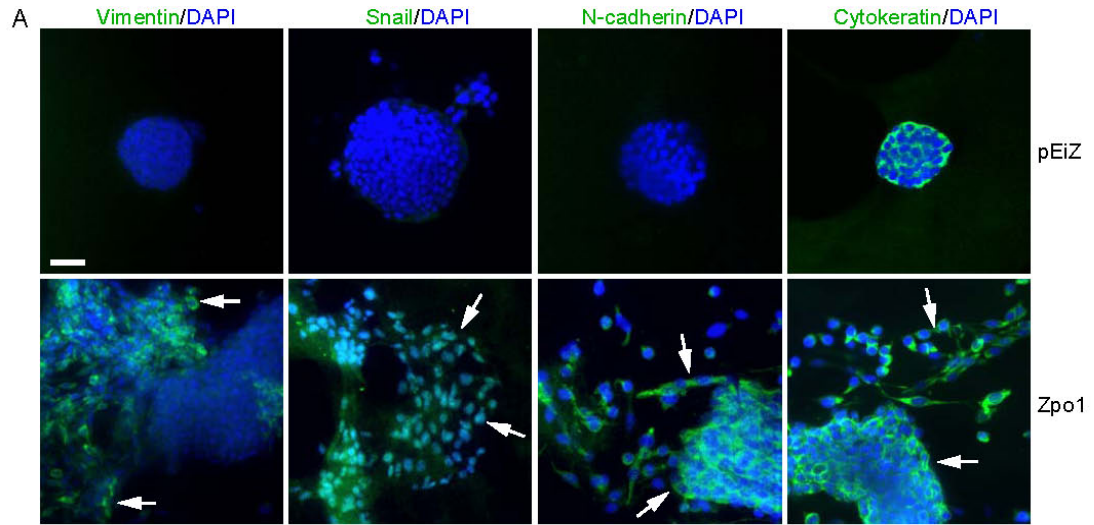


Figure A.4

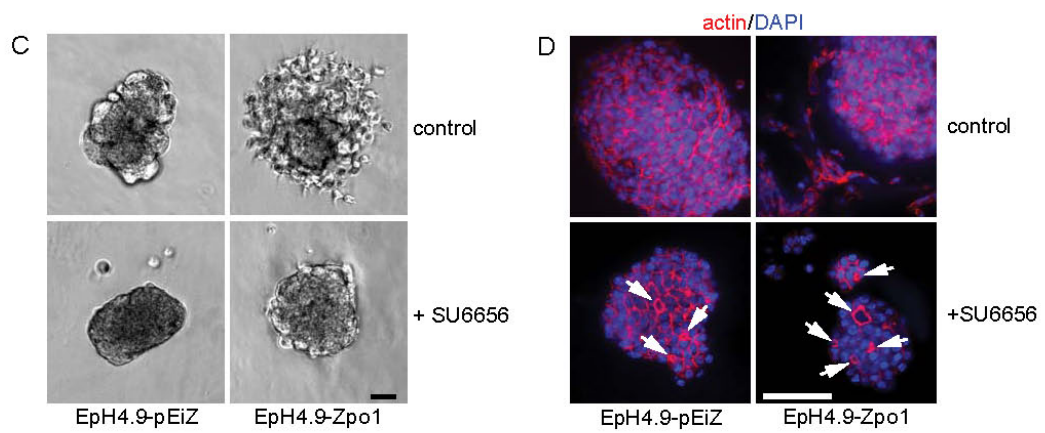
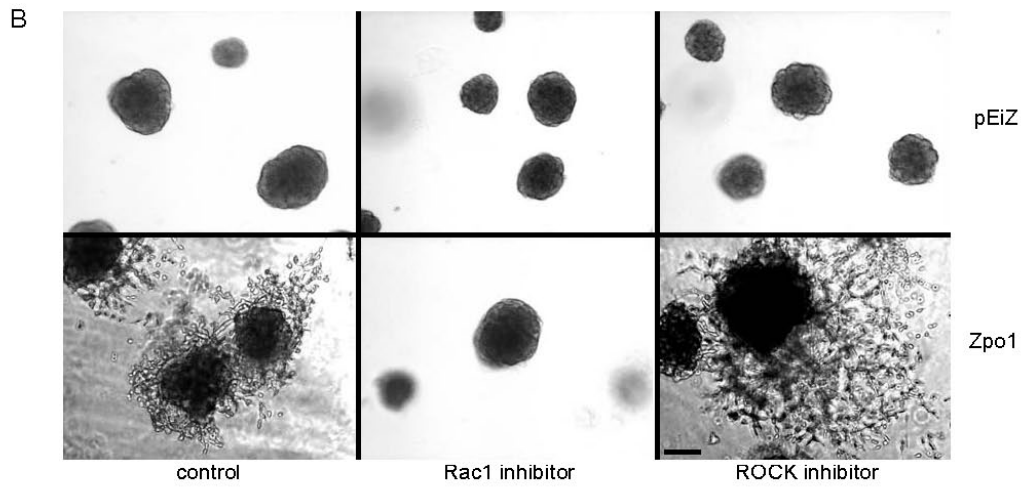
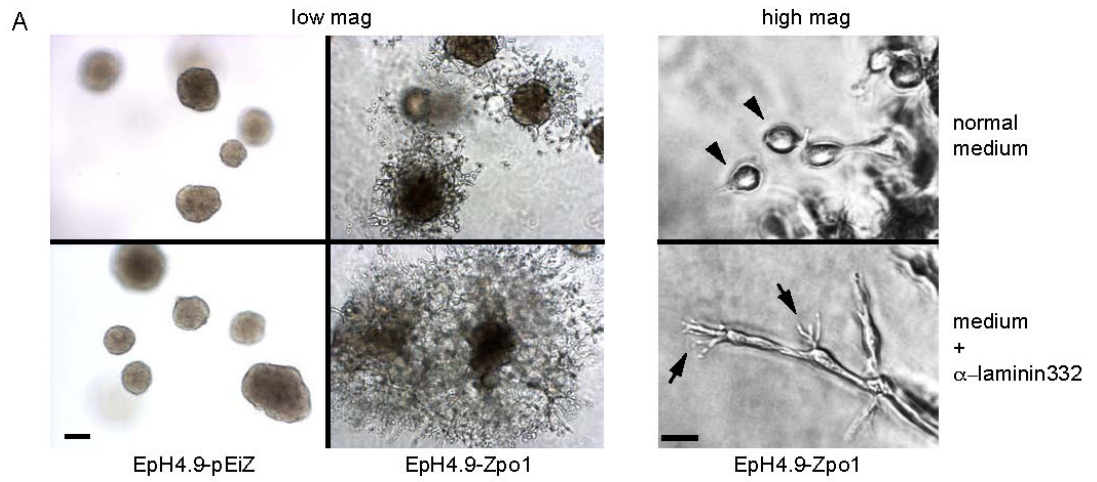


Figure A.5

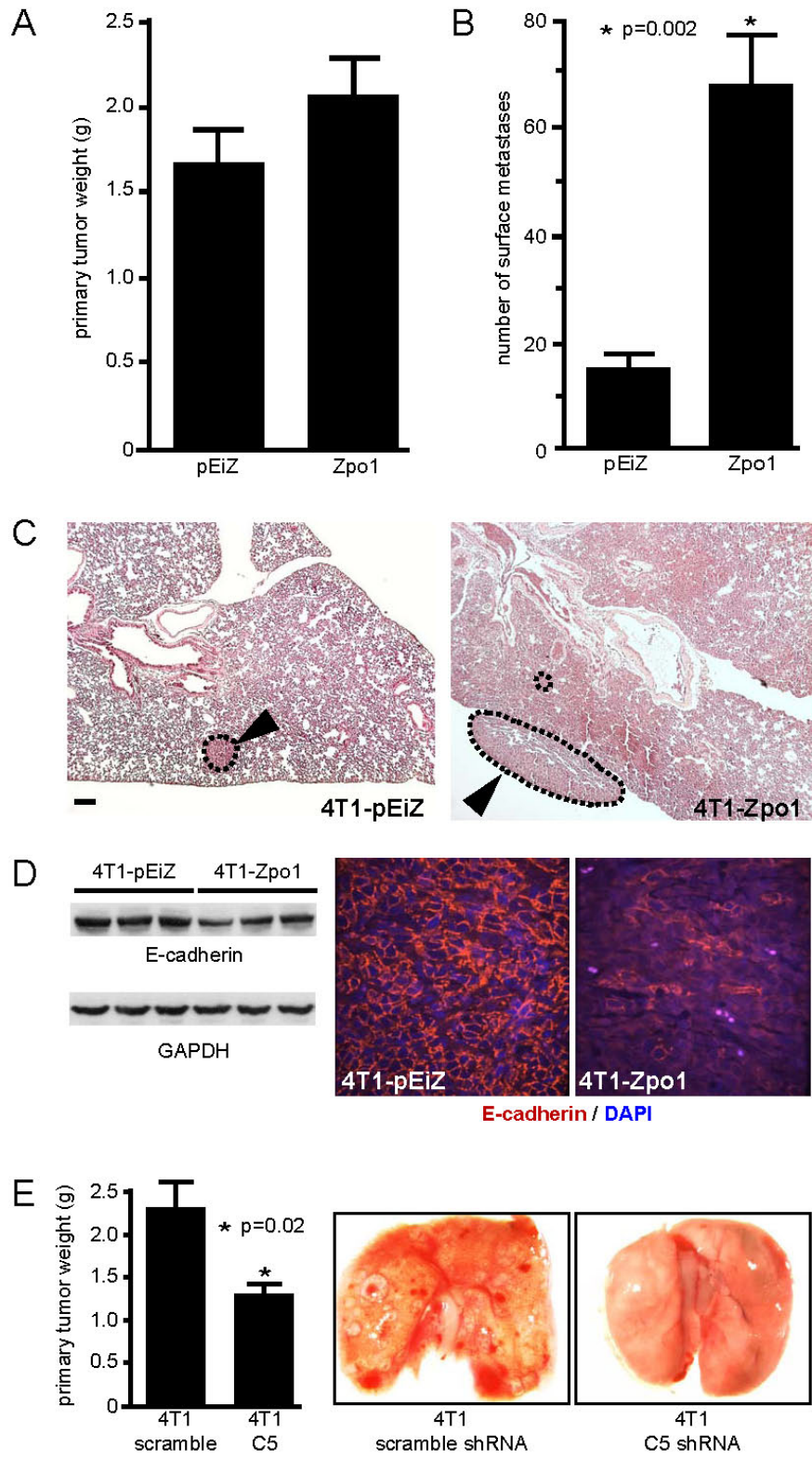


Figure A.6

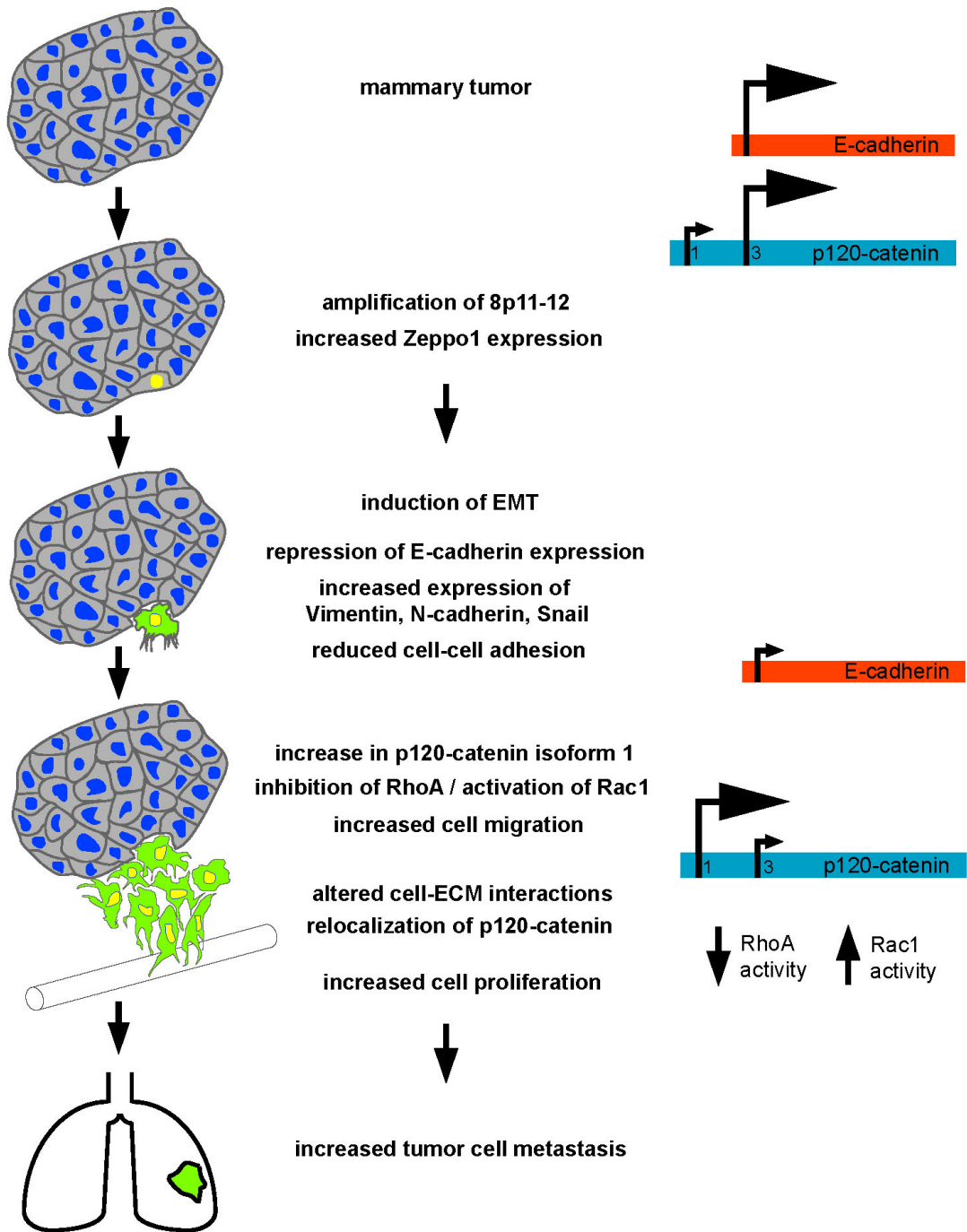
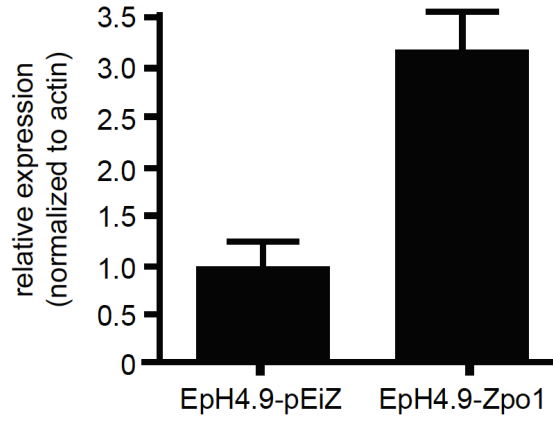
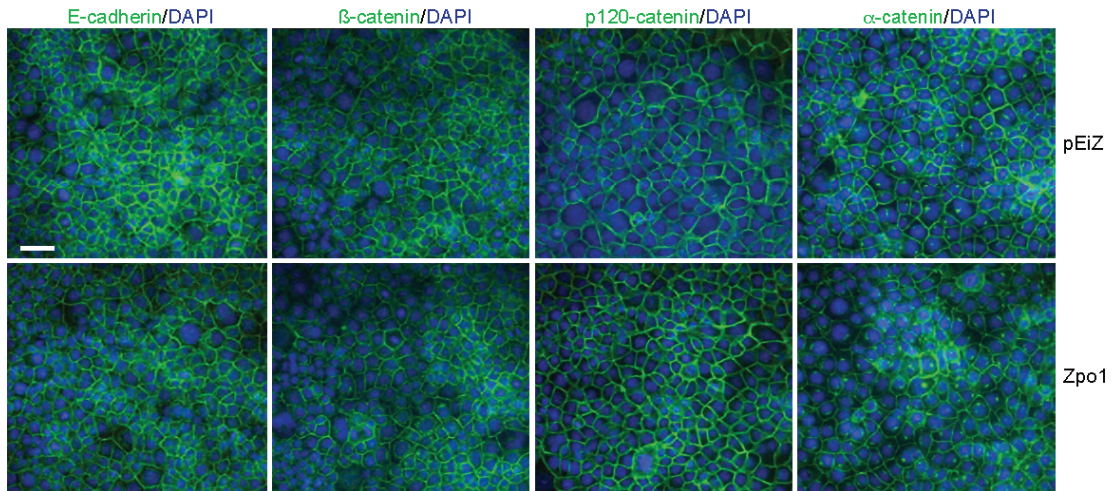


Figure A.7



Supplemental Figure A.S1



Supplemental Figure A.S2

Appendix B: The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression

Source: The following appendix was published as a manuscript from: Littlepage LE, et al. (2012) The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression. *Cancer Discov* 2(7), 1-14.

Contributions: For this project, I discussed project ideas with Laurie Littlepage, who conceived the project and performed most of the experiments described. I performed some of the qPCR experiments, helped generate the VO-PyMT cell line used in the in vivo experiments, helped treat mice with triciribine, performed 3D organoid cultures and provided help with statistical analysis. Zena Werb supervised the project. The entire manuscript is included here to give context to the data.

The transcription factor ZNF217 is a prognostic biomarker and therapeutic target during breast cancer progression

Laurie E. Littlepage^{1,2}, Adam S. Adler^{3,7}, Hosein Kouros-Mehr^{1,7}, Guiqing Huang^{1,2}, Jonathan Chou^{1,2}, Sheryl R. Krig⁴, Obi L. Griffith⁵, James E. Korkola⁵, Kun Qu³, Devon A. Lawson^{1,2}, Qing Xue¹, Mark D. Sternlicht^{1,2,8}, Gerrit J. P. Dijkgraaf^{1,7}, Paul Yaswen⁵, Hope S. Rugo², Colleen A. Sweeney⁴, Colin C. Collins^{2,9}, Joe W. Gray^{2,5,10}, Howard Y. Chang^{3,6}, Zena Werb^{1,2}

¹Department of Anatomy, University of California, San Francisco, California 94143-0452

²Helen Diller Family Comprehensive Cancer Center, University of California, San Francisco, California 94143

³Program in Epithelial Biology, Stanford University School of Medicine, Stanford, California 94305

⁴Division of Basic Sciences, University of California at Davis Cancer Center, Sacramento, California, 95817

⁵Life Sciences Division, Lawrence Berkeley National Laboratory, Berkeley, California 94720

⁶Howard Hughes Medical Institute

Present addresses:

⁷Genentech, South San Francisco, CA 94080

⁸FibroGen, 409 Illinois Street, San Francisco, CA 94158

⁹The Vancouver Prostate Centre, Vancouver, BC V5Z 1M9, Canada

¹⁰Department of Biomedical Engineering, Oregon Health & Science University, Portland, OR 97239

Corresponding Author:

Zena Werb, Ph.D.

Department of Anatomy, Box 0452

513 Parnassus Avenue

San Francisco, CA 94143-0452

E-mail: zena.werb@ucsf.edu;

Phone: 415-476-4622; Fax: 415-476-4565

Disclosure of Potential Conflicts of Interest:

Authors disclosed no potential conflicts of interest.

Keywords:

Breast cancer; metastasis; gene amplification; cancer stem cells; epithelial-mesenchymal transition; AKT; chemotherapy resistance

ABSTRACT

The transcription factor ZNF217 is a candidate oncogene in the amplicon on chromosome 20q13 that occurs in 20-30% of primary human breast cancers and that correlates with poor prognosis. We show Znf217 overexpression drives aberrant differentiation and signaling events, promotes increased self-renewal capacity, mesenchymal marker expression, motility and metastasis, and represses an adult tissue stem cell gene signature downregulated in cancers. By in silico screening, we identified candidate therapeutics that at low concentrations inhibit growth of cancer cells expressing high ZNF217. We demonstrate that the nucleoside analog tricyridine inhibits ZNF217-induced tumor growth and chemotherapy resistance, and inhibits signaling events (e.g., P-AKT, P-MAPK) in vivo. Our data suggest ZNF217 is a biomarker of poor prognosis and a therapeutic target in breast cancer patients, and tricyridine may be part of a personalized treatment strategy in patients overexpressing ZNF217. Since ZNF217 is amplified in numerous cancers, these results have implications for other cancers.

SIGNIFICANCE

This study finds that ZNF217 is a poor prognostic indicator and therapeutic target in breast cancer patients and may be a strong biomarker of tricyridine treatment efficacy in patients. Because previous clinical trials for tricyridine did not include biomarkers of treatment efficacy, this study provides a rationale for revisiting tricyridine in the clinical setting as a therapy for breast cancer patients who overexpress ZNF217.

INTRODUCTION

In the most aggressive breast tumors, neoplastic cells activate or amplify oncogenes, or inactivate or delete tumor suppressor genes to promote invasive growth and poor prognosis in patients. Amplification of the human chromosomal region 20q13 occurs in 20-30% of primary human breast cancers, as well as in other cancers, and its amplification correlates with poor patient prognosis (Collins et al. 1998; Collins et al. 2001).

The ZNF217 gene on human 20q13.2 encodes a transcription factor that is overexpressed in all breast tumors and cell lines in which the gene is amplified, as compared to normal mammary tissue and epithelial cells (Collins et al. 1998; Collins et al. 2001). The ZNF217 protein is a member of the C₂H₂ family of transcription factors and contains eight predicted Kruppel-like C₂H₂ zinc finger motifs and a proline-rich region. It is a component of a human histone deacetylase complex (CoREST-HDAC) and is found in complexes with the transcriptional co-repressor C-terminal binding protein (CtBP), the histone demethylases LSD1 (H3K4, H3K9) and KDM5B/JARID1B/PLU-1 (H3K4) and the methyltransferases G9a (H3K9, H3K27) and EZH2 (H3K27) (You et al. 2001; Shi et al. 2003; Quinlan et al. 2006; Cowger et al. 2007; Thillainadesan et al. 2008; Banck et al. 2009). Its overexpression in human mammary epithelial cells (MECs) overcomes senescence and promotes immortalization accompanied by increased telomerase activity, increased resistance to TGFβ-induced growth inhibition and amplification of c-Myc (Nonet et al. 2001). ZNF217 binds to the promoters of genes involved in differentiation and is repressed following retinoic

acid treatment of pluripotent embryonal cells (Krig et al. 2007).

In this study we investigated whether and how Znf217 promotes tumor progression and poor prognosis using cultured cells, in vivo transplant models and human patient expression datasets.

RESULTS

ZNF217 is Prognostic of Poor Survival in Breast Cancer Patients

Using microarray expression data from primary breast tumors and corresponding clinical data (Chin et al. 2006; Sorlie et al. 2006), we found that high ZNF217 amplification and expression correlate with shorter overall survival, disease-specific survival and relapse-free survival (**Figures B.1A–B, B.S1A–B**). To determine if ZNF217 overexpression overlapped with another poor prognostic subtype, we compared ZNF217 expression levels across patient subtypes (e.g., ER+, ER-, ERBB2/HER2+, ERBB2/HER2-, luminal and basal patient cohorts) and found that ZNF217 expression levels are highest in ER+ tumors and lowest in basal subtype tumors (data not shown).

We next determined if ZNF217 had prognostic value across breast cancer subtypes. We compared survival and ZNF217 expression by univariate analysis across ER+, ER-, HER2+, luminal and basal subtypes. Patients with tumors expressing high ZNF217 consistently had reduced survival compared to patients with tumors expressing lower ZNF217 across multiple breast cancer subtypes (**Figure B.1C** and data not shown). For example, in a meta-analysis of relapse-free survival across nine published studies that included 858 patients (ER+HER2-LN-) with ZNF217 expression, we found that ZNF217 expression was significantly associated with 5 yr ($P=0.012$) and 10 yr ($P=0.023$) relapse status (Mann-Whitney), and patients with relapse had higher ZNF217 expression. Similarly, patients grouped into the high expression tertile had significantly worse survival than low expression groups. These data show that ZNF217 is prognostic

of poor survival in patients by univariate analysis. Moreover, ZNF217 was a better predictor of survival than ER status by multivariate analysis (**Figure B.S1C**).

Overexpression of Znf217 Accelerates Loss of Adhesion and Increased Motility in Mouse Mammary Epithelial Cells

To determine the consequences of Znf217 overexpression, we generated mouse mammary epithelial cell lines that overexpressed Znf217 by retroviral and lentiviral infection. Mouse mammary epithelial cell lines (SCp2, NMuMG, EpH4) overexpressing Znf217 had altered motility showing a more scattered phenotype than adherent, clustered control cells (**Figures B.2A–C, B.S2A–C**). In a wound healing/scratch assay, individual SCp2 cells (**Figure B.2D**) and NMuMG cells (data not shown) overexpressing Znf217 showed increased motility, forward extended lamellipodia and independent migration (i.e., separate from other cells) towards the middle of the scratch (**Figure B.2D, Supplemental Movie B.S1**). In contrast, vector-treated cells migrated as a sheet predominantly with a single leading edge. In keeping with the increased motility, cells overexpressing Znf217 reorganized their actin cytoskeleton with reduced cortical actin and increased actin stress fibers (**Figure B.2E**) and upregulated EMT markers including Snail1 and Twist (**Figure B.2F**). Consistently, we found ZNF217 expression levels correlated with expression of EMT markers, including Snail1, Snail2 and Vimentin, genes that have ZNF217 enriched at their promoters in human breast cancer cell lines and tumors (**Table B.S1**). Taken together, these results indicate

that the early effects of increased ZNF217 expression would lead to premalignant changes of enhanced mammary epithelial migration.

We used gene expression microarrays of the mouse SCp2 MECs overexpressing Znf217 to identify altered processes and molecular targets of Znf217 (**Figure B.2G**). In these cells, 176 genes were upregulated and 243 genes downregulated following Znf217 overexpression. We used DAVID software to classify the gene sets with gene ontology (GO) terms (**Table B.S2**). The GO terms suggested that Znf217 overexpression promoted increased cell motility, decreased epithelial differentiation, increased vasculature development and changes at the membrane.

We also assayed for clonogenicity and transformation potential in vitro. Znf217 overexpression in NIH3T3 cells stimulated anchorage independent growth in a soft agar assay, with increased number and size of colonies (**Figures B.3A–C**).

Znf217 Overexpression in Normal Primary Mammary Epithelium Promotes Increased Mammosphere Formation in Culture

Because we found that Znf217 overexpression in culture promoted increased motility, decreased epithelial marker expression/increased mesenchymal marker expression and increased clonogenicity/transformation potential in vitro, we reasoned that these changes were consistent with a change in differentiation towards a less differentiated or more mesenchymal phenotype. Znf217 gene expression was enriched in CD24^{Med}CD49f^{High} cell population, which includes

basal, myoepithelial and progenitor cells, compared to CD24^{High}CD49f^{Low} cells, which include luminal and luminal progenitor cells (**Figure B.3D–E**).

To determine if Znf217 could promote progenitor cell phenotypes, we overexpressed Znf217 in normal primary MECs and analyzed clonogenicity potential in the mammosphere assay. Primary MECs infected with Znf217-overexpressing lentivirus and grown in serum-free nonadherent culture conditions demonstrated increased self-renewal capacity (**Figure B.3F–G**).

Gene Expression Analysis of Mammary Epithelial Cells Following Znf217 Overexpression Predicts Changes in Epithelial Proliferation, Cell Adhesion and Motility

We next determined the impact of Znf217 overexpression on global gene expression in normal primary mouse MECs by gene expression microarrays (**Figure B.3H**). Primary MECs were infected with a Znf217-overexpressing lentivirus, passaged to expand the population, and confirmed to overexpress Znf217 by reverse transcription quantitative PCR (rt-qPCR) (**Figure B.3I**). In these MECs, 340 genes were upregulated and 401 genes downregulated following Znf217 overexpression (**Figure B.3H**). The GO terms classified by DAVID software (**Table B.S2**) suggested that Znf217 overexpression altered the gene expression profile of genes involved in cell proliferation, cell adhesion, cell migration, G-protein coupled receptor signaling pathways and ribosomal function.

Genes identified by microarray analysis suggested that overexpressing Znf217 in vivo would promote increased epithelial growth or progenitor cell

expansion (**Figure B.3H, Table B.S2**). These genes included a number of TGF β and Wnt pathway genes (Axud1/Csrnp-1, Bcl9l, Bmper, Bmpr2, follistatin, Samd9l, Sfrp1, Tcf4, Tgfb2, Tgfbr3, Wnt5a). We validated selected genes including Wnt5a and Sfrp1 by rt-qPCR (**Figure B.3J**). These results are consistent with Znf217 promoting differentiation towards a less differentiated cell-like phenotype via aberrant signaling in the TGF β and Wnt pathways. We also found that ZNF217 was required for cell proliferation (**Figure B.S2D**).

We also examined epithelial marker expression in our microarray dataset. We found increased expression of both K17 (myoepithelial marker) and K18 (luminal epithelial marker) expression (**Figure B.3H**). Consistent with these results, we found ZNF217 expression levels correlated with expression of K19 and K8/18, genes with ZNF217 enriched at their promoters in human breast cancer cell lines and tumors and with downregulated expression after ZNF217 knockdown in MCF7 cells (**Tables B.S1, B.S3**).

Znf217 Overexpression in Primary Mammary Epithelial Cells Represses an Adult Stem Cell Expression Signature Downregulated in Cancers

Because the aberrant differentiation markers seen both in vivo and in culture suggest that Znf217 may push mammary epithelial cells to a more progenitor cell-like phenotype, we compared our microarray dataset with our previously defined gene expression signature that we found expressed in adult stem cells (Wong et al. 2008). Many of these adult stem cells are slowly cycling and show balanced differentiation versus self-renewal during normal homeostasis. Adult

stemness as defined by this adult stem cell signature is not correlated with self-renewal. This signature has high expression in normal tissues, where cells predominantly are quiescent and have limited self-renewal, but low expression in cancerous tissues, where cells can self-renew. Therefore, reduced expression of the signature correlates with increased self-renewal (Wong et al. 2008).

Similar to that seen in cancers, primary MECs overexpressing Znf217 significantly repressed genes of the adult stem cell signature (**Table B.S4**; $p=1.89 \times 10^{-10}$), thus making normal cells like self-renewing cancer cells at the expression level. Consistent with the increased clonogenicity by mammosphere assay, this alteration in the expression signature suggests that Znf217 may block differentiation and/or promote self-renewal.

Tumors Overexpressing Znf217 Have a More Basal Pathology with Increased Dual Positive Luminal and Basal Cell Marker Expression

Our finding that Znf217 overexpression promoted decreased differentiation in normal and immortalized mammary epithelium prompted us to determine if these changes are also followed in tumorigenic mammary epithelium both in culture and in vivo. We overexpressed Znf217 by lentiviral delivery of Znf217 into primary luminal-type mammary epithelial tumor cells isolated from 18-week-old MMTV-PyMT mice (PyMT MEC) or a MMTV-PyMT cell line (Vo-PyMT), sorted the cells for the IRES-Tomato reporter gene and confirmed Znf217 overexpression by rt-qPCR (**Figure B.4A**). In culture, Znf217 induced mesenchymal marker expression, reduced expression of E-cadherin, increased

expression of EMT markers Snail2 and Twist (**Figure B.4A**), and produced a more scattered phenotype (**Figure B.S2C**). In addition, cells overexpressing Znf217 readily formed increased numbers of mammospheres compared to vector control cells (**Figure B.4B–C**).

When the sorted cells were transplanted into syngeneic mouse mammary fat pads cleared of epithelium, Znf217 overexpression accelerated the rate of tumor formation, reduced the tumor-free survival and increased both tumor volume and final tumor weight (**Figures B.4D–F, B.S3A**). Tumors overexpressing Znf217 had a markedly altered, heterogeneous histology compared to tumors from vector-treated cells (**Figure B.4G**). Control tumors had little to no smooth muscle actin (SMA) staining with predominantly luminal K8⁺ epithelium, while Znf217-overexpressing tumors expressed higher levels of myoepithelial and myofibroblast SMA protein (**Figure B.4H**).

Tumor cells from vector-treated cohorts were predominantly K8⁺ with few K14⁺ cells; most of the K14⁺ cells were K8⁻ and were located basal to the K8⁺ tumor cells. Znf217 overexpression increased numbers of K14⁺ cells in tumors, with many double-positive K8⁺K14⁺ cells (**Figures B.4I–J** and data not shown). The K8⁺K14⁺ cells may be a bipotent progenitor population capable of forming both luminal and myoepithelial cells, also seen by others as K18⁺K19⁺ cells (Gudjonsson et al. 2002; Woodward et al. 2005; Villadsen et al. 2007; Wang et al. 2008b; McCaffrey and Macara 2009).

We also assayed for epithelial E-cadherin expression in these tumors. In vivo Znf217 overexpression resulted in heterogeneous staining of E-cadherin

with a large number of regions containing a marked reduction in E-cadherin expression, whereas control tumors had E-cadherin localized to the plasma membrane throughout the epithelium (**Figures B.5A, B.S3B**).

Taken together, these phenotypic changes seen within tumors are consistent with Znf217 promoting the acquisition of a mesenchymal/progenitor cell phenotype.

Overexpression of Znf217 Promotes Metastasis

Since metastasis in vivo often is accompanied by increased motility, invasive, and mesenchymal/basal phenotypes, we next asked if Znf217 promotes metastasis in vivo. In mice transplanted with either PyMT or Vo-PyMT cells, Znf217 overexpression significantly increased the percentage of mice with lung metastases, increased metastatic burden and increased the number of spontaneous lung metastases per mouse (**Figures B.5B–D, B.S3C**). In keeping with these results, high ZNF217 expression was prognostic of reduced metastasis-free survival in breast cancer patients (**Figure B.5E**) (Minn et al. 2005).

ZNF217 Promotes Resistance to Chemotherapy

Patients with tumors expressing stem cell-like/progenitor cell markers have increased resistance to chemotherapy (Li et al. 2008). We determined if ZNF217 was a prognostic predictor of treatment response by comparing clinical data to expression data from patients who received neoadjuvant chemotherapy. In one

dataset, patients received either doxorubicin or a combination of 5-fluorouracil and mitomycin (FUMI) before surgical removal and microarray gene expression analysis of the tissue (Sorlie et al. 2006). Tumors that responded to treatment (i.e., became smaller) expressed less ZNF217 than did nonresponsive tumors (**Figure B.6A**).

We also found that tumors with low ZNF217 expression responded better to treatment (pathological complete response) than did tumors with high ZNF217 expression. In a second dataset, fine needle aspirate samples were collected prior to treatment and used for gene expression (Hess et al. 2006). All patients received similar preoperative treatments (paclitaxel and fluorouracil-doxorubicin-cyclophosphamide). We found ZNF217 expression levels higher in nonresponsive tumors that did not respond to therapy than in tumors that responded (**Figure B.6A**). Therefore, ZNF217 is a prognostic predictor of patient outcome in response to chemotherapy.

ZNF217 Levels are Related to Levels of Activated AKT, MAPK and ERBB3

ERBB3 is a direct target of ZNF217 (Krig et al.). We found significant correlation between ZNF217 and ERBB3 expression levels in human breast tumors (**Figure B.6B**). To determine the mechanism underlying the effects of ZNF217, we analyzed several downstream signaling molecules downstream of ERBB3 in two human breast cancer cell lines (MCF7, ZR-75-1) that express high levels of ZNF217. Using both wild type cells and cells after knockdown for ZNF217 expression by siRNA, as described previously (Krig et al.), we treated serum-

starved cells with the growth factor ligand for ERBB3, heregulin/neuregulin (HRG). Cells containing ZNF217-siRNAs consistently required higher concentrations of HRG to induce ERBB3 signaling, phospho-AKT and phospho-MAPK (**Figures B.6C, B.S4A–D**). These data indicate that these pathways are downstream of ZNF217 and that ZNF217 sensitizes cells to HRG.

ZNF217 is a Drug Target for Individualized Therapy

Since ZNF217 is overexpressed in poor prognostic and chemoresistant breast cancer patients, we sought to identify drugs that kill tumor cells overexpressing ZNF217. We first used a candidate approach to determine if AKT pathway inhibitors promoted cell death in a ZNF217-dependent manner, since ZNF217 is required for and promotes AKT activation (**Figure B.6C** and (Huang et al. 2005)). MCF7 cells were infected with virus expressing shRNAs to ZNF217 and validated for reduced protein expression (**Figure B.S4E–F**). Assaying cell death of MCF7 cells infected with lentivirus expressing scrambled or ZNF217 shRNA, we observed that the PI3K inhibitor GDC0941 and the AKT inhibitor MK2206 did not induce ZNF217-dependent cell death (**Figure B.6D**).

We next used an in silico screening approach to identify candidate therapeutics that inhibit growth of cancer cells expressing high ZNF217 at low drug concentrations. We determined ZNF217 expression by rt-qPCR in the NCI60 panel of cell lines. We then used a drug dataset from a panel of ~50,000 drugs generated by the NCI Developmental Therapeutics Program (dtp.nci.nih.gov). Correlation of ZNF217 expression in the cell line panel with the

drug panel identified 15 drugs that selectively inhibited growth of cells expressing high levels of ZNF217, as assessed by GI50, with a low drug concentration (**Table B.S5**).

To determine if ZNF217 contributes to the drug-induced growth inhibition, we assayed cell death of MCF7 cells infected with lentivirus expressing stably integrated scrambled or ZNF217 shRNA knockdown constructs. As proof-of-concept, we tested bisacodyl, triciribine, nogalamycin and 2E3E for their ability to influence cell death in culture in a ZNF217-dependent manner. Cells expressing reduced levels of ZNF217 (shRNA-ZNF217) required higher concentrations of bisacodyl or triciribine for cell killing (**Figure B.6E–F**). Three different ZNF217-shRNA constructs gave similar results (**Figure B.6** and data not shown). Nogalamycin or 2E2E treatment did not promote ZNF217-dependent cell death at a therapeutically possible concentration range (data not shown).

We focused on triciribine (also known as API-2), which is a nucleoside analog and DNA synthesis inhibitor that has been tested in Phase I clinical trials in cancer patients as well as in one Phase II clinical trial in metastatic breast cancer patients (Feun et al. 1984; Schilcher et al. 1986; Hoffman et al. 1996; Garrett et al. 2011). Cancer cells expressing high levels of ZNF217 required lower concentrations of triciribine to inhibit growth than cells with low ZNF217 expression (**Figure B.6G–H**). We then assayed triciribine on a panel of breast cancer cell lines that we previously analyzed for gene expression (Neve et al. 2006). The GI50s significantly correlated with ZNF217 expression levels ($r=-0.39$, $p=0.035$ Spearman) (**Figure B.6I**), consistent with a selective effect of triciribine

on cell lines that express the highest ZNF217 levels.

Triciribine Kills Cells that Overexpress Znf217 in vivo

To test the effects of triciribine in vivo, we transplanted vector control and Znf217-overexpressing tumorigenic Vo-PyMT cells orthotopically to contralateral mammary fat pads cleared of epithelium. At three weeks, we injected mice with either triciribine or vehicle solution for five days per week. Triciribine treatment significantly reduced to baseline levels the increase in tumor burden seen as a result of Znf217 overexpression, and led to reduced phospho-AKT expression, reduced phospho-MAPK expression and increased cell death in vivo (**Figures B.7A–C, B.S4G**). In culture, triciribine inhibited only phospho-AKT and did not inhibit phospho-MAPK or ErbB3 activation after heregulin stimulation (**Figure B.7D**). The observed differences between signaling events inhibited by triciribine in culture versus in vivo suggest that cells within a tumor microenvironment respond differently to triciribine than do cells in culture.

Triciribine also was effective in vivo at inhibiting tumor growth in mice xenografted with the human tumorigenic cell line MCF7 compared to control treated mice (**Figure B.7E**).

Triciribine Overcomes ZNF217-induced Doxorubicin Resistance

Tumor cells overexpressing ZNF217 are resistant to doxorubicin-induced cell death (Huang et al. 2005). Several groups have found triciribine to be an effective, synergistic therapy in combination with other drugs (e.g., trastuzumab,

farnesyltransferase inhibitors) to reduce tumor burden (Lu et al. 2007; Balasis et al. 2011). Similarly, we found that the addition of triciribine with doxorubicin to cells in culture generated a synthetic lethality in which cells overexpressing ZNF217 were no longer resistant to doxorubicin and instead were killed (**Figure B.7F**). Interestingly, the parent mammary epithelial cell line HBL100 expresses low levels of adenosine kinase, which is required for the phosphorylation and activation of triciribine in patients. This suggests that ZNF217 may be a sufficiently predictive biomarker of triciribine efficacy, even in patients with low adenosine kinase, if patients are also treated in combination with a drug such as doxorubicin.

DISCUSSION

ZNF217 is a Biomarker of Disease Progression and Treatment Response and is a Therapeutic Target Inhibited by Triciribine

In this study we identified ZNF217 as a prognostic biomarker of reduced survival, metastasis and chemoresistance in breast cancer patients. Using both cultured cells and in vivo mouse transplant models, we found that ZNF217 overexpression contributes to multiple aspects of carcinogenesis including increased proliferation/decreased cell death, increased invasiveness, increased motility, immortalization, chemotherapy resistance, metastasis and progenitor cell expansion. Our data demonstrate that Znf217 promotes carcinogenesis by driving a differentiation gene expression signature towards a less differentiated/progenitor state indicative of expanding a multipotent progenitor population.

We identified a panel of drugs that inhibit the growth of cell lines that overexpress ZNF217 and validated two that induced ZNF217-dependent cell death.

Triciribine: a Component of Therapy for Poor Prognostic Breast Cancer Patients

Triciribine is a nucleoside analog and DNA synthesis inhibitor that has been tested in Phase I clinical trials in cancer patients as well as in one Phase II clinical trial in metastatic breast cancer patients (Feun et al. 1984; Schilcher et al. 1986; Hoffman et al. 1996; Garrett et al. 2011). In the Phase II study, one of 14

patients had stable disease and the other patients progressed (Hoffman et al. 1996). The Phase II studies and subsequent studies found that triciribine was not readily bioactive in all patients, possibly due to the requirement for expression of multiple genes for triciribine bioactivation. Triciribine is also an allosteric inhibitor of AKT activation: it physically interacts with AKT to prevent AKT recruitment to the plasma membrane and to block the phosphorylation and activation of AKT (Yang et al. 2004b; Berndt et al. 2010).

While several studies identified triciribine as an AKT inhibitor, triciribine is not always redundant with other AKT pathway inhibitors in treating tumors: often triciribine is more effective in vivo than other PI3K/AKT pathway inhibitors at inhibiting tumor progression (Lu et al. 2007). Indeed, we found that triciribine inhibited not only AKT activation but also MAPK activation in vivo (**Figures B.7B–C**). We hypothesize that triciribine inhibits both AKT and MAPK pathways, both of which are downstream of ErbB3/ErbB2 activation. This could provide a rationale for inhibition of ZNF217-induced tumor burden by triciribine, since ZNF217 drives the overexpression of ErbB3 and leads to the activation of both AKT and MAPK pathways. Alternatively, triciribine could inhibit these pathways after activation of other receptor tyrosine kinases (RTKs) (**Figure B.7G**).

Our study suggests that ZNF217 may be a sufficiently predictive biomarker of triciribine efficacy if patients are also treated in combination with a drug such as doxorubicin or another drug that offers synergy with triciribine. In part, ZNF217 may act by inducing upregulation of its target ERBB3 (Krig et al.). Thus, cells resistant to triciribine treatment might independently activate multiple

signaling pathways, making them less responsive to inhibitors that act upstream in the signaling pathway.

Combinatorial pathway activation may be therapeutically important in treating patients with high ZNF217 expression, as concurrent activation of the PI3K/AKT and RAS/MAPK pathways causes resistance to AKT inhibition in cells (She et al.). Interestingly, in the panel of immortal cell lines expressing ZNF217 and tested for triciribine sensitivity, all outlier cell lines (i.e., lines with high GI50s and high ZNF217 expression) contained previously identified mutations in the PI3K/AKT and/or RAS/MAPK pathways (She et al. 2008; Hoeflich et al. 2009; Hollestelle et al.). Since triciribine does not inhibit upstream activators PI3K or PDK1 or related family members directly (Yang et al. 2004b), future studies will be required to sort out mechanistically how ZNF217 activates and triciribine inhibits signaling. Whether combination therapies will be more effective in vivo remains to be tested.

ZNF217 Reprograms Tumor Cells to Express Luminal and Myoepithelial Cell Markers

We found that Znf217 promotes phenotypes suggestive of expansion of progenitor cells in vivo and in culture and drives repression of an adult stem cell gene expression signature that is also downregulated in many epithelial cancers. Consistent with a progenitor phenotype, ZNF217 promotes increased telomerase, resistance to TGF β growth inhibition, amplified c-MYC (Nonet et al. 2001; Li et al. 2007), as well as chemotherapy resistance (Huang et al. 2005).

That Znf217 may drive a less differentiated gene expression signature is supported by the observation that Znf217 is upregulated in the somite following the transition from the presomitic mesoderm and prior to the differentiation into the skeleton, muscle and dermis (Buttitta et al. 2003). Moreover, Znf217 is repressed concurrently with Oct4 following differentiation of a teratocarcinoma cell line to neuronal cells and binds to the promoters of a number of genes involved in differentiation (Krig et al. 2007). Thus, in tumors ZNF217 may promote transdifferentiation to or expansion of a pool of progenitor-like cells by aberrantly suppressing differentiation pathways.

Znf217 overexpression in tumor cells derived from mice expressing the oncogene PyMT switched their phenotype from a predominantly luminal to a more heterogeneous pathology characterized by expression of both luminal and myoepithelial epithelial cell markers. This phenotype is similar to that seen following Wnt1 overexpression or activation of the AKT pathway by PTEN deletion in vivo (Li et al. 2003; Korkaya et al. 2009). Interestingly, the PyMT oncogene can give rise to tumors expressing both luminal and myoepithelial markers, depending on the cell type into which it is introduced. Expression of PyMT by intraductal injection of avian retrovirus (RCAS-PyMT) induces tumors with markers of luminal, myoepithelial and progenitor cells (Du et al. 2006). Recently a connection has emerged between the undifferentiated, stem cell-like phenotype in breast cancer cells and transdifferentiation of the tumor cells towards a mesenchymal phenotype [reviewed in (Polyak and Weinberg 2009; Thiery et al. 2009; Blick et al.)]. Induction of epithelial-to-mesenchymal transition

(EMT) in cultured MECs not only increases the population of cells with mesenchymal markers but also increases those with progenitor cell characteristics (CD44^{high}/CD24^{low}) (Mani et al. 2008).

CONCLUSION

We used an integrated biological approach to model the multiple contributions of ZNF217 to carcinogenesis during tumor progression, metastasis and neoadjuvant treatment. We propose that ZNF217 is a biomarker that is prognostic of disease progression and is a therapeutic target. Our data suggest that triciribine may be a component of an effective treatment strategy in patients who have tumors expressing high ZNF217, possibly by targeting a progenitor population and reducing signaling in the AKT and MAPK pathways. Since ZNF217 is amplified in numerous cancers, this work has implications for other cancers as well.

MATERIALS AND METHODS

Cell Lines

Cell lines used in this study include mouse mammary epithelial cells NMuMG (source: Rik Derynck), SCp2 (source: Mina Bissell), EpH4 (source: Mina Bissell) and Vo-PyMT-Luc (source: Conor Lynch) and human mammary epithelial cell lines MCF7, ZR-75-1 and HBL100 (source: ATCC for all three lines). MCF7 was authenticated by SNP6.0 copy number analysis. Other cell lines were not authenticated.

Mouse Lines

Mice used in this study were maintained on the FVB/n background under pathogen-free conditions in the UCSF barrier facility. Our animal protocols were reviewed and approved by the UCSF Institution Animal Care and Use Committee.

Metastasis analysis

Both PyMT MEC and Vo-PyMT transplants were analyzed for lung metastasis. To determine metastasis frequency, lung tissue blocks were sectioned into 5-mm sections and stained by hematoxylin and eosin (H&E). For each mouse analyzed, one section was scored for number of metastases seen at 100x magnification in 3 (PyMT MECs) or 5 (Vo-PyMT cell line) high-powered fields in regions of the tissue section with the highest density of metastases. Each cohort had 6-11 mice analyzed.

Statistical analysis

Statistical analysis was performed using Prism 4 software (GraphPad Software, Inc.) or SPSS Statistics software (IBM) for Cox proportional hazard tests. Cohorts of three or more samples were compared using one-way analysis of variance (ANOVA). All tests used and *P* values are specified in the figure legends. $P < 0.05$ was considered significant.

Human Microarray Datasets

Survival data based on ZNF217 expression levels was determined using our previously published dataset (Chin et al. 2006), published studies (Sorlie et al. 2001; Sorlie et al. 2003; Sorlie et al. 2006). The meta-analysis was completed using previously published studies (Miller et al. 2005; Wang et al. 2005; Ivshina et al. 2006; Sotiriou et al. 2006; Desmedt et al. 2007; Loi et al. 2007; Schmidt et al. 2008; Zhang et al. 2009; Symmans et al. 2010). Metastasis-free survival was determined using the dataset from (Minn et al. 2005). Patients with ZNF217 expression data were separated into high and low ZNF217 expression. Chemotherapy resistance was determined using datasets compiled from (Sorlie et al. 2001; Sorlie et al. 2003; Hess et al. 2006; Sorlie et al. 2006). Patients were sorted by response to chemotherapy and by ZNF217 expression levels. If the tumor became smaller following neoadjuvant treatment, then patients were considered responsive and otherwise nonresponsive.

Transplants

Primary cells from wildtype mammary glands and MMTV-PyMT tumors were

isolated as described previously (Welm et al. 2008). In two experiments these cells were infected with lentivirus overexpressing Znf217 or vector (pEiT) alone upstream of an IRES-Tomato red fluorescent reporter gene. After one month of expression these cells were sorted by flow cytometry for collection of fluorescent cells. Cells resuspended in 10 μ l PBS were injected orthotopically on contralateral mammary fat pads cleared of epithelium on the same day following FACS sorting. 1×10^4 cells were injected per gland in the first experiment and 1.7×10^4 cells injected in the second experiment with both experiments yielding similar results. Vo-PyMT transplants: For transplants using the Vo-PyMT cells described below for triciribine treatment 1×10^5 cells were injected orthotopically in mammary glands of adult female mice.

Triciribine Treatment in vivo

Vo-PyMT cells were infected with lentivirus Znf217 or vector (pEiT) alone carrying a Tomato red fluorescent reporter gene (MOI=3) (Halpern et al. 2006). Within one week of expression these cells were sorted by flow cytometry for collection of fluorescent cells. The sorted cells were expanded and injected (1×10^6 cells in 15 μ l) orthotopically on contralateral mammary fat pads cleared of epithelium in two separate experiments. For both experiments, tumors formed by two weeks post-transplant. For one experiment, at three weeks post-transplant mice were injected with triciribine or DMSO solution for five days per week at a concentration of 1 mg/kg/day until 42 days post-transplant. (Note: In clinical trials in cancer patients, the dosage range was 20-48 mg/m²) (Schilcher et al. 1986;

Hoffman et al. 1996). This dosage is equivalent to 6.2-16 mg/kg in mouse, using an FDA-recommended conversion formula described previously (Reagan-Shaw et al. 2008). This is considerably higher (~6.2-16 times more) than the amount used in our studies. Five mice with contralateral transplants were included in the DMSO cohort, and eight mice were included in the triciribine cohort. In a second experiment, beginning at 10 days post-transplant, mice were injected i.p. daily for two weeks with triciribine or DMSO solution with seven mice per cohort and collected at 28 days post-transplant. Tissues were processed and scored for cell death by morphology by H&E. In both experiments, tumor burden was measured by a caliper.

For human xenograft experiments, 1×10^7 MCF7-M1 cells/0.1 ml were injected subcutaneously in nude (nu/nu) mice in five cohorts with five mice per cohort for a total of 40 mice. Estradiol pellets (Innovative Research, 0.72 μ g, 60 day release) were implanted at the time of transplant. The target starting volume was 200 mm⁻³. Doxorubicin was used at a dosage of 5 mg/kg once per week for three weeks alone or in combination with triciribine. A control (no drug) or triciribine dosage (50 mg/kg) were given bid for five days. These xenografts were completed in the Preclinical Therapeutics Core, UCSF Helen Diller Family Comprehensive Cancer Center.

RNA Isolation and qPCR

Tissues used for RNA isolation first were crushed in liquid nitrogen using a mortar and pestle and then resuspended in RNA-Bee (Tel-Test). RNA for tissue

and cell lines was isolated following manufacturer's instructions and both quantified and tested for purity using a NanoDrop 1000 spectrophotometer (NanoDrop Technologies, www.nanodrop.com). Reactions were assayed on an Eppendorf Mastercycler EP Realplex machine.

Supplementary Data

Additional descriptions of materials and methods, including cell lines, antibodies, and staining procedures can be found online with this article. In addition, Supplementary Data includes four Supplemental Figures, one movie and five Supplemental Tables.

Accession Numbers

Microarray data were deposited to the NCBI's GEO Repository and are accessible to reviewers through GEO series accession number GSE24727.

ACKNOWLEDGEMENTS

The authors thank members of the Werb laboratory, Lewis Cantley and Gordon Mills for helpful discussions throughout this project. We thank Ying Yu for genotyping, Elena Atamaniuc for tumor measurements and Jimmy Hwang for statistical advice. The authors disclose no conflicts of interest.

Grant Support

This work was supported by grants from the National Institutes of Health (CA129523 and CA129523-02S1 to Z.W., CA058207 to J.W.G. and Z.W., and ES019458 to Z.W. and P.Y.), a Stand Up to Cancer-American Association for Cancer Research Dream Team Translational Cancer Research Grant SU2C-AACR-DT0409 (to Z.W. and J.W.G.), an American Cancer Society Postdoctoral fellowship and a Ruth L. Kirschstein National Research Service Award CA103534 (to L.E.L.), a Canadian Institutes of Health Research Postdoctoral fellowship (to O.L.G.) and a Career Development Award from Bay Area Breast Specialized Programs of Research Excellence CA058207 (to M.D.S.).

FIGURE LEGENDS

Figure B.1. ZNF217 overexpression is a prognostic indicator in breast cancer patients. (A) Patients (n=118) were separated into high (n=59) vs. low (n=59) ZNF217 expression and analyzed for overall survival ($P=0.003$; Logrank). (B) Relapse-free survival based on high (n=40) vs. low (n=41) ZNF217 expression ($P=0.01$; Logrank). (C) Patients (n=858) were separated into low (n=286), intermediate (n=286) and high (n=286) ZNF217 expression and analyzed for relapse-free survival. Patients with high ZNF217 expression had worse survival than low ZNF217 patients ($P=0.03$; Logrank).

Figure B.2. Znf217 overexpression promotes increased cell motility and aberrant epithelial marker expression. (A) Relative Znf217 expression levels by rt-qPCR in SCp2 mammary epithelial cell lines infected with virus to overexpress vector or Znf217 with comparable results in three experiments. Each sample was tested by rt-qPCR in triplicate relative to the reference TBP, with similar results for other reference genes. Graphs show the mean \pm S.E.M. (B) Western blot analysis of Znf217 protein (anti-Znf217) and loading control (anti-HDAC1) in SCp2 cells. Images are representative of multiple experiments using retrovirus or lentivirus overexpression of Znf217. Arrows mark the indicated proteins. (C) Brightfield images of SCp2 cells \pm Znf217 display increased cell scattering in culture after Znf217 overexpression. (D) Frames from movies of SCp2 cells infected with vector or Znf217 following a scratch with a pipette tip. The movies ran 20.25 hours. Note the lamellipodia (arrow) extending from the cells by 5.5 hours and

the increased number of Znf217-expressing cells in the middle of the scratch by 10.5 hours (arrow). (E) Phalloidin staining of SCp2 cells \pm Znf217. (F) Relative expression of Znf217 and selected genes by rt-qPCR from NMuMG (top) and SCp2 (bottom) cells \pm Znf217 in vitro. Graph shows the mean \pm S.E.M., relative to the reference GAPDH. Similar results were seen with the reference HPRT. For each gene, samples for Znf217 were compared to vector by Mann-Whitney tests, and significant P values <0.02 were marked with *. (G) Heat map of selected genes enriched following Znf217 overexpression in SCp2 cells.

Figure B.3. Znf217 overexpression causes an increase in soft agar colonies and in mammosphere formation. (A) Western blot analysis of Znf217 protein in NIH3T3 cells infected with vector or Znf217 retrovirus. (B) Znf217 overexpression increases the number of colonies by anchorage-independent growth in soft agar assay. Relative number of colonies per well by soft agar with vector or Znf217 overexpression ($P=0.001$; Student's t test). Graph compiles results from three experiments, each done in triplicate. (C) Brightfield images of anchorage-independent colonies from soft agar assay \pm Znf217 in (B). Arrows mark examples of colonies. The large colonies were only seen with the Znf217-overexpressing cells, while much smaller colonies were seen with vector-expressing cells. (D) Relative expression of Znf217 by rt-qPCR from normal adult mammary gland (FVB/n), relative to the reference HPRT with line marking the mean. Glands were sorted by flow cytometry for CD24^{Med}CD49f^{High} (basal/myoepithelial/progenitor cells) and CD24^{High}CD49f^{Low} (luminal/luminal

progenitor) fractions. RNA was isolated and used to generate cDNA from each population. Each dot represents one mouse sorted, collected and processed by rt-qPCR. Graph shows relative epithelial Znf217 expression in the CD24^{Med}CD49f^{High} vs. CD24^{High}CD49f^{Low} populations. Similar results were seen with the reference GAPDH. (E) Relative expression of Znf217 expression by rt-qPCR in primary mouse mammary epithelial cells following lentiviral infection with either pEiT vector or Znf217-pEiT in three separate samples. Each sample was tested by qPCR in triplicate relative to the reference TBP. These samples were used for microarray analysis. Graph shows the mean \pm S.E.M. (F) Quantification and (G) brightfield images of mammosphere assay of Vo-PyMT cells overexpressing vector or Znf217. (H) Heat map of selected genes from gene expression microarray analysis enriched in primary MECs overexpressing Znf217. (I) Relative expression of Znf217 expression by rt-qPCR in primary mouse mammary epithelial cells following lentiviral infection with either vector or Znf217 in three separate samples. Each sample was tested by qPCR in triplicate relative to the reference TBP and used for microarray analysis. Graph shows the mean \pm S.E.M. (J) Rt-qPCR to validate microarray targets using the same samples used in (H) with HPRT as a reference in qPCR reactions. Similar results were obtained with GAPDH used as a reference (data not shown).

Figure B.4. Znf217 overexpression in vivo increases rate of tumor progression, tumor heterogeneity and differentiation state. (A) Relative expression of Znf217 and EMT genes by RT-qPCR in the Vo-PyMT cell line overexpressing either

vector or Znf217. Assay used the reference GAPDH. Similar results were seen using HPRT or TBP references. The cells used in this experiment had previously been sorted for fluorescent marker expression and were used for the Vo-PyMT transplants throughout this study. (B) Mammosphere assay of primary MECs infected with vector or Znf217-overexpressing lentivirus. (C) Quantification of mammosphere formation in primary MECs expressing vector or Znf217 after one week. Graph shows mean \pm S.E.M., and samples were compared by unpaired t test. (D) Tumor-free survival over time in Vo-PyMT transplants ($P=0.01$; Logrank). (E) Tumor volume over time in Vo-PyMT transplants of Znf217 ($n=8$) versus vector ($n=10$) ($P=0.007$; ANOVA, repeated measures). (F) Final tumor weight in Vo-PyMT transplants ($P=0.02$; Mann-Whitney). Line represents median of vector ($n=9$) versus Znf217 ($n=8$). (G) H&E staining of MMTV-PyMT (PyMT MEC) tumors from transplants overexpressing vector (top) or Znf217 (center, bottom panels). Inserts are enlarged images of boxed regions and demonstrate heterogeneous pathology. (H) Immunofluorescence staining with anti-Keratin-8 (green), anti-alpha-smooth muscle actin (red)(arrows) and DNA (Hoechst; blue) in tumors derived from PyMT MEC transplants. (I) Immunofluorescence staining with anti-Keratin-8 (green), Keratin-14 (red) and DNA (blue) from PyMT MEC transplants. Arrows mark cells double positive for K8 and K14. (J) Quantification of a progenitor cell population: K8+K14+ ($P=0.002$), % K8+K14+ ($P=0.0002$)(unpaired t-tests). Bar graphs show mean representation (number/percentage) of K8+K14+ cells \pm S.E.M. per HPF. HPF= 3 high-powered fields.

Figure B.5. Znf217 overexpression in vivo increases lung metastasis. (A) Immunofluorescence with anti-E-cadherin (green) and DNA (blue) from Vo-PyMT transplants. Arrows mark regions with low E-cadherin expression. (B) Number of lung metastases per three (a) or five (b) high powered fields from (a) PyMT MEC ($P=0.008$) or (b) Vo-PyMT transplants ($P=0.01$; Mann-Whitney) with vector or Znf217 overexpression. Bar graph shows the mean \pm S.E.M. (C) Metastatic burden from PyMT and Vo-PyMT transplants. Number of lung metastases per three (PyMT) or five (Vo-PyMT) high-powered fields divided by tumor weight from (a) PyMT MEC ($P=0.003$; Mann-Whitney) or (b) Vo-PyMT ($P=0.32$; Mann-Whitney) transplants with vector or Znf217 overexpression. Bar graph shows the mean \pm S.E.M. Similar results were obtained using final tumor volume (data not shown). (D) H&E staining of lung metastases from PyMT MEC transplants. Arrows mark examples of metastases. (E) Metastasis-free survival based on high ($n=41$) vs. low ($n=41$) ZNF217 expression ($P=0.01$; Logrank) from (Minn et al. 2005).

Figure B.6. Identification of triciribine as a candidate inhibitor of ZNF217-induced growth. (A) Response to neoadjuvant chemotherapy in breast cancer patients with high vs. low ZNF217 expression in tumors (a) from (Sorlie et al. 2006). Patients had responsive ($n=27$) or nonresponsive (stable/progressive) disease ($n=28$) in response to treatment ($P=0.01$; Mann-Whitney). Lines mark means. (b) from (Hess et al. 2006) ($P<0.001$; Mann-Whitney). Tumors were responsive

(pathological complete response; n=34) or nonresponsive (residual disease; n=34) to treatment. Lines mark means. (B) ZNF217 and ERBB3 expression levels in human breast tumors (n=118) from (Chin et al. 2006). ZNF217 and ERBB3 strongly correlate (Pearson $r=0.47$; $R^2=0.22$; by linear regression, $P<0.0001$). (C) MCF7 cells (left) or ZR-75-1 cells (right) were transiently transfected with scrambled or ZNF217 siRNA. 48 hours after transfection, cells were serum starved 24 hours and treated for 15 minutes with heregulin. Lysates were blotted for the indicated proteins. (D) PI3K and AKT inhibitors do not promote ZNF217-dependent cell death. FACS analysis of cell death by Annexin V staining in MCF7 cells \pm ZNF217-shRNA or scramble control and treated for two days with control, 2 μ M GDC0941, or 10 μ M MK2206. (E) Treatment of MCF7 cells \pm ZNF217shRNA with bisacodyl in triplicate at the indicated concentrations ($P=0.001$; ANOVA). Similar results were obtained in at least three experiments. (F) Treatment of MCF7 cells \pm ZNF217-shRNA with 10 μ M triciribine at the indicated concentrations ($P=0.001$; ANOVA). Similar results were obtained with a second shRNA and in at least three experiments. (G) ZNF217 expression levels and (H) related triciribine GI50 concentrations in NCI60 panel breast cancer cell lines. Inset: Chemical structure of triciribine. (I) ZNF217 expression levels across triciribine GI50s in 30 breast cancer cell lines (15 each of cell lines expressing highest/lowest ZNF217)($r=-0.39$; $P=0.035$; Spearman correlation). Two outliers circled are identified by cell type and relevant mutations.

Figure B.7. Triciribine inhibits Znf217 in vivo and in human cells. (A) Tumor burden growth rate of Vo-PyMT transplants treated with DMSO solution (solid lines) or triciribine (dotted line) ($P < 0.0001$ by genotype; $P = 0.02$, genotype over time; ANOVA). Vo-PyMT transplants overexpressed vector (blue) or Znf217 (orange). Shown are the mean \pm S.E.M. (B) Phospho-AKT (left) and phospho-MAPK (right) protein expression by immunohistochemistry in tissues from Vo-PyMT transplants treated with either control or triciribine. (C) Model of pathways downstream from Znf217. Znf217 overexpression promotes phospho-AKT and phospho-MAPK. This activation is associated with increased tumor burden, chemotherapy resistance and mammosphere formation. Triciribine can block these phenotypes of Znf217 overexpression. (D) MCF7 cells \pm triciribine were serum starved overnight and stimulated with heregulin/neuregulin-1 β for the indicated times. Cell lysates were blotted for the indicated proteins. (E) Human MCF7-M1 subcutaneous xenografts treated with control or triciribine (50 mg/kg) at the indicated time post-transplant. Ticks show mean tumor burden \pm standard deviation. (F) Triciribine induces synthetic lethality with doxorubicin in culture. Stable HBL100 MECs (low ZNF217, low adenosine kinase expression) (\pm ZNF217) were treated with triciribine and doxorubicin at the indicated concentrations and monitored for cell death using Annexin V staining ($P = 0.0002$; ANOVA). All doxorubicin-treated samples were statistically different ($P < 0.05$; Bonferroni posttest), whereas triciribine treatment alone did not promote statistically significant results. Graph shows mean \pm S.E.M. (G) Model of Znf217 function. Increased Znf217 promotes increased ErbB3 expression and activation

of downstream signaling events during tumor progression. Znf217 may also activate other receptor tyrosine kinases (RTKs) that in turn lead to activation of AKT or MAPK pathways. In vivo during tumor progression, triciribine can block signaling events downstream of Znf217 overexpression.

Supplemental Figure B.S1. High ZNF217 expression in breast cancer patients is prognostic of poor survival. (A) Patients (n=88) were separated into high (n=44) vs. low (n=44) ZNF217 DNA amplification and analyzed for overall survival ($P=0.004$; Logrank) using Chin dataset. (B) Disease-specific survival. Patient survival based on ZNF217 expression levels using Chin dataset. Patients (n=118) were separated into two cohorts (high vs. low) based on ZNF217 expression and analyzed for disease-specific survival ($P=0.01$; Logrank test). (C) Multivariate Cox proportional hazard analysis of ZNF217 expression, ER status and tumor size in relation to survival. N=number of patients per cohort; ER=estrogen receptor; CI=confidence interval. ZNF217 expression is given in log2 values. Tumor size is the length of the tumor's diameter. Endpoint was based on overall survival from the Chin (2006) expression dataset.

Supplemental Figure B.S2. Znf217 overexpression in mammary epithelial cells induces a scattered morphology. (A) Relative expression of Znf217 expression levels by RT-qPCR in mammary epithelial cell lines (NMuMG and EpH4) infected with virus to overexpress vector or Znf217. Results are comparable to results seen in multiple experiments. Results compile the results from 8 experiments

(NMuMG) and 1 experiment (EpH4). EpH4 results are consistent with experiments using other Znf217 viral vectors. Each sample was tested by qPCR in triplicate relative to the reference TBP, with similar results for other reference genes. Graphs show the mean and S.E.M. (B) Western blot analysis of Znf217 protein levels (anti-Znf217) and a loading control (anti-cdc2-PSTAIRE or crossreacting lower migrating band). These images are representative of multiple experiments using retrovirus or lentivirus overexpression of Znf217. Arrows mark the indicated proteins. (C) Brightfield images of mammary epithelial cells \pm Znf217 display increased cell scattering in culture. Cell lines used included NMuMG, EpH4 and MMTV-PyMT primary mammary epithelial cells. (D) Proliferation assay of MCF7 cells expressing either scrambled control shRNA or one of two ZNF217 shRNAs. Results are typical of at least three experiments. See Figure S4 for validation of knockdown expression of ZNF217 following shRNA expression.

Supplemental Figure B.S3. Znf217 alters epithelial differentiation state and promotes metastasis in vivo. (A) Tumor burden over time in tumors from transplants of primary MMTV-PyMT mammary epithelial tumor cells infected with virus overexpressing Znf217 or vector. Ticks represent the mean, and error bars represent S.E.M. ($P=0.87$; Two-way ANOVA, repeated measures). (B) Immunofluorescence with anti-E-cadherin (green), Keratin-14 (red), and DNA (blue). Arrows mark regions with membrane-localized, low expression, or cytoplasmic E-cadherin staining, respectively. (C) Metastasis rate from

transplants of primary MMTV-PyMT mammary epithelial tumor cells infected with virus overexpressing Znf217 or vector. Percentage of mice per cohort that had metastases ($P=0.01$; Fisher's Exact). Bar graph shows the mean for each cohort.

Supplemental Figure B.S4. ZNF217 sensitizes MCF7 and ZR-75-1 mammary epithelial cells to heregulin. (A) Densitometric analysis of phospho-MAPK signal from MCF7 cells in Figure 6C. (B) Densitometric analysis of phospho-AKT signal from MCF7 cells in Figure 6C. (C) Densitometric analysis of phospho-MAPK signal from ZR-75-1 cells in Figure 6C. (D) Densitometric analysis of phospho-AKT signal from ZR-75-1 cells in Figure 6C. (E) Western blot and (F) densitometry analysis using anti-ZNF217 and actin after lentivirus infection followed with the indicated scrambled or ZNF217-targeted shRNAs on MCF7 lysates. (G) Triciribine treatment of mice with tumors derived from transplants of Vo-PyMT cells infected with virus overexpressing Znf217 or vector. H&E stains of tumors in Figure 7. Arrow marks necrotic tissue.

SUPPLEMENTAL MOVIE B.S1

Movie of SCp2 cells infected with vector (left movie) or Znf217 (right movie) following a scratch with a pipette tip. This movie ran for 20.25 hours. This file can be found online with the published article.

SUPPLEMENTAL TABLE LEGENDS

All Supplemental Tables can be found online with the published article.

Supplemental Table B.S1. Human microarray summary

Summary of correlations between ZNF217 and selected genes from ChIP-chip and gene expression microarray datasets. ZNF217 expression levels correlated with expression of K19 and K8/18, genes that have ZNF217 enriched at their promoters (Krig et al. 2007), in human breast cancer cell lines, tumors, and after knockdown in MCF7 cells.

Supplemental Table B.S2. Gene ontology for microarray gene sets

List of gene ontology (GO) terms from the microarray gene expression datasets of primary MECs \pm Znf217 using DAVID software.

Supplemental Table B.S3. Microarray target genes from MCF7 siRNA

Target genes identified by microarray expression analysis from MCF7 cells \pm Znf217 siRNA.

Supplemental Table B.S4. Overlapping adult stem cell signature genes

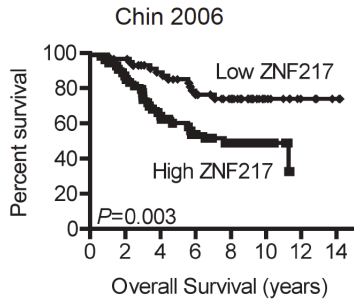
List of overlapping genes between our dataset (this study) and our adult stem cell signature, which is downregulated in many epithelial cancers relative to normal tissue (Wong et al. 2008). Primary MECs overexpressing Znf217 significantly repressed genes of the adult stem cell signature ($P=1.89 \times 10^{-10}$) (Wong et al. 2008).

Supplemental Table B.S5. Drug list identified in silico

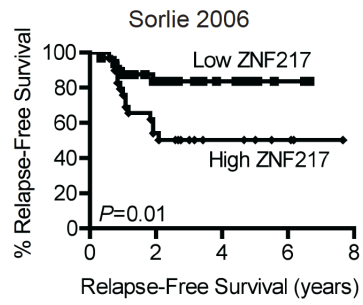
List of drugs identified in silico to inhibit breast cancer cell lines that overexpress ZNF217. Correlation of ZNF217 expression in the cell line panel with the drug panel of ~50,000 drugs generated by the NCI Developmental Therapeutics Program (ntp.nci.nih.gov) identified several drugs that selectively inhibited growth of cells, assessed by GI50, expressing high levels of ZNF217 with a low drug concentration.

FIGURES

A Overall survival



B Relapse-Free Survival



C ER+ Her2- LN- patients

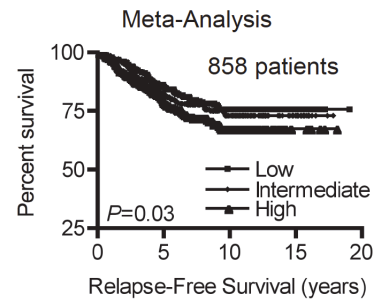


Figure B.1

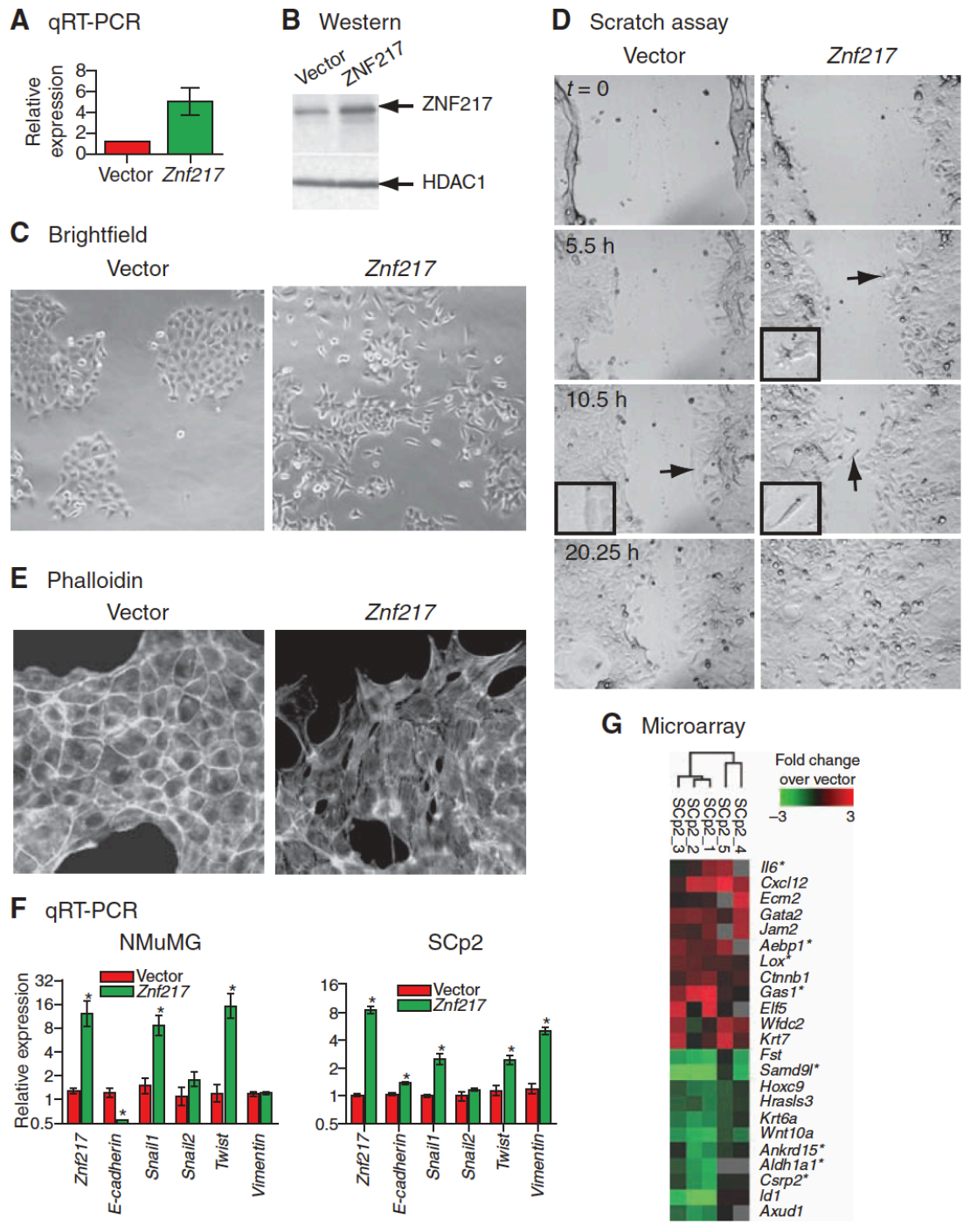


Figure B.2

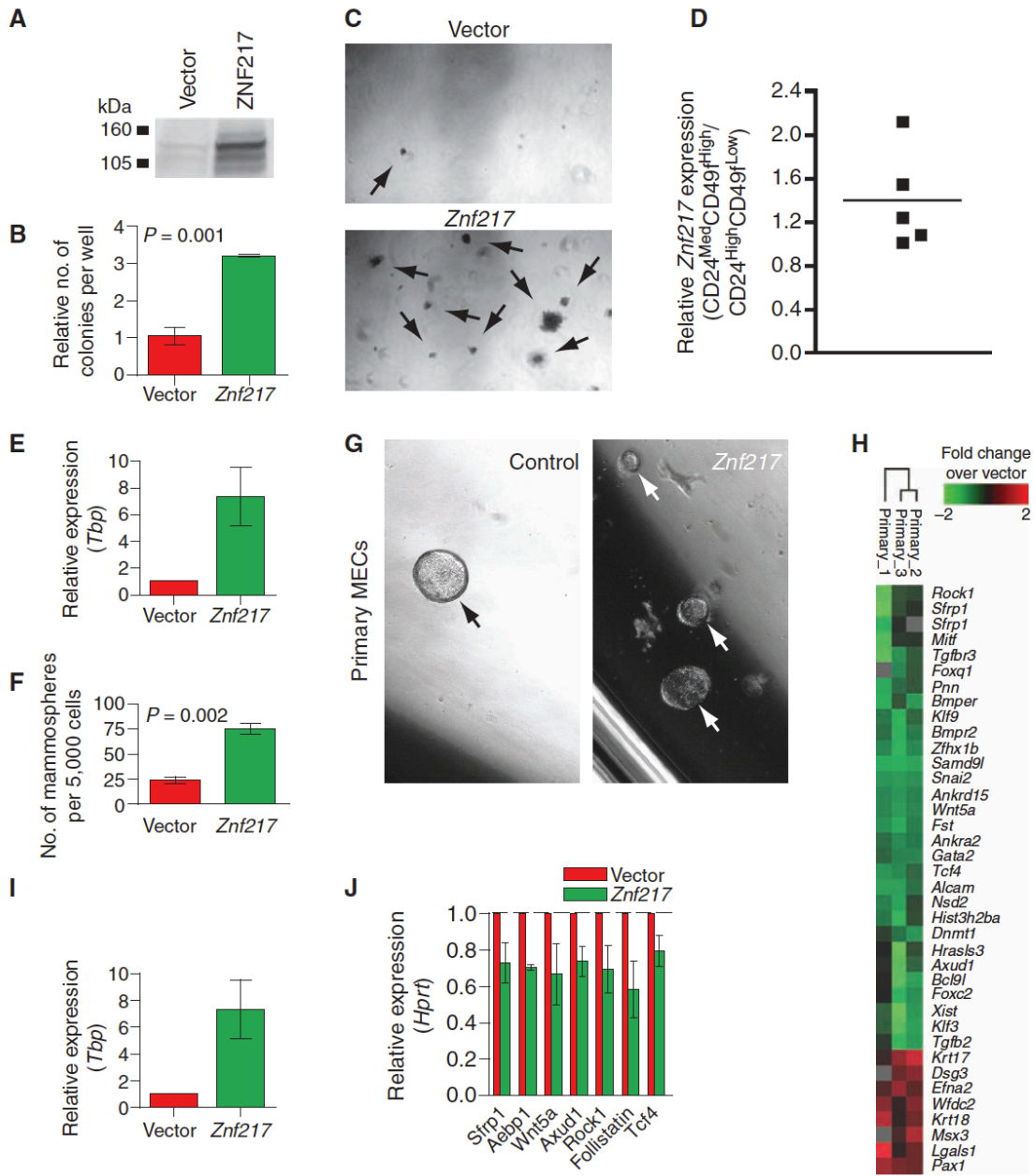


Figure B.3

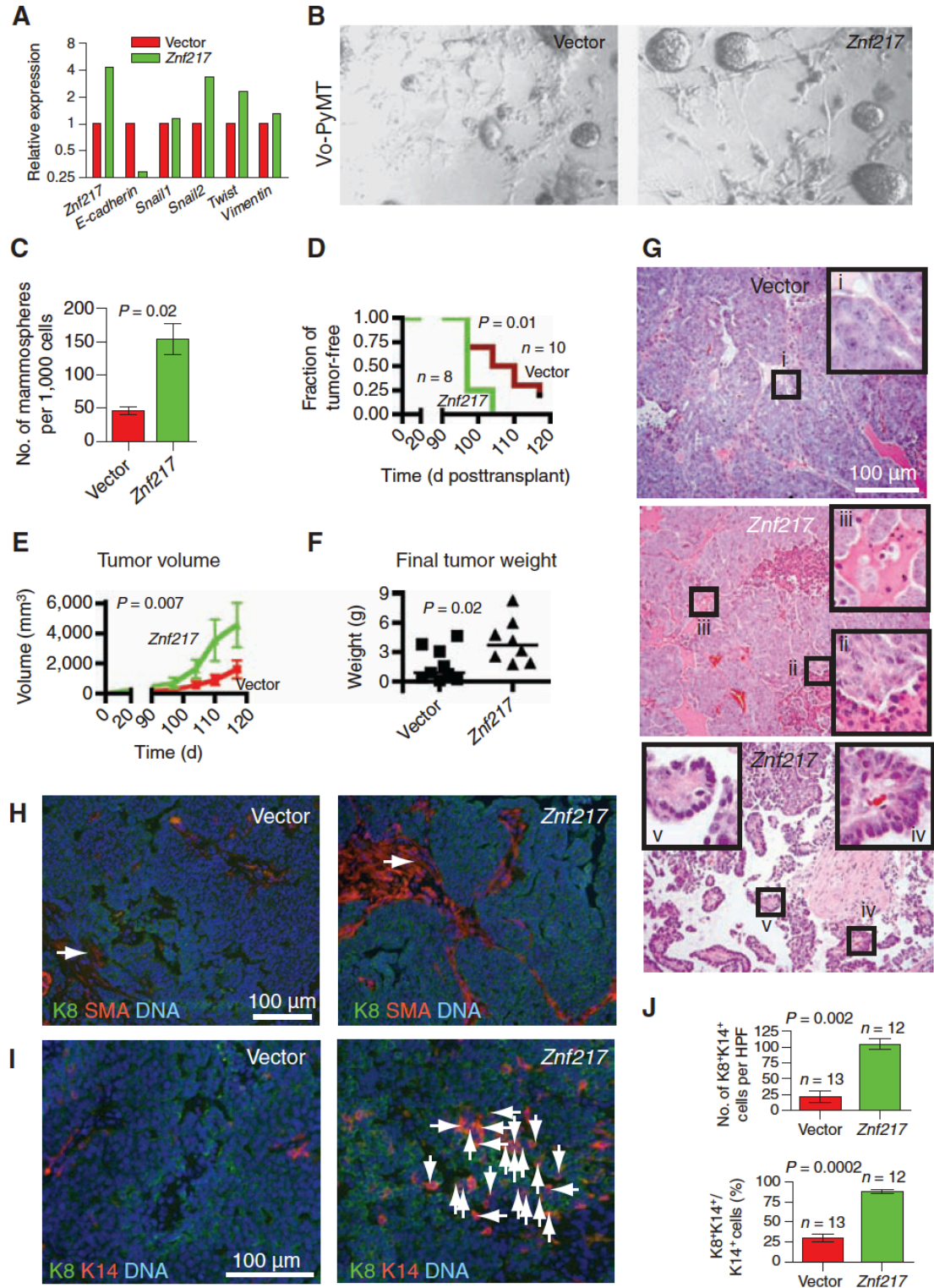


Figure B.4

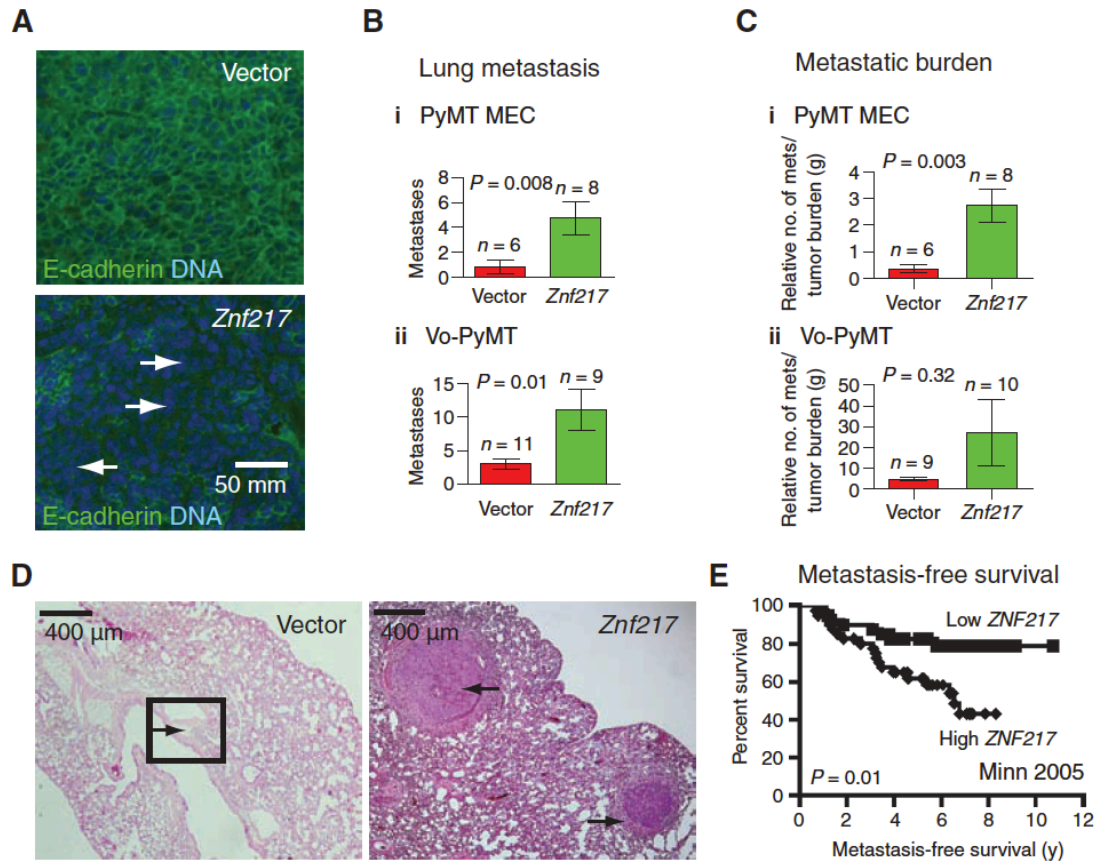


Figure B.5

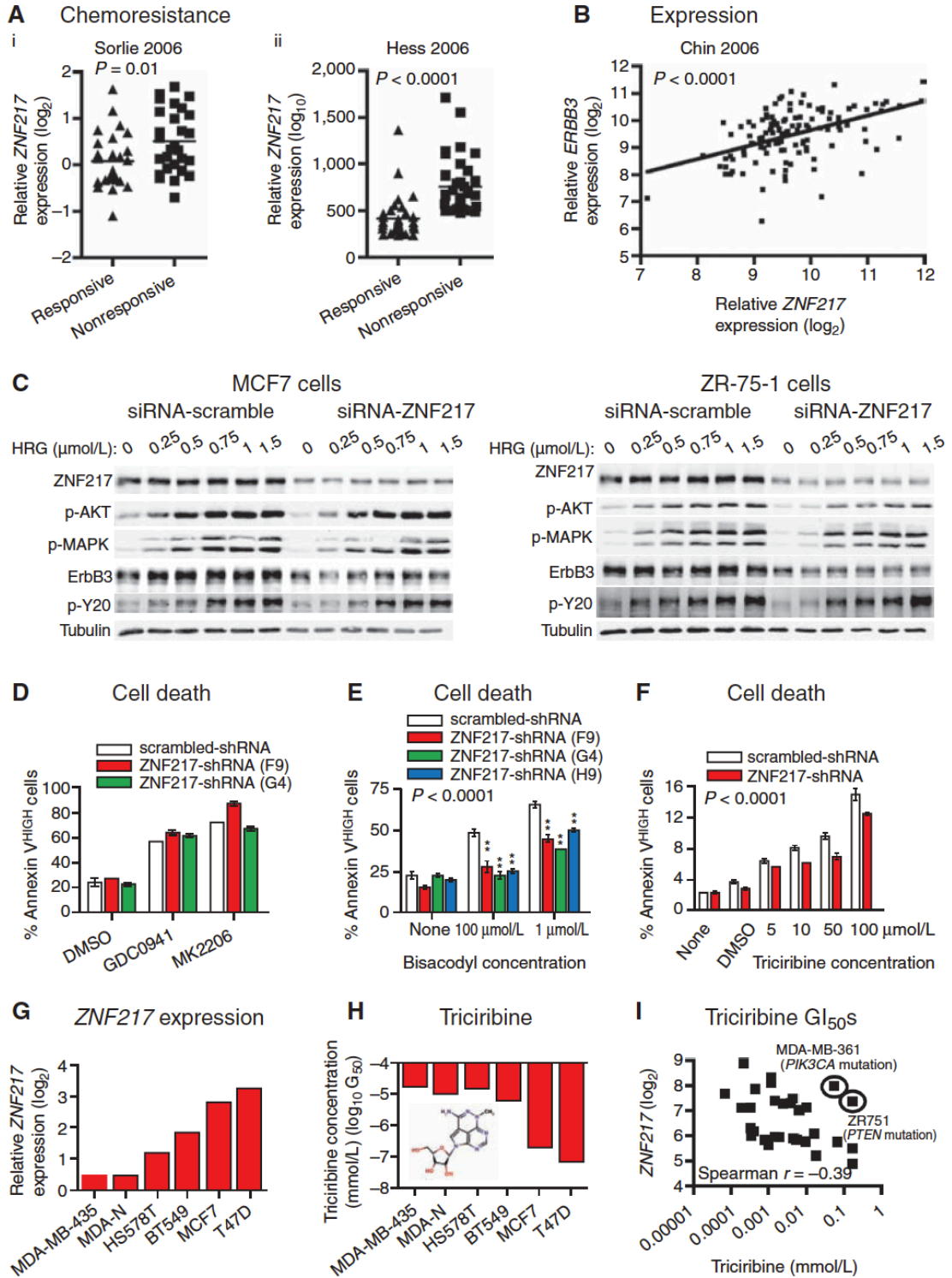


Figure B.6

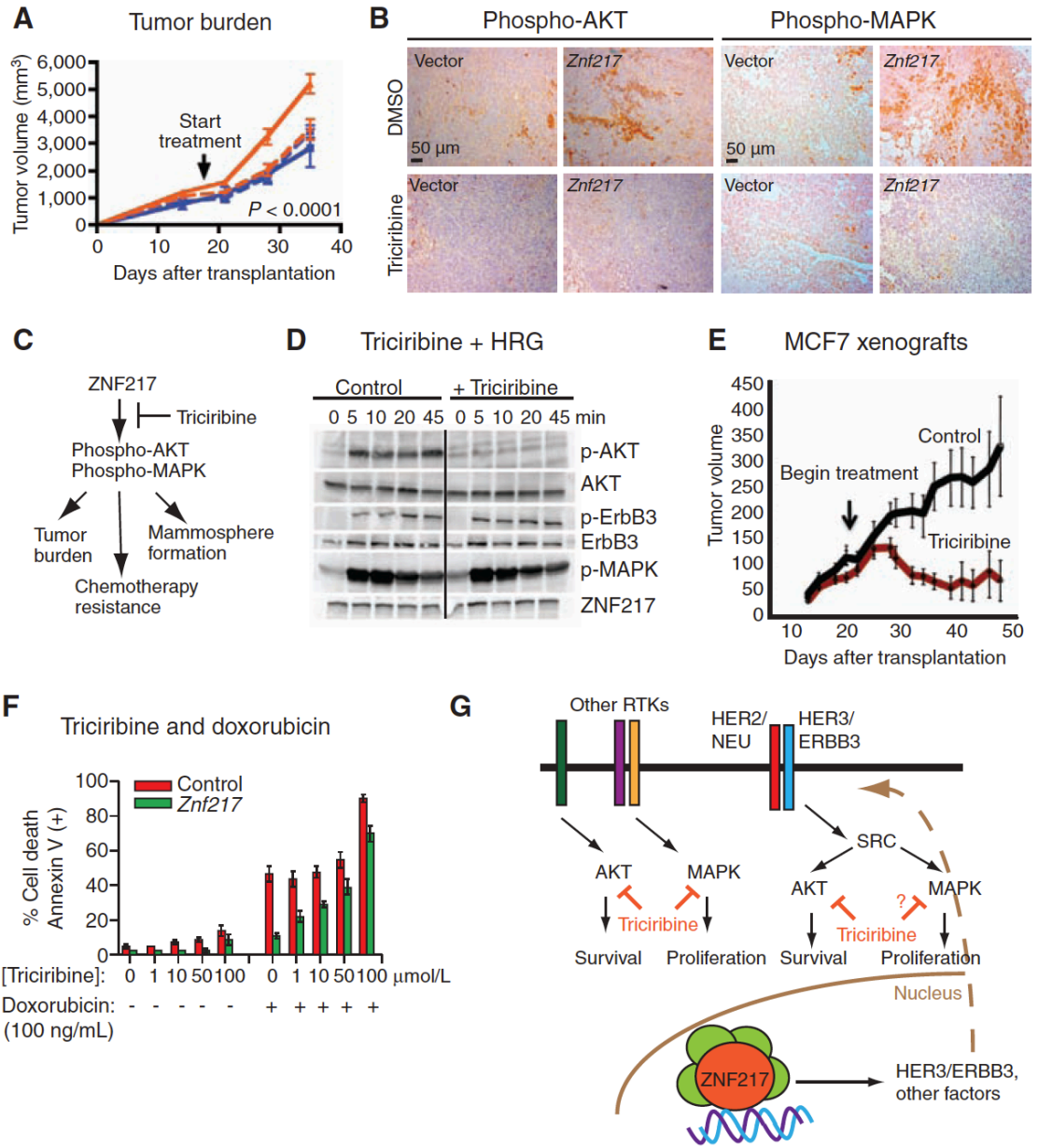
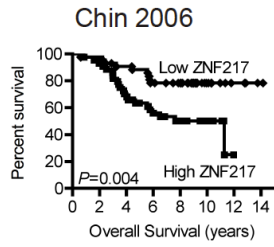
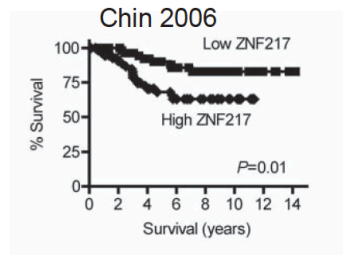


Figure B.7

A Overall survival
(DNA amplification)



B Disease-specific survival



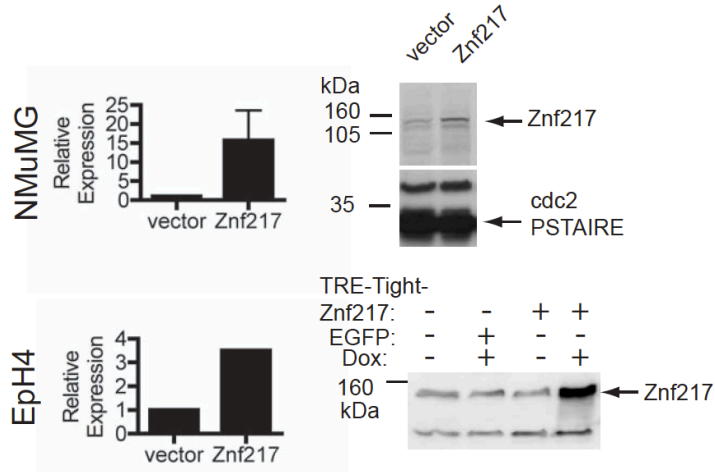
C Multivariate Cox proportional hazard analysis in relation to survival

Variable	N	B	SE	Wald	df	P	Exp(B) risk ratio	95% CI (Risk)
ER positive	74	0.23	0.33	0.49	1	0.48	1.26	0.66 to 2.40
ER negative	43							
ZNF217 Low ZNF217 (<9.5)	54	0.71	0.36	4.03	1	0.04	2.04	1.02 to 4.10
ZNF217 High ZNF217 (>9.5)	63							
Tumor Size Small (<2.1 cm diameter)	47	0.90	0.37	5.83	1	0.02	2.47	1.19 to 5.14
Tumor Size Large (>2.1 cm diameter)	68							

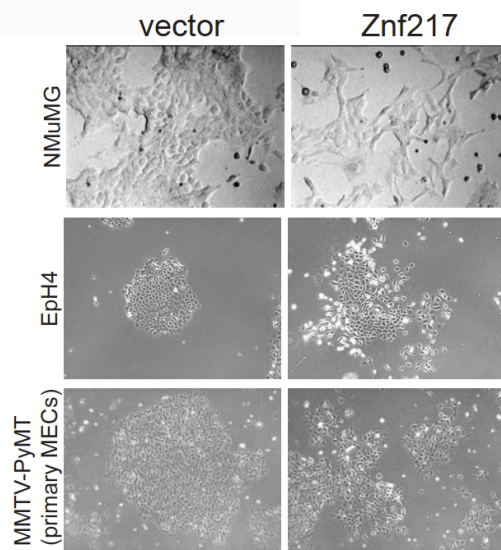
Supplemental Figure B.S1

A rt-qPCR

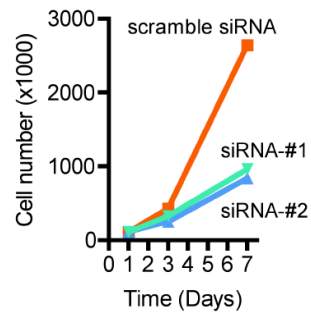
B Western analysis



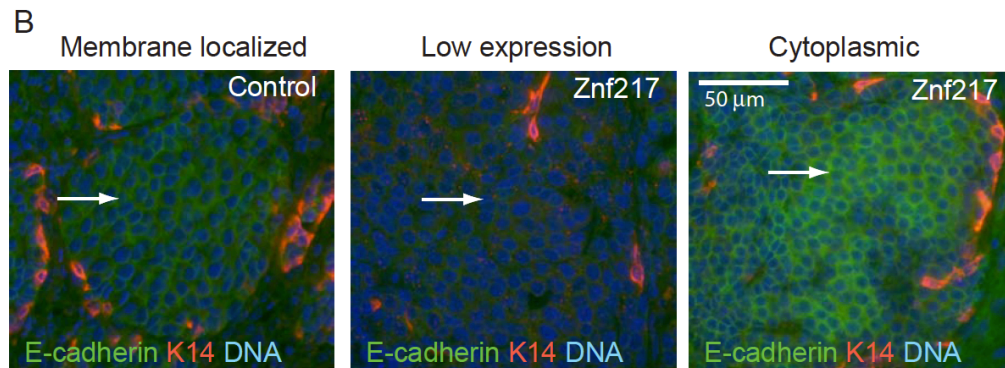
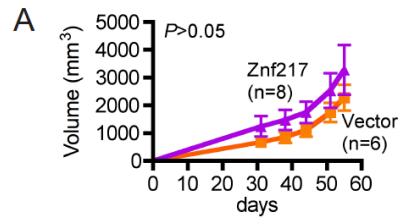
C



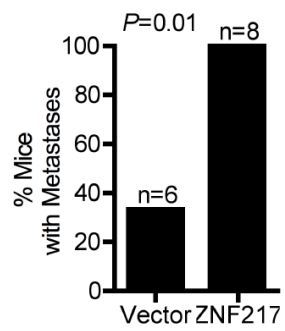
D



Supplemental Figure B.S2

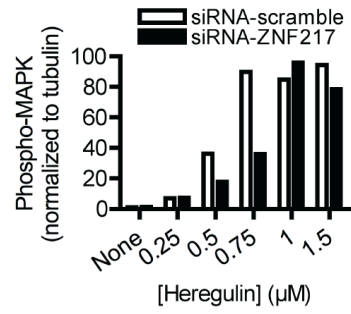


C Metastasis Rate

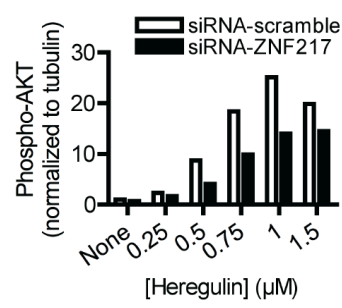


Supplemental Figure B.S3

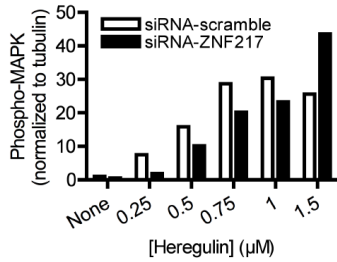
A Phospho-MAPK



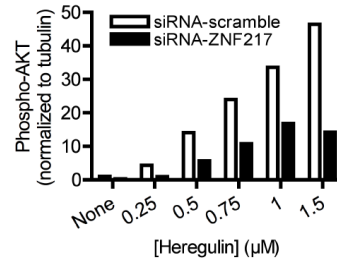
B Phospho-AKT



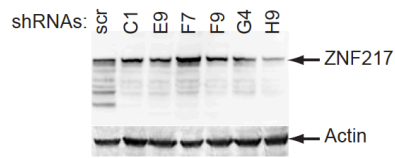
C Phospho-MAPK



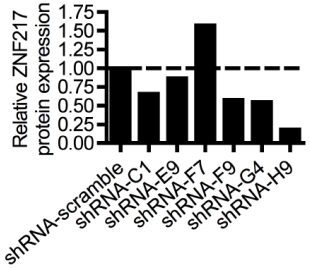
D Phospho-AKT



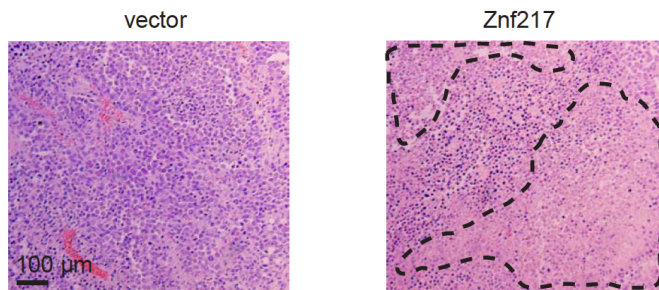
E shRNA-ZNF217



F Protein Quantification



G Cell Death after Triciribine Treatment



Supplemental Figure B.S4

Appendix C: A nanocrystal sensor for luminescence detection of cellular forces

Source: The following appendix is unpublished work currently being prepared for submission.

Contributions: Charina Choi, Paul Alivisatos, Zena Werb and I conceived the idea for this project. Charina Choi characterized and synthesized the tetrapod quantum dots, calibrated the strain gauge, and built the microscope, with help from Katie Lutker. I cultured the HL-1 cells, and assisted with the imaging experiments. Charina Choi analyzed the data. Charina Choi and I wrote the manuscript. Zena Werb and Paul Alivisatos supervised the project.

A nanocrystal sensor for luminescence detection of cellular forces

Charina Choi^{1,2}, Jonathan Chou^{3,4}, Katie Lutker^{1,2}, Zena Werb^{3,4} and A. Paul Alivisatos^{1,2,5}

¹Department of Chemistry, University of California, Berkeley CA, 94720

²Materials Science Division, Lawrence Berkeley National Laboratory, Berkeley,

CA 94720, ³Department of Anatomy, and ⁴Biomedical Sciences Program,

University of California, San Francisco CA 94143

⁵Corresponding author:

A. Paul Alivisatos, Ph.D.

Department of Chemistry and

Materials Science Division

Lawrence Berkeley National Laboratory

Berkeley, CA 94720

Email APAAlivisatos@lbl.gov

ABSTRACT

Quantum dots have been used as bright fluorescent tags with high photostability to probe numerous biological systems. In this work we present the tetrapod quantum dot as a dynamic, next-generation nanocrystal probe that fluorescently reports cellular forces with spatial and temporal resolution. Its small size and colloidal state suggest that the tetrapod may be further developed as a tool to measure cellular forces in vivo and with macromolecular spatial resolution.

INTRODUCTION

Quantum dots were first developed as luminescent biological tags more than a decade ago in the laboratories of Paul Alivisatos and Shuming Nie (Bruchez et al. 1998; Chan and Nie 1998). Quantum dots are highly fluorescent with bright and narrow emission spectra and high photostability, and their colloidal nature enables them to be incorporated into a variety of systems. In the years since their first demonstration as fluorescent biolabels, quantum dots have been used to probe numerous biological processes, including fluorescent assays for drug discovery, disease detection, single protein tracking, and intracellular reporting.

Quantum dots have five distinct properties that give them their unique capabilities. First, the dots are small, ranging from 4 to 12 nm in diameter. Second, they have size-tunable, narrow, Gaussian emission spectra that can be excited at a single wavelength, enabling multiplexed experiments. Third, they have enormous absorption extinction coefficients and high fluorescent quantum yields, making them exceptionally bright. Fourth, they are composed of inorganic semiconductor materials and are thus photochemically robust. Their resistance to photobleaching enables extended dynamic imaging. Finally, because quantum dots “blink,” the fluorescence intermittency assures the observation of a single dot event, which translates to the observation of a single protein. These properties enable fluorescence imaging experiments in biology to unravel biological function at the molecular level (Rosenthal et al. 2011).

In recent years, researchers have aimed to develop more complex nanocrystals to probe important biological questions that cannot be answered by

static fluorescent tags or other microscopic techniques. For example, cell-generated forces on the surrounding extracellular matrix play a crucial role in important processes ranging from stem cell differentiation to cancer metastasis (Engler et al. 2004; Butcher et al. 2009; Vogel and Sheetz 2009). These forces have been measured directly with spatial resolution (Chen et al. 2004) by quantifying cell-induced displacements in microfabricated arrays of fluorescent particles (Balaban et al. 2001) and elastomeric vertical cantilevers (Tan et al. 2003). Such methods have yielded important insights into cell mechanical behavior and mechanotransduction (Chen et al. 2004). However, while cellular mechanics are ultimately generated by structures at the size scale of a single protein, current techniques cannot map forces with nanoscale spatial resolution. Furthermore, current techniques are not adaptable to the geometries of many biological systems, precluding their use to make in vivo measurements.

We have recently designed a more complex luminescent nanocrystal, the tetrapod quantum dot. The CdSe/CdS tetrapod nanocrystal consists of a zinc blende CdSe quantum dot core with four tetrahedrally protruding wurtzite CdS arms. The tetrapod has dimensions of 4 nm for the core and arm diameters and 15-50 nm for the arm length. Like its spherical quantum dot predecessor, the CdSe/CdS tetrapod exhibits bright and narrow photoluminescence, with a Gaussian-shaped ensemble emission spectrum. Previous work showed that the peak emission wavelength of the tetrapod shifts as a function of applied stress (Choi et al. 2009; Choi et al. 2010) and is sensitive to perturbations on the nanonewton (nN) scale (Schrier et al. 2008). Deformation of the tetrapod due to

arm bending alters the electronic structure of the tetrapod and shifts the fluorescence emission to longer wavelength. Because cell-generated forces are also on the order of nanonewtons (Chen et al. 2004), we postulated that tetrapods might be able to sense and fluorescently report cellular mechanical stresses.

Here we demonstrate that the luminescence wavelength of a colloidal nanocrystal, the tetrapod quantum dot, shifts in response to cell-generated stresses. We designed a tetrapod monolayer array and cultured beating cardiomyocytes on top of the array. Using a home-built acousto-optic tunable filter microscope, we measured spatially resolved luminescence spectra of the array in real time and observed clear time-dependent color shifts of the tetrapod array emission due to cardiomyocyte contractions. This result creates a path towards studies of cellular forces with nanoscale resolution and geometric flexibility. Additionally, because the tetrapod responds optically to cell-generated stresses, it can also be used to convert cell-generated inputs into optical signals, which may be useful towards more complex synthetic biology systems.

RESULTS AND DISCUSSION

To demonstrate that the luminescence wavelength of the tetrapod quantum dots shift in response to cell-generated stresses, we designed a luminescent tetrapod array on which cells could be cultured (**Fig. C.1a, C.1b**), using tetrapods of 25 nm arm length. An ensemble of these tetrapods exhibits luminescence peak emission at 1.94 eV with a peak full-width-half-maximum of 0.09 eV. As-synthesized tetrapods are not water soluble, so we first made them biocompatible using an amphiphilic polymer wrapping technique (Pellegrino et al. 2004). The resultant carboxylate-functionalized tetrapods were covalently attached to an amine-functionalized transparent substrate via amide bond formation chemistry, achieving a tetrapod monolayer array (**Fig. C.1c**). We chose to culture HL-1 rat cardiomyocytes, which exhibit spontaneous and periodic contractions in vitro, on top of a tetrapod monolayer to apply periodic stresses to the luminescent array (**Fig. C.1d**) (Claycomb et al. 1998)). The morphological and contractile phenotype of the cells was similar to those grown in culture without the tetrapod monolayer.

To determine whether cell contractions could switch the luminescence emitted by the tetrapod array, we built an acousto-optic tunable filter (AOTF) microscope to collect spatially resolved luminescence spectra of the array in real time (**Fig. C.2**). The AOTF microscope (Wachman et al. 1997) images the intensity of light emitted by the sample as a function of band pass. Images taken at multiple band pass frequencies are stacked together to create a map with spectral data at each image pixel. Luminescence spectra of the tetrapod array

were collected by imaging the emission intensity of the array at 100 band pass frequencies ranging from 1.81-2.14 eV. We collected 324 spectra simultaneously in 350 ms over a total area of 18 x 18 pixels with a pixel resolution of 5.1 μm^2 . Each spectrum was fit to a Gaussian curve, and the emission peak energy at each pixel was used to create a spectral map. Twenty successive spectral maps were imaged for a given spot to observe luminescence behavior over time.

We observed clear time-dependent color shifts in the tetrapod array emission due to cardiomyocyte contractions (**Fig. C.3a, C3.b**), which were not observed in arrays without cardiomyocytes (**Fig. C.3c**). Fluorescence red shifts relative to a control array were observed, indicating a breaking of tetrahedral symmetry in the nanocrystals induced by cell contractions. Utilizing results from electronic structure modeling of uniaxially compressed tetrapods (Schrier et al. 2008), we calculate an average shift-inducing force of 0.7 ± 0.4 nN per tetrapod. This range of values is consistent with previous measurements of contractile forces from cardiomyocytes (Balaban et al. 2001). Although the force behavior of cells on a stiff, two-dimensional culture substrate likely differs from that under physiologic conditions, this geometry provides a simple system for the first biological demonstration of this tool. Importantly, the observed emission shifts demonstrate that tetrapod nanocrystals can sense cellular forces and respond with luminescent readout.

The small size and colloidal nature of the tetrapod allow its further development towards imaging cellular stresses with macromolecular spatial resolution and in a variety of biological geometries. Cellular forces are generated

by cytoskeletal proteins such as microtubules and actin filaments, and are localized to subcellular structures such as focal adhesions. Nanoscale resolution will allow quantitative investigations of how stresses are sensed and relayed at single protein complexes consisting of integrin clusters to the cytoskeleton. Furthermore, colloidal particles can be treated as a chemical reagent, enabling use of the tetrapods to study processes that occur on flat surfaces, as shown in the present work, in three-dimensional matrices, or even within tissues or between cells. The key to measurements in more complex geometries will be careful anchoring of the nanocrystal through smart surface functionalization for maximal tetrapod deformation by the stress of interest.

Because the tetrapod responds optically to cell-generated stresses, it can also be used to convert biochemical inputs into optical signals, which may be useful in a synthetic biology circuit, for example. Future work will enable fully quantitative force measurements through a detailed characterization of the tetrapod response to biologically relevant stress states and more physiologic studies of cell mechanics by attachment of tetrapods within softer extracellular matrix-like materials, such as collagen or Matrigel. These studies will provide an important complement to measurements of *in vitro* cellular stress and tension (Chen et al. 2004; Legant et al. 2010), inter-macromolecular forces (Streichfuss et al. 2011), and extracellular matrix strain (Stone et al. 2007).

ACKNOWLEDGEMENTS

We gratefully acknowledge Arron Campi and John Kump (Crystal Technology, LLC) and Dale Gifford (Impact Zone) for AOTF and PCB design. We also thank Andrew Olson for synthesis of CdSe/CdS tetrapods, Sarah Swisher and Jesse Engel for electronics assistance, Joseph Shieh for providing HL-1 cells, and Dan Fletcher, Marc Shuman, and Philip Jess for helpful discussions. Work on AOTF microscope construction was supported by the Physical Chemistry of Semiconductor Nanocrystals Program, KC3105, Director, Office of Science, Office of Basic Energy Sciences, of the United States Department of Energy under Contract No. DE-AC02-05CH11231. Work on spectral imaging and analysis was supported by the National Institutes of Health through a pilot project of the Bay Area Physical Sciences Oncology Center, U54 CA143836. Work on tetrapod array fabrication was supported by the Dauben Fellowship (CLC). Work on cardiomyocyte culture was supported by a Department of Defense predoctoral fellowship, W81XWH-10-1-0168 (JC) and by funds from the National Institutes of Health U01 ES019458.

FIGURE LEGENDS

Figure C.1. Fabrication of luminescent tetrapod substrates. **(a)** Schematic of a cell grown on a fibronectin-coated monolayer of tetrapod nanocrystals. **(b)** Fluorescence image of a CdSe/CdS tetrapod monolayer substrate upon excitation with a 2.54 eV Ar⁺ laser. Scale bar, 100 μm. **(c)** As-synthesized CdSe/CdS tetrapod nanocrystals are surface-functionalized with phosphonic acid ligands (left); a schematic of a single ligand is shown for clarity. Following an amphiphilic polymer wrapping procedure the tetrapods are water soluble, with exposed carboxylic acid functional groups (center). The tetrapods are covalently linked to a transparent amine-functionalized coverglass substrate utilizing 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC)/N-Hydroxysulfosuccinimide (Sulfo-NHS) coupling chemistry. Incubation of the coupling reaction followed by thorough washing of the substrate results in a tetrapod nanocrystal monolayer covalently attached to the coverglass surface. **(d)** Brightfield snapshots depicting HL-1 cells growing and beating on a tetrapod substrate. Synchronized contractions are exhibited over the entire area (**Supplementary Movie C.1**). Yellow lines highlight one area that contracts (right) and relaxes (left).

Figure C.2. Diagram of the AOTF microscope used to collect photoluminescence spectra of a two-dimensional tetrapod array with spatial and temporal resolution.

Figure C.3. Beating heart cells induce shifts in the luminescence color emitted by a tetrapod quantum dot array. **(a–b)** Spectral map snapshots of peak shifts from a cardiomyocyte-perturbed tetrapod quantum dot array, periodically stressed by contractions over time. Two different areas on the luminescent array are shown in **(a)** and **(b)**. The data in **(a)** corresponds to the same area shown in **Fig. C.1d**. **(c)** Snapshots of the change in peak emission over time of a fibronectin-coated tetrapod array without cardiomyocytes as a control. Each map **(a–c)** shows ten successively collected frames with a spectral integration time of 350 ms. The change in peak energy at each pixel is plotted relative to a standard spectral map containing the highest peak energies observed. The side length of each frame is 40.7 μm .

FIGURES

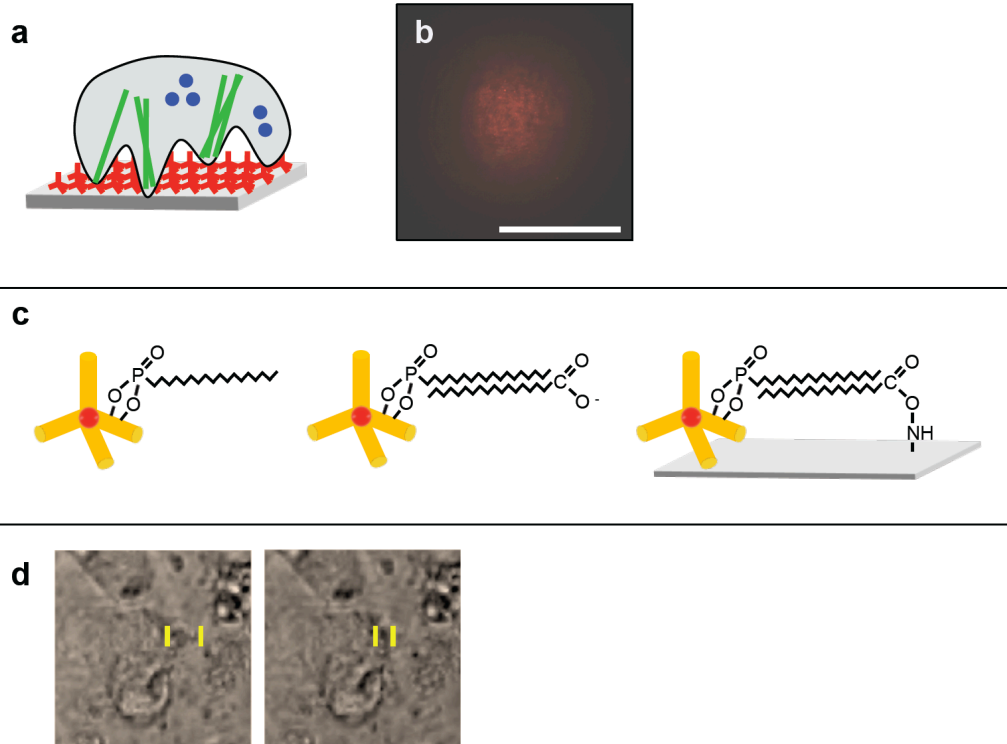


Figure C.1

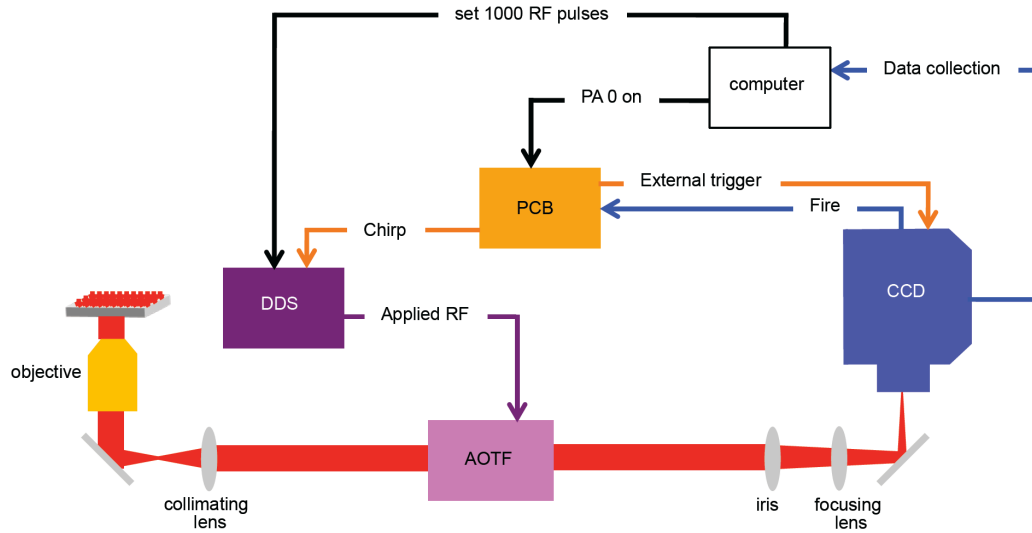


Figure C.2

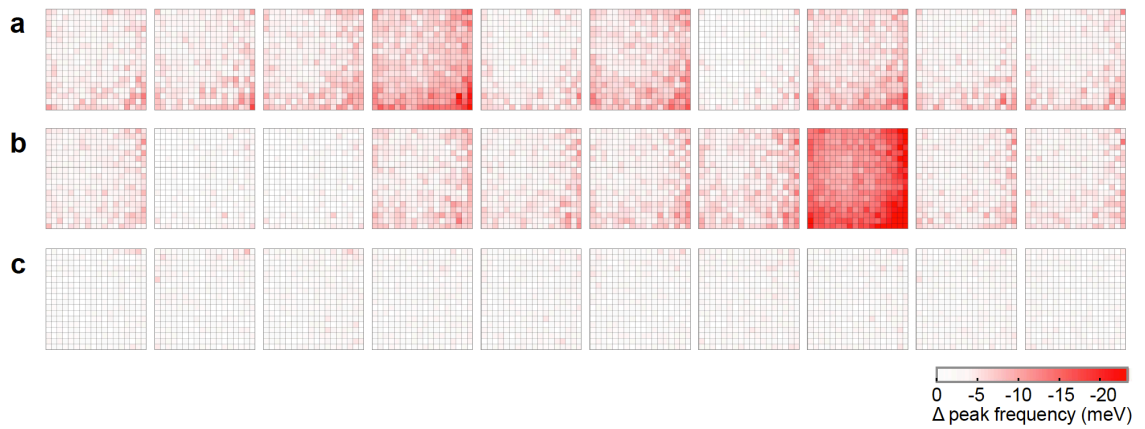


Figure C.3

MATERIALS AND METHODS

HL-1 cell culture on luminescent tetrapod substrates.

CdSe/CdS tetrapods were synthesized following previously reported methods (Fiore et al. 2009). A polymer-wrapping procedure was used to water-solubilize the tetrapods and create carboxylic acid groups on the tetrapod surface (Pellegrino et al. 2004). Tetrapod monolayer substrates were fabricated by covalently linking the tetrapods to transparent coverglass slides (no. 1.0) functionalized with a dense amine surface (custom made SuperAmine slides, ArrayIt) via a 1-Ethyl-3-[3-dimethylaminopropyl]carbodiimide hydrochloride (EDC)/*N*-Hydroxysulfo-succinimide (Sulfo-NHS) coupling reaction. CdSe/CdS tetrapods (4.5×10^{-8} M in PBS buffer), EDC (0.24 mg, Sigma), and sulfo-NHS (0.66 mg, Pierce) were reacted overnight with the amine-coated coverglass to achieve monolayer coverage. Each slide was washed thoroughly with PBS buffer to ensure that only covalently bound tetrapods remained.

Tetrapod monolayer substrates were placed inside cell culture wells, and HL-1 cells were cultured on the tetrapod monolayer according to culture conditions previously described (Claycomb et al. 1998). Briefly, the nanocrystal substrate was first coated with a dilute solution of fibronectin (Sigma) for 1-2 hours. HL-1 cells were grown in Claycomb media (Sigma) supplemented with norepinephrine (100 μ M), FBS (10%), L-glutamine (4 mM), and penicillin/streptomycin. The cells were plated onto the nanocrystal monolayer and allowed to adhere overnight. One to two days later, the medium was changed, supplemented with HEPES buffer (25 mM), and the cells were imaged. HL-1

cells cultured on fibronectin-coated tetrapod substrates exhibited qualitatively similar morphological and contractile phenotypes as those grown on normal tissue culture plates.

Acousto-optic tunable filter (AOTF) microscope.

A schematic of the microscope is shown in **Fig. C.2**. Our design is based on previous work utilizing an AOTF for spatially resolved spectral imaging (Wachman et al. 1997; Gupta 1999). Our samples were excited at 2.54 eV with an Ar⁺ laser (3 mW total power, 300 μm spot size). Luminescence from the sample is collected through an objective (Zeiss 40x LD Plan-Neofluar, NA = 0.6 with correction collar) and this light is collimated using achromatic doublet lenses and sent through the AOTF (97-02838-01, Crystal Technology, LLC). An iris is used to block the undeflected beam, and the deflected beam is focused using an additional achromatic doublet lens onto a CCD (Andor iXon 897).

The computer sets 1000 radio frequencies (RFs) to be applied by the direct digital synthesizer (DDS) onto the AOTF. We calibrated the AOTF and found that $x_{\text{eV}} = .0205 \cdot (x_{\text{MHz}} - 91.41) + 1.942$, where x_{MHz} = the applied RF (MHz) and x_{eV} = the light frequency allowed to deflect through the crystal (eV). Typically, we applied a sweep of 100 different frequencies ranging from 85-101 MHz, which corresponds to 1.81-2.14 eV. The CCD collects an intensity image at each RF frequency with a total time (including exposure time and transfer time) of 3.5 ms per frequency, for a total spectral collection time of 350 ms. We used a custom

made printed circuit board (PCB, Impact Zone) to synchronize the AOTF and CCD; the time offset between the two was 10 ns.

Spectral data collection and analysis.

Samples were taken from a 37° C incubator and imaged on our microscope at room temperature. HL-1 cells exhibited beating on the tetrapod substrate throughout data collection. At each spot imaged, a brightfield movie (25 s, 10 frames per s), brightfield image (9.2 ms exposure time), and fluorescence image (9.2 ms exposure time) were collected using a camera (PaxCam 2+, MIS, Inc.) mounted on top of the microscope. Emitted light from the tetrapods was then sent through the AOTF to obtain spatially resolved photoluminescence spectra of the substrate (20 spectra at each pixel simultaneously, 350 ms total integration time per spectra). A brightfield movie was taken after spectra collection to determine whether any perturbations to cardiomyocyte contractions had been induced by laser excitation. In all spots studied, no changes in beating behavior following laser excitation were observed.

The time resolution for spectral collection was limited to 350 ms total integration time by the luminescence intensity reaching the CCD. Although the rate of cardiomyocyte beating was ~0.5-1 Hz, we were unable to consistently optimally capture in succession the stress changes within one beat cycle due to the rapid timescale of contraction (~100 ms) and variations in beating periodicity (**Supplementary Movies C.1 and C.2**). Future work will increase spatial and temporal resolution via improvements to tetrapod quantum yield and microscope

design. Use of electrical stimuli and addition of a microscope stage with temperature and CO₂ control may further optimize sub-cycle stress imaging through synchronization of beating with spectral measurements and a more consistent beat frequency.

Spectral maps of shifts in peak emission were plotted to highlight changes in array luminescence. The change in energy at each pixel is calculated relative to a standard spectral map for a given area containing the highest peak energies observed at each pixel. A per pixel offset calculation is required since emission from the AOTF deflected beam was slightly spatially dispersed. The highest peak energies emitted by cardiomyocyte-perturbed array areas correspond to the average frequency emitted by an unperturbed control array, suggesting that the highest energies observed in a perturbed array represent a relatively unstressed state. This observation strongly advocates that cell-generated forces red shift tetrapod luminescence in the experimental geometry, in concordance with previous calculations (Schrier et al. 2008) and demonstrations (Choi et al. 2009; Choi et al. 2010) that symmetry-breaking deformations induce red shifting of the tetrapod energy gap.

To calculate the average force, the average and standard deviation in offset value over all pixels in a given area (6480 total, from 324 pixels per frame and 20 frames) were calculated and converted from MHz to eV using our determined calibration and eV to nN following previous electronic structure calculations (Schrier et al. 2008). The force values reported in the main text were calculated using data from the area depicted in **Fig. C.3a–C3.b**. An identical

calculation on data from a control area determined an average force of 0.3 ± 0.2 nN, suggesting the minimum threshold of force detectable in our system. Work to thoroughly characterize the luminescence response of tetrapods to a set of all relevant stress states will provide an experimental eV to nN conversion, enabling more precise quantitative measurements of cell-generated forces.

Appendix D: Generation of novel FVB cell lines for breast cancer studies

Source: The following appendix is unpublished work, and describes the generation and some preliminary characterization of new breast cancer cell lines.

Contributions: I generated these new cell lines, with help and advice from Sylvain Provot and Audrey Brenot. Some of the qPCR experiments were performed with the assistance of Jennifer Tai and Joanne Dai. Vicki Plaks and I performed the in vivo characterization of these new cell lines. Zena Werb supervised the research.

ABSTRACT

The FVB/n strain has been widely used to study breast cancer. Indeed, a number of transgenic breast cancer mouse lines carrying various oncogenes have been generated on this background and are widely studied. However, few FVB/n breast cancer lines exist, which limits the ability to do orthotopic transplants and experimental metastasis assays in immunocompetent hosts. Here, I describe the generation and initial characterization of four new breast cancer lines that carry various oncogenes, representing both luminal and basal breast cancer subtypes. Because I have generated most of these cells to be both luminescent and fluorescent, these lines should be a valuable resource for future researchers performing single tumor transplant experiments, as well as spontaneous and experimental metastasis assays, in fully immunocompetent FVB/n mice.

INTRODUCTION

The inbred FVB/n mouse strain is widely used in cancer research. This strain was created in the early 1970s, and is characterized by vigorous reproductive performance and consistently large litters. Moreover, fertilized FVB/N eggs contain large and prominent pronuclei, which facilitate microinjection of DNA (Taketo et al. 1991). Indeed, transgenic mice bearing various oncogenes driven by the mouse mammary tumor virus (MMTV) promoter, the whey acidic protein (WAP) promoter, and the prostatic steroid binding protein (C3(1)) promoter have been generated on the FVB/n background to model various aspects of luminal and basal breast cancers. However, there exists few established breast tumor lines on this particular strain, precluding the use of syngeneic transplants of mammary tumors into immunocompetent hosts to study primary tumor formation and metastasis.

One breast cancer line that has previously been described was derived from an osteolytic inducing MMTV-PyMT carcinoma (Halpern et al. 2006). In addition, our lab has generated another cell line from an MMTV-PyMT primary carcinoma. This line displays typical epithelial features in two dimensional (2D) tissue culture, forms primary tumors when injected orthotopically into the mammary fat pad, and is metastatic when injected i.v. into the tail vein. Here, I describe the creation of additional cell lines derived from other strains of mice, including the MMTV-PyMT-Luc/mCherry mouse, which expresses luciferase and mCherry under the β -actin promoter, the MMTV-PyMT-DB mouse, which expresses a mutant form of the PyMT oncogene that cannot activate the PI3K

pathway (Webster et al. 1998), the MMTV-Neu mouse, which over-expresses the rat Her2/Neu oncogene, and the C3(1)-Tag mouse, which more closely resembles the basal subtype of breast cancer. In addition, I describe three additional lines generated by other investigators, which I have made luminescent and fluorescent for metastasis studies: a liver tumor line driven by the Myc and Ras oncogenes (LT2MR, derived by Andrei Goga), a spontaneous lung cancer line (LAP0297, derived by Peigen Huang), and a spontaneous mammary adenocarcinoma line (MCP0008, also derived by Peigen Huang) (Huang et al. 2008). I hope that these lines will be useful for future investigators to study tumor growth and metastasis in the FVB/n mouse strain.

RESULTS AND DISCUSSION

A cell line generated from an MMTV-PyMT-Luc/mCherry tumor

The Luciferase-mCherry mice (Line 614) were originally designed and generated by Peter Dijkgraaf. These mice carry a dual-function reporter, consisting of firefly luciferase (Luc) fused to mCherry via a seven-alanine linker, and under the control of the pCA promoter. This promoter is a hybrid of the cytomegalovirus (CMV) enhancer and the chicken β -actin promoter, which provides robust expression in most mouse tissues, including the mammary gland (**Fig. D.1A**). We crossed the MMTV-PyMT mice to homozygous Luciferase-mCherry mice to obtain PyMT females with a single copy of this reporter. These MMTV-PyMT-Luc/mCherry mice are essentially identical to the original MMTV-PyMT mice previously described (Guy et al. 1992), and display similar tumor growth kinetics. The advantage is that these cells will express Luc and mCherry under the pCA promoter instead of a viral promoter, which has a higher potential of being silenced during long-term in vivo studies.

We isolated several tumors from 15-week old mice and pooled the tumors. We attempted to generate four independent cell lines from these tumors (Y193, Y194, Y197, and YMG4). After plating these cells in 2D culture, they formed tightly adherent epithelial colonies (**Fig. D.1B**). We validated that the cells we were growing in culture expressed high levels of luciferase by IVIS imaging (**Fig. D.1C**), and that these cells expressed mCherry by flow cytometry (**Fig. D.1D**). We also verified that we were growing epithelial cells by co-staining for EpCAM,

an epithelial marker expressed on PyMT cells, and sorted the cells to isolate the mCherry^{hi}-expressors (**Fig. D.1D**).

These cells were growing very robustly for the first two to three passages. However, after the fourth passage, the cells started to proliferate more slowly, likely due to a senescence crisis previously described by Hayflick (Kuilman et al. 2010). Previous researchers have described observing premature senescence in vitro because of inadequate culturing conditions. When cells are explanted from an organism and placed in culture, they have to adapt to an artificial environment, characterized by abnormal concentrations of nutrients and growth factors and the presence of ambient O₂ levels, as well as the absence of surrounding cell types and extracellular matrix components (Sherr and DePinho 2000). These conditions can induce a “culture shock,” resulting in stress-induced senescence without changes in telomere length. We have cultured these cells in cells containing 10% fetal bovine serum (FBS) and in 21% O₂ incubators. We have not tried culturing these cells in serum-free medium supplemented with a number of defined growth factors or under low oxygen conditions (e.g., 3% O₂).

The PyMT-Luc/mCherry cells appear to exhibit a more heterogeneous morphology over time. Both large epithelial clusters as well as more fibroblastic cells can be seen. We are continuing to culture and expand these cells, even three months after first isolating the tumors. We should consider lowering the serum concentration to 2 – 5% and lowering to the O₂ concentration to determine if the cells will grow better. We have not yet performed in vivo experiments with this cell line to determine the tumorigenicity and metastatic capacity.

A cell line generated from an MMTV-PyMT-DB tumor

The PyMT oncogene has no known catalytic activity, but rather acts as a scaffold for signal transduction molecules. Protein phosphatase 2A, protein tyrosine kinases of the Src family, phosphatidylinositol 3' kinase (PI3K), phospholipase C γ 1 as well as the Shc/Grb2 adaptors are all assembled on PyMT (Schaffhausen and Roberts 2009). Mutations of these different sites have helped elucidate different signaling pathways important for different aspects of transformation. The PyMT-DB mutant harbors a mutation in PyMT that renders it incapable of activating downstream PI3K signaling (Webster et al. 1998). Although mice carrying the mutant PyMT-DB oncogene develop hyperplasias with similar latency as the normal PyMT mice, these mutant cells are highly apoptotic, suggesting that activation of the PI3K is necessary for cell survival (Webster et al. 1998). Genetic and biochemical analyses of tumors in the PyMT-DB mice demonstrated no evidence of reversion of the PI3K binding site in tumors, in contrast to revertants that were identified in other PyMT mutants (Webster et al. 1998). However, these tumors eventually grow after an additional 8 – 12 week period, suggesting that additional mutations develop to permit cell survival. Indeed, upregulation of the epidermal growth factor receptor (EGFR) family members, which are known to couple to the PI3K and Shc signaling pathways, are observed in tumors that progress.

Cells generated from PyMT-DB tumors retained a luminal epithelial morphology in culture (**Fig. D.2A**). Although these cells also went through a

senescence crisis, cells eventually succeeded in growing after a two month period in culture, in media containing 10% FBS and grown in normal 21% O₂ incubators. We also infected these cells with pMIG-GFP and sorted them for GFP (**Fig. D.2B**). Future studies to determine the degree of PI3K activation in this cell line will be important to identify distinguishing characteristics between this PyMT-DB line and the normal PyMT lines generated. In addition, we will need to characterize the tumorigenicity and metastatic capacity in vivo.

A cell line generated from an MMTV-Neu tumor

The MMTV-Neu model also represents the luminal subtype of breast cancer. These mice over-express the un-activated version of the rat ErbB2/Neu oncogene under the MMTV promoter. In 2D culture, these cells exhibit a luminal morphology (**Fig. D.3A**). We transduced these cells with luciferase (pMSCV-Luc) so that we could conduct metastasis studies with these cells, and confirmed that these cells expressed high levels of luciferase (**Fig. D.3B**). We also confirmed that the NeuC1 cells over-expressed the Her2/Neu oncogene by quantitative PCR (qPCR) (**Fig. D.3C**) and by western blot (**Fig. D.3D**).

We sought to determine their metastatic potential after intravenous (i.v.) injection into the tail vein. We injected varying concentrations of cells (1×10^5 , 5×10^5 and 1×10^6) into FVB/n mice, and monitored metastatic growth in the lungs by bioluminescence imaging. We found that mice injected with 1×10^6 cells had detectable bioluminescence signal by 2 weeks post-injection (**Fig. D.3E**). After four weeks, the luminescence signal did not significantly increase, so we decided

to sacrifice the mice. We did not observe any macroscopic metastases on the surface, but when we examined the lungs histologically, we found several micro-metastases (**Fig. D.3F**). Interestingly, these metastases appear well-differentiated. We have not confirmed that these cells continue to over-express Her2/Neu in vivo. In addition, we need to conduct further studies to determine the tumorigenicity of the cells when injected into the fat pad, and to determine if these cells will spontaneously metastasize to the lungs.

A cell line generated from a C3(1)-Tag tumor

The C3(1)-Tag mouse represents the basal subtype of breast cancer (Zhu et al. 2011). In this model, the C3(1) component of the rat prostate steroid binding protein controls expression of the SV40 large T-antigen (Tag) to the epithelium of the mammary and prostate glands. Atypia of the mammary ductal epithelium develops at about 8 weeks of age, progressing to mammary intraepithelial neoplasia (which resembles human ductal carcinoma in situ (DCIS)) at about 12 weeks of age. At about 16 weeks of age, all female mice have invasive carcinomas, with about a 15% incidence of lung metastases (Green et al. 2000). The invasive carcinomas from C3(1)-Tag mice are hormone-independent, which corresponds to the loss of estrogen receptor-alpha ($ER\alpha$) expression during tumor progression.

These cells displayed a heterogeneous phenotype when first cultured, but were more elongated and mesenchymal looking than the luminal cell lines (**Fig. D.4A**), consistent with the basal features of this model. Because of the large

number of fibroblasts that contaminated the culture, we decided to enrich for epithelial cells by staining and sorting for EpCAM. We isolated both the EpCAM-positive and EpCAM-negative cells. Not surprisingly, we found that the EpCAM-positive and EpCAM-negative populations had different morphologies, with the more epithelial looking cells in the EpCAM-positive population (**Fig. D.4B**). Because it has been previously described that EpCAM-positive and negative cells might interconvert (with one population being the more “stem-like” population giving rise to the other population), we sought to determine whether C3Tag cells exhibited this phenomenon. After three weeks in culture, we re-stained the EpCAM-positive and negative cells for EpCAM. We found that the cell surface phenotypes were very stable over time, with greater than 90% of the cells retaining their original phenotype (**Fig. D.4C**). We further confirmed that these cells expressed the SV49 large T-antigen by qPCR (**Fig. D.4D**). We also transduced these cells with luciferase virus, and confirmed their luminescence signal (data not shown).

When injected 1×10^6 cells i.v., we found that only the EpCAM-positive cells formed detectable micro-metastases (**Fig. D.4E**). In contrast, we could not detect any metastases in the mice injected with EpCAM-negative cells (**Fig. D.4F**). This suggests that the EpCAM-positive cells contain the bona fide tumor cells. Further experiments will need to be performed to determine whether the EpCAM-positive and negative populations can equally form primary tumors when injected into the mammary fat pad. We also need to determine if orthotopic primary tumors will spontaneously metastasize to the lungs.

Cell lines from miscellaneous tumors: LAP0297, MaP0008 and LT2MR cells

We were additionally interested in two cell lines derived from spontaneous FVB/n tumors previously described by Huang and colleagues (Huang et al. 2008). The LAP0297 line was derived from spontaneous lung adenocarcinoma from a male mouse, and the MaP0008 line was from a spontaneous breast tumor. We found that the LAP0297 cells were elongated in culture (**Fig. D.5A**). We transduced the cells with luciferase and GFP, confirmed their expression (data not shown), and injected these cells i.v. at concentrations between $1 - 10 \times 10^5$. We found that even at low concentrations, these cells very rapidly formed lung tumors, which were detectable both by bioluminescence imaging (**Fig. D.5B**), as well as histologically (**Fig. D.5C**). The lung tumors were poorly differentiated adenocarcinomas, consistent with prior studies. Many of these mice had such aggressive lung tumors that they were found dead within two weeks of the experiment. When the mice were sacrificed, several macroscopic metastases were readily apparent on the lung surface.

We also did some initial characterization of the MaP0008 line. In culture, these cells also appeared very elongated (**Fig. D.5D**). We also transduced these cells with luciferase and GFP. However, when injected i.v., we could not detect any metastases either by bioluminescence imaging (**Fig. D.5E**) or histologically (**Fig. D.5F**), even when 1×10^6 cells were injected. These data suggest that the MaP0008 cells do not form experimental metastases in vivo. We will need to determine whether these cells form primary tumors in our hands.

We also obtained the LT2MR liver cancer line generated by Goga and colleagues, which was derived from a male primary liver tumor expressing Myc and then retrovirally transduced with Ras. In culture, these cells appear elongated and fibroblastic (**Fig. D.6A**). We also transduced these cells with luciferase and injected the cells i.v. to test their ability to form experimental lung metastases. We found that cells in the lung were readily detectable one week post-injection by bioluminescence imaging. However, the signal was progressively lost when we imaged at two weeks post-injection, and was non-detectable by three weeks post-injection (**Fig. D.6B**). We initially hypothesized that perhaps the luciferase was being silenced over time, although all of the other cell lines have not had this problem within this time frame. However, when we sacrificed the mice four weeks post-injection, we did not find any surface lung metastases, nor did we find any micro-metastases when we examined the lungs histologically. Our results suggest that perhaps the cells are capable of seeding of the lungs initially (as evidenced by a robust signal one week post-injection), but are unable to successfully colonize and grow. This may be due to the fact that these cells were derived from a male mouse, which might cause tumor cell rejection, although the LAP0297 cells were also from a male mouse. Therefore, we conclude that LT2MR cells cannot experimental metastases, at least when injected i.v. into female mice.

Preliminary analysis of differentiation markers

Finally, we have done a preliminary analysis of differentiation markers expressed by the various cell lines. In particular, we have focused on the PyMT, NeuC1, C3Tag, MaP0008 and LAP0297 cells. Consistent with the luminal subtype of the PyMT and NeuC1 cells, these cells express high levels of luminal markers (e.g., *Krt8*, *Krt18*, *Gata3*) and low levels of basal markers (e.g., *Krt14*, *Snail1*, *Snail2*, *Zeb1*, *vimentin*). Conversely, the C3Tag, MaP0008 and LAP0297 cells express higher levels of basal markers (**Fig. D.7A**). We also analyzed E-cadherin and vimentin protein expression by western blot and found that the luminal cells expressed E-cadherin and low levels of vimentin, whereas the basal cells expressed low levels of E-cadherin and high levels of vimentin (**Fig. D.7B**). One of the C3Tag-EpCAM-positive samples did express E-cadherin. Taken together, our analysis demonstrate that the cell lines that we generated from luminal type tumors retain their luminal characteristics in culture over several months in culture, while the cell lines from basal type tumors retain their basal characteristics. We anticipate that these cell lines will be useful for future studies to interrogate differences between luminal and basal cancer cells in vivo (e.g., in response to chemotherapy, or in response to over-expression or knockdown of particular genes) using syngeneic, fully immuno-competent FVB/n mice. Moreover, the preliminary experiments demonstrate that some of these lines will be useful for experimental metastasis studies.

ACKNOWLEDGMENTS

We thank Elena Atamaniuc and Ying Yu for excellent technical assistance, as well as Marja Lohela for maintaining C3(1)-Tag mice in the lab. We thank Mark Moasser for helping to generate the Her2/NeuC1 cell line, Andrei Goga for sharing the LT2MR liver cancer line, and Peigen Huang for sharing the LAP0297 and MaP0008 lines.

EXPERIMENTAL PROCEDURES

Cell culture

The cell lines were generated by isolating late stage tumors from the following FVB/n mice: MMTV-PyMT-Luc-mCherry, MMTV-Neu, MMTV-PyMT-DB and C3(1)-Tag. The tumors were dissociated in collagenase and trypsin, and the epithelial fraction was enriched by differential centrifugation (to eliminate the stromal cells e.g., RBCs, fibroblasts, myeloid cells). Tumor cells were cultured in DMEM/F12 media supplemented with 10% FBS, 1 $\mu\text{g}/\text{mL}$ insulin, 1 $\mu\text{g}/\text{mL}$ hydrocortisone and penicillin-streptomycin. Cells were passaged by trypsinization, and grown for at least 2 months before enough cells escaped senescence to proliferate well in culture.

Lentiviral and Retroviral Production

Viral production was carried out using calcium phosphate-mediated transfection of HEK 293T or GP2 cells. Virus was concentrated by ultracentrifugation. Cells were transduced with pMSCV-Luciferase (retrovirus, puromycin resistance), pLV-Luciferase (lentivirus, neomycin/G418 resistance) or pMIG-GFP (retrovirus, no resistance) and then selected in 2 $\mu\text{g}/\text{mL}$ puromycin or 200 $\mu\text{g}/\text{mL}$ G148 for at least 5 days or by FACS.

Quantitative real-time PCR (qPCR)

Total RNA was isolated from cells using the miRNeasy Mini Kit (Qiagen). cDNA was synthesized using the Superscript III RT First Strand Kit (Invitrogen). qPCR

was performed using FastStart Universal SYBR Green master mix (Roche) in an Eppendorf Mastercycler realplex machine. Ct values were normalized to actin and GAPDH, and relative expression was calculated using the $2^{-\Delta\Delta Ct}$ method. Primer sequences for qPCR were found using the Harvard Primer Bank.

Animal Studies

All animal experiments were performed at UCSF, and reviewed and approved by the UCSF IACUC. Mice were housed under pathogen-free conditions in the UCSF barrier facility. FVB/n mice, originally from Jackson Laboratories, were bred in-house. MMTV-PyMT-DB (originally from W. Muller), MMTV-Neu (originally from Jackson), C3(1)-Tag, and MMTV-PyMT-Luc-mCherry mice were bred in-house. For experimental metastasis experiments, age-matched female mice were injected i.v. (via tail vein) with $1-10 \times 10^5$ cells, depending on the cell line. Bioluminescence imaging was performed using an IVIS Spectrum (Caliper Life Science). Image radiance values were normalized using the Living Image software.

FIGURE LEGENDS

Figure D.1. A cell line generated from an MMTV-PyMT-Luc/mCherry tumor.

(A) Schematic of construct used to generate Line 614, which carries the Luc/mCherry transgene under the pCA promoter.

(B) PyMT-Luc/mCherry tumors from 15 week old mice were digested in collagenase and plated on tissue culture plates. Phase contrast image of the cells after one passage in culture

(C) PyMT-Luc/mCherry tumor cells were transduced with pMSCV-luciferase virus and the luminescence was verified by bioluminescent imaging. Cells were plated at high density (top) and low density (bottom).

(D) PyMT-Luc/mCherry tumor cells were stained for EpCAM and analyzed by flow cytometry. Approximately 95% of the cells were positive for EpCAM expression. The endogenous mCherry shows heterogeneous expression and so the mCherry^{hi} cells were isolated by FACS and re-plated to expand in culture.

Figure D.2. A cell line generated from an MMTV-PyMT-DB tumor.

(A) PyMT-DB tumors from 24 week old mice were digested in collagenase and plated on tissue culture plates. Phase contrast image of the cells after three passages in culture.

(B) PyMT-DB tumor cells were infected with pMIG-GFP virus and sorted for GFP expression by FACS.

Figure D.3. A cell line generated from an MMTV-Neu tumor.

(A) NeuC1 tumors from a 47 week old mouse were digested in collagenase and plated on tissue culture plates. Phase contrast image of the cells after three passages in culture.

(B) NeuC1 tumor cells were transduced with pMSCV-luciferase virus and the luminescence was verified by bioluminescent imaging. Cells were plated at high density (top), medium density (middle) and low density (bottom).

(C–D) Over-expression of Her2/Neu/ErbB2 was verified by quantitative PCR (C) and by western blot (D). PyMT tumor cells, which do not over-express the Her2 oncogene, were used as controls for the qPCR and western blot.

(E) Bioluminescent imaging of FVB/n mice i.v. injected with 1×10^6 NeuC1 cells after 2 weeks. A very weak signal barely above background was detected in one of two mice.

(F) Lungs were sectioned, stained with H&E and examined histologically for micro-metastases. Several micro-metastases were identified (outlined), which appeared to be well-differentiated epithelial clusters of cells.

Figure D.4. A cell line generated from a C3(1)-Tag tumor.

(A) C3(1)-Tag tumors from a 20 week old mouse were digested in collagenase and plated on tissue culture plates. Phase contrast images of two representative cell lines after three passages in culture shows the heterogeneity within the culture, including more luminal and more mesenchymal epithelial cells and fibroblasts.

(B) C3(1)-Tag tumor cells were stained for EpCAM and sorted into EpCAM-positive and EpCAM-negative populations by FACS. Cells were re-plated in culture and phase contrast images of the two populations were taken, which demonstrates the more epithelial EpCAM-positive and more mesenchymal and fibroblastic EpCAM-negative population.

(C) After three weeks in culture, the EpCAM-positive and EpCAM-negative C3(1)-Tag cells were stained for EpCAM again to determine whether cells had transdifferentiated. EpCAM-positive cells retained their EpCAM surface expression, with 91% of the cells within the EpCAM^{pos} gate, while EpCAM-negative cells remained negative, with >99% of the cells within the EpCAM^{neg} gate.

(D) Expression of SV40 large T antigen was verified by qPCR.

(E) EpCAM-positive and EpCAM-negative cells were i.v. injected into FVB/n mice and allowed to grow for four weeks in vivo. Lungs were sectioned and stained with H&E, and examined for micro-metastases. Representative H&E images show that only EpCAM-positive cells were capable of forming micro-metastases.

Figure D.5. Cell lines generated from spontaneous lung and breast tumors.

(A) Phase contrast image of LAP0297 cells derived from a spontaneous lung adenocarcinoma from a male FVB/n mouse.

(B) Bioluminescence images of female mice i.v. injected with LAP0297 cells. Lung tumors formed as soon as 2 weeks after injection of 1×10^5 cells.

(C) Representative H&E images of the lung tumors from i.v. injected LAP0297 cells shows poorly differentiated lung adenocarcinoma.

(D) Phase contrast image of MaP0008 cells derived from a spontaneous breast adenocarcinoma from an FVB/n mouse.

(E) Bioluminescence images of female mice i.v. injected with 1×10^6 MaP0008 cells.

(F) Representative H&E image of the lungs of mice i.v. injected with MaP0008 demonstrates that these cells do not form experimental lung metastases.

Figure D.6. An FVB/n liver tumor cell line expressing Myc and Ras.

(A) Phase contrast image of LT2MR cells derived from a Myc-expressing hepatocellular carcinoma from a male FVB/n mouse.

(B) Bioluminescence images of female mice i.v. injected with 1×10^6 LT2MR cells. Signal in the lungs were detected a week 1 post-injection, but was then progressive lost by week 2.

Figure D.7. Analysis of epithelial and mesenchymal markers expressed by these new cell lines.

(A) qPCR expression analysis of PyMT, NeuC1, C3Tag, MaP0008 and LAP0297 cells for luminal epithelial markers (*Krt8*, *Krt18*, *Gata3*, *Cdh1*) and mesenchymal markers (*p63delN*, *Snail1*, *Zeb2*, *vimentin*, *Krt14*).

(B) Western blot analysis of PyMT, NeuC1, C3Tag, MaP0008 and LAP0297 cells for E-cadherin and vimentin expression.

FIGURES

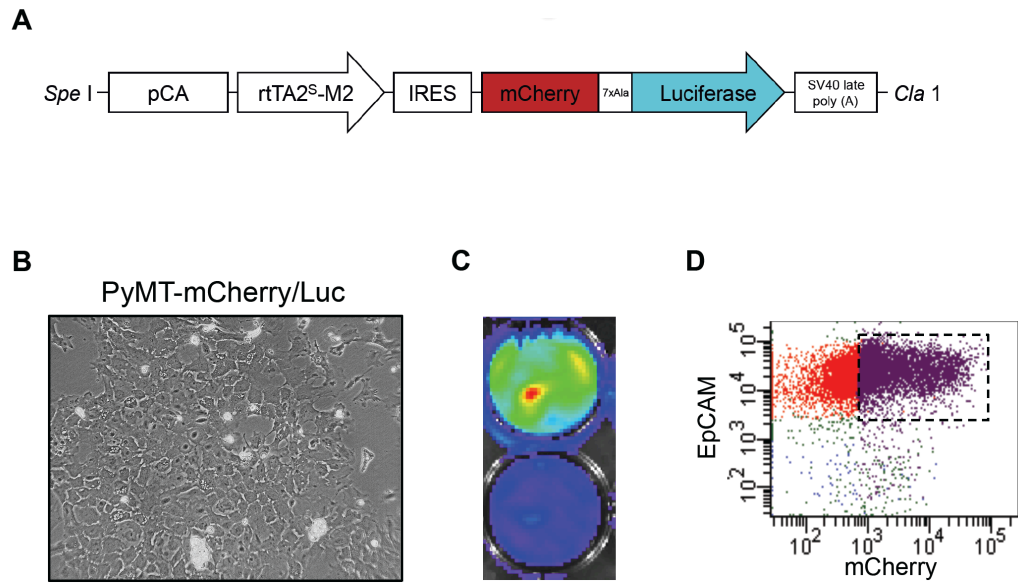


Figure D.1

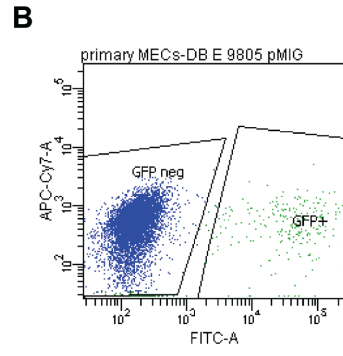
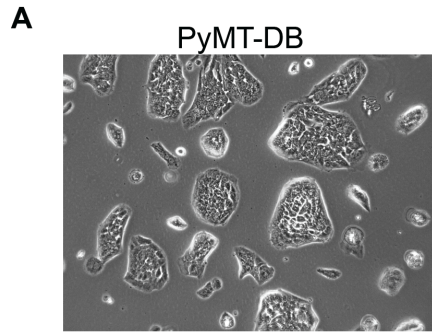


Figure D.2

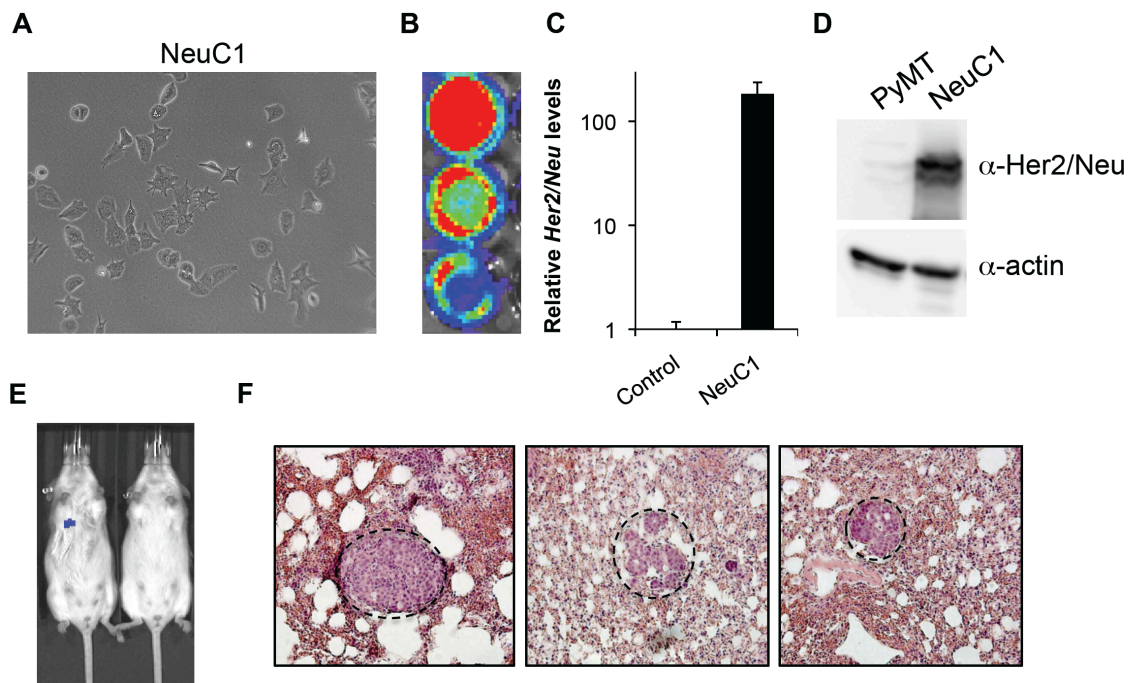


Figure D.3

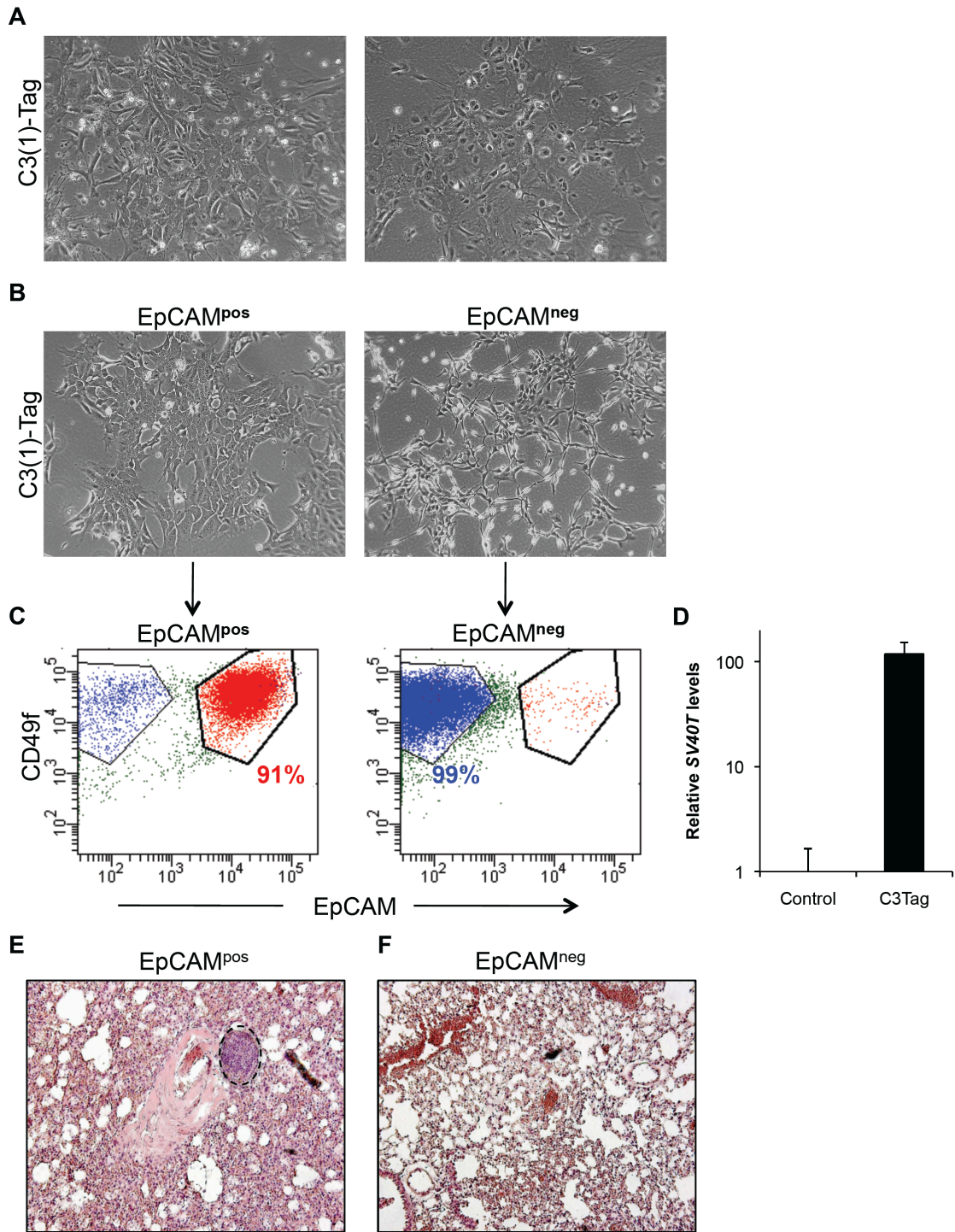


Figure D.4

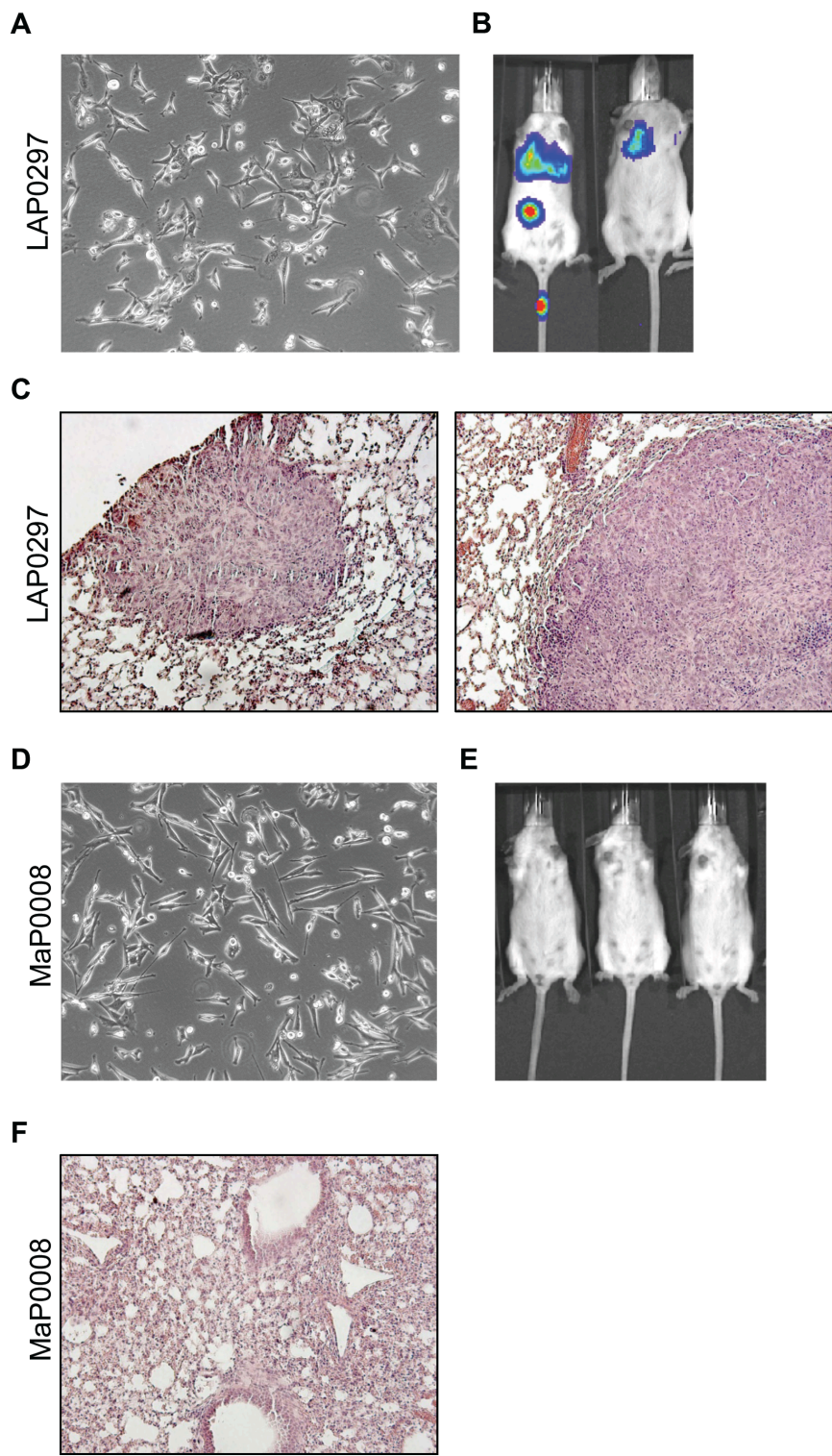


Figure D.5

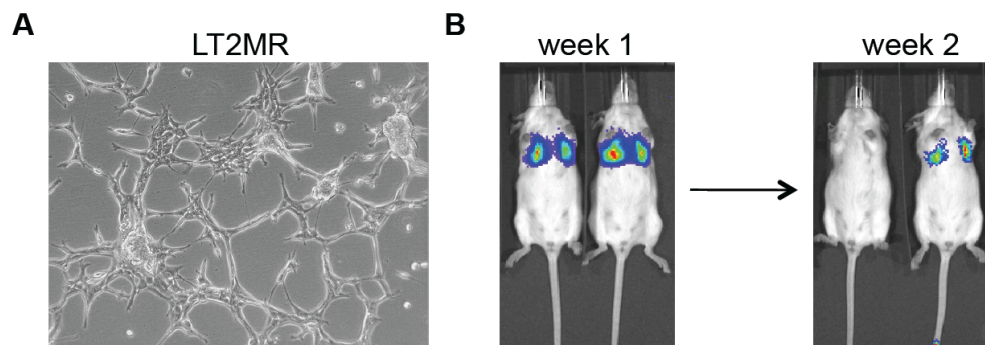


Figure D.6

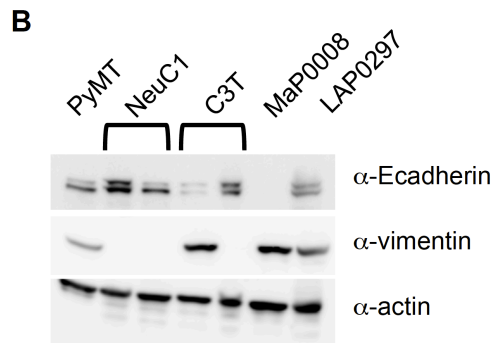
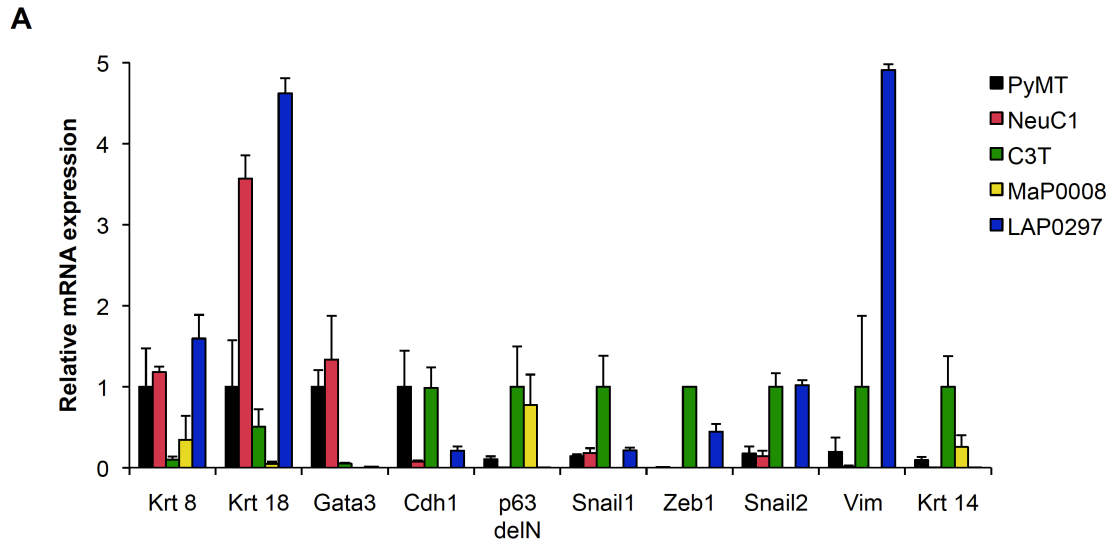


Figure D.7

REFERENCES

- Abba, M.C., Nunez, M.I., Colussi, A.G., Croce, M.V., Segal-Eiras, A., and Aldaz, C.M. 2006. GATA3 protein as a MUC1 transcriptional regulator in breast cancer cells. *Breast Cancer Res* **8**(6): R64.
- Acharyya, S., Oskarsson, T., Vanharanta, S., Malladi, S., Kim, J., Morris, P.G., Manova-Todorova, K., Leversha, M., Hogg, N., Seshan, V.E. et al. 2012. A CXCL1 Paracrine Network Links Cancer Chemoresistance and Metastasis. *Cell* **150**(1): 165-178.
- Adelaide, J., Finetti, P., Bekhouche, I., Repellini, L., Geneix, J., Sircoulomb, F., Charafe-Jauffret, E., Cervera, N., Desplans, J., Parzy, D. et al. 2007. Integrated profiling of basal and luminal breast cancers. *Cancer Res* **67**(24): 11565-11575.
- Aigner, K., Dampier, B., Descovich, L., Mikula, M., Sultan, A., Schreiber, M., Mikulits, W., Brabletz, T., Strand, D., Obrist, P. et al. 2007. The transcription factor ZEB1 (deltaEF1) promotes tumour cell dedifferentiation by repressing master regulators of epithelial polarity. *Oncogene* **26**(49): 6979-6988.
- Akbari, A., Razzaghi, Z., Homaei, F., Khayamzadeh, M., Movahedi, M., and Akbari, M.E. 2010. Parity and breastfeeding are preventive measures against breast cancer in Iranian women. *Breast Cancer* **18**(1): 51-55.
- Albiges-Rizo, C., Destaing, O., Fourcade, B., Planus, E., and Block, M.R. 2009. Actin machinery and mechanosensitivity in invadopodia, podosomes and focal adhesions. *J Cell Sci* **122**(Pt 17): 3037-3049.
- Amsen, D., Antov, A., Jankovic, D., Sher, A., Radtke, F., Souabni, A., Busslinger, M., McCright, B., Gridley, T., and Flavell, R.A. 2007. Direct regulation of Gata3 expression determines the T helper differentiation potential of Notch. *Immunity* **27**(1): 89-99.
- Anastasiadis, P.Z. 2007. p120-ctn: A nexus for contextual signaling via Rho GTPases. *Biochim Biophys Acta* **1773**(1): 34-46.
- Andreu, P., Johansson, M., Affara, N.I., Pucci, F., Tan, T., Junankar, S., Korets, L., Lam, J., Tawfik, D., DeNardo, D.G. et al. 2010. FcRgamma activation regulates inflammation-associated squamous carcinogenesis. *Cancer Cell* **17**(2): 121-134.
- Ansel, K.M., Djuretic, I., Tanasa, B., and Rao, A. 2006. Regulation of Th2 differentiation and Il4 locus accessibility. *Annu Rev Immunol* **24**: 607-656.
- Aslakson, C.J. and Miller, F.R. 1992. Selective events in the metastatic process defined by analysis of the sequential dissemination of subpopulations of a mouse mammary tumor. *Cancer Res* **52**(6): 1399-1405.
- Asselin-Labat, M.L., Shackleton, M., Stingl, J., Vaillant, F., Forrest, N.C., Eaves, C.J., Visvader, J.E., and Lindeman, G.J. 2006. Steroid hormone receptor status of mouse mammary stem cells. *J Natl Cancer Inst* **98**(14): 1011-1014.
- Asselin-Labat, M.L., Sutherland, K.D., Barker, H., Thomas, R., Shackleton, M., Forrest, N.C., Hartley, L., Robb, L., Grosveld, F.G., van der Wees, J. et al.

2007. Gata-3 is an essential regulator of mammary-gland morphogenesis and luminal-cell differentiation. *Nat Cell Biol* **9**(2): 201-209.
- Asselin-Labat, M.L., Sutherland, K.D., Vaillant, F., Gyorki, D.E., Wu, D., Holroyd, S., Breslin, K., Ward, T., Shi, W., Bath, M.L. et al. 2011. Gata-3 negatively regulates the tumor-initiating capacity of mammary luminal progenitor cells and targets the putative tumor suppressor caspase-14. *Mol Cell Biol* **31**(22): 4609-4622.
- Avni, O., Lee, D., Macian, F., Szabo, S.J., Glimcher, L.H., and Rao, A. 2002. T(H) cell differentiation is accompanied by dynamic changes in histone acetylation of cytokine genes. *Nat Immunol* **3**(7): 643-651.
- Avril-Sassen, S., Goldstein, L.D., Stingl, J., Blenkiron, C., Le Quesne, J., Spiteri, I., Karagavriliidou, K., Watson, C.J., Tavare, S., Miska, E.A. et al. 2009. Characterisation of microRNA expression in post-natal mouse mammary gland development. *BMC Genomics* **10**: 548.
- Baek, D., Villen, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. 2008. The impact of microRNAs on protein output. *Nature* **455**(7209): 64-71.
- Balaban, N.Q., Schwarz, U.S., Riveline, D., Goichberg, P., Tzur, G., Sabanay, I., Mahalu, D., Safran, S., Bershadsky, A., Addadi, L. et al. 2001. Force and focal adhesion assembly: a close relationship studied using elastic micropatterned substrates. *Nat Cell Biol* **3**: 466-472.
- Balasis, M.E., Forinash, K.D., Chen, Y.A., Fulp, W.J., Coppola, D., Hamilton, A.D., Cheng, J.Q., and Sebt, S.M. 2011. Combination of farnesyltransferase and Akt inhibitors is synergistic in breast cancer cells and causes significant breast tumor regression in ErbB2 transgenic mice. *Clin Cancer Res* **17**(9): 2852-2862.
- Banck, M.S., Li, S., Nishio, H., Wang, C., Beutler, A.S., and Walsh, M.J. 2009. The ZNF217 oncogene is a candidate organizer of repressive histone modifiers. *Epigenetics* **4**(2): 100-106.
- Barker, H.E., Cox, T.R., and Erler, J.T. 2012. The rationale for targeting the LOX family in cancer. *Nat Rev Cancer*.
- Bartel, D.P. 2009. MicroRNAs: target recognition and regulatory functions. *Cell* **136**(2): 215-233.
- Bates, D.L., Chen, Y., Kim, G., Guo, L., and Chen, L. 2008. Crystal structures of multiple GATA zinc fingers bound to DNA reveal new insights into DNA recognition and self-association by GATA. *J Mol Biol* **381**(5): 1292-1306.
- Ben-Porath, I., Thomson, M.W., Carey, V.J., Ge, R., Bell, G.W., Regev, A., and Weinberg, R.A. 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet* **40**(5): 499-507.
- Bergers, G., Brekken, R., McMahon, G., Vu, T.H., Itoh, T., Tamaki, K., Tanzawa, K., Thorpe, P., Itohara, S., Werb, Z. et al. 2000. Matrix metalloproteinase-9 triggers the angiogenic switch during carcinogenesis. *Nat Cell Biol* **2**(10): 737-744.
- Bermudez, O., Pages, G., and Gimond, C. 2010. The dual-specificity MAP kinase phosphatases: critical roles in development and cancer. *Am J Physiol Cell Physiol* **299**(2): C189-202.

- Berndt, N., Yang, H., Trinczek, B., Betzi, S., Zhang, Z., Wu, B., Lawrence, N.J., Pellecchia, M., Schonbrunn, E., Cheng, J.Q. et al. 2010. The Akt activation inhibitor TCN-P inhibits Akt phosphorylation by binding to the PH domain of Akt and blocking its recruitment to the plasma membrane. *Cell Death Differ* **17**(11): 1795-1804.
- Betel, D., Wilson, M., Gabow, A., Marks, D.S., and Sander, C. 2008. The microRNA.org resource: targets and expression. *Nucleic Acids Res* **36**(Database issue): D149-153.
- Bingle, L., Brown, N.J., and Lewis, C.E. 2002. The role of tumour-associated macrophages in tumour progression: implications for new anticancer therapies. *J Pathol* **196**(3): 254-265.
- Birnbaum, D., Adelaide, J., Popovici, C., Charafe-Jauffret, E., Mozziconacci, M.J., and Chaffanet, M. 2003. Chromosome arm 8p and cancer: a fragile hypothesis. *Lancet Oncol* **4**(10): 639-642.
- Blenkiron, C., Goldstein, L.D., Thorne, N.P., Spiteri, I., Chin, S.F., Dunning, M.J., Barbosa-Morais, N.L., Teschendorff, A.E., Green, A.R., Ellis, I.O. et al. 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. *Genome Biol* **8**(10): R214.
- Blick, T., Hugo, H., Widodo, E., Waltham, M., Pinto, C., Mani, S.A., Weinberg, R.A., Neve, R.M., Lenburg, M.E., and Thompson, E.W. 2010. Epithelial mesenchymal transition traits in human breast cancer cell lines parallel the CD44(hi)/CD24 (lo/-) stem cell phenotype in human breast cancer. *J Mammary Gland Biol Neoplasia* **15**(2): 235-252.
- Blokzijl, A., ten Dijke, P., and Ibanez, C.F. 2002. Physical and functional interaction between GATA-3 and Smad3 allows TGF-beta regulation of GATA target genes. *Curr Biol* **12**(1): 35-45.
- Bos, P.D., Zhang, X.H., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A. et al. 2009. Genes that mediate breast cancer metastasis to the brain. *Nature* **459**(7249): 1005-1009.
- Brown, J.D., Dutta, S., Bharti, K., Bonner, R.F., Munson, P.J., Dawid, I.B., Akhtar, A.L., Onojafe, I.F., Alur, R.P., Gross, J.M. et al. 2009. Expression profiling during ocular development identifies 2 Nlz genes with a critical role in optic fissure closure. *Proc Natl Acad Sci U S A* **106**(5): 1462-1467.
- Bruchez, M., Jr., Moronne, M., Gin, P., Weiss, S., and Alivisatos, A.P. 1998. Semiconductor nanocrystals as fluorescent biological labels. *Science* **281**(5385): 2013-2016.
- Bu, Y., Lu, C., Bian, C., Wang, J., Li, J., Zhang, B., Li, Z., Brewer, G., and Zhao, R.C. 2009. Knockdown of Dicer in MCF-7 human breast carcinoma cells results in G1 arrest and increased sensitivity to cisplatin. *Oncol Rep* **21**(1): 13-17.
- Buffa, F.M., Camps, C., Winchester, L., Snell, C.E., Gee, H.E., Sheldon, H., Taylor, M., Harris, A.L., and Ragoussis, J. 2011. microRNA-associated progression pathways and potential therapeutic targets identified by integrated mRNA and microRNA expression profiling in breast cancer. *Cancer Res* **71**(17): 5635-5645.

- Butcher, D.T., Alliston, T., and Weaver, V.M. 2009. A tense situation: Forcing tumour progression. *Nat Rev Cancer* **9**: 108-122.
- Buttitta, L., Tanaka, T.S., Chen, A.E., Ko, M.S., and Fan, C.M. 2003. Microarray analysis of somitogenesis reveals novel targets of different WNT signaling pathways in the somitic mesoderm. *Dev Biol* **258**(1): 91-104.
- Calin, G.A. and Croce, C.M. 2006. MicroRNA signatures in human cancers. *Nat Rev Cancer* **6**(11): 857-866.
- Castano, J., Solanas, G., Casagolda, D., Raurell, I., Villagrasa, P., Bustelo, X.R., Garcia de Herreros, A., and Dunach, M. 2007. Specific phosphorylation of p120-catenin regulatory domain differently modulates its binding to RhoA. *Mol Cell Biol* **27**(5): 1745-1757.
- Chan, W.C. and Nie, S. 1998. Quantum dot bioconjugates for ultrasensitive nonisotopic detection. *Science* **281**(5385): 2016-2018.
- Chang, H.C., Zhang, S., Thieu, V.T., Slee, R.B., Bruns, H.A., Laribee, R.N., Klemsz, M.J., and Kaplan, M.H. 2005. PU.1 expression delineates heterogeneity in primary Th2 cells. *Immunity* **22**(6): 693-703.
- Chapnik, E., Sasson, V., Belloch, R., and Hornstein, E. 2012. Dgcr8 controls neural crest cells survival in cardiovascular development. *Dev Biol* **362**(1): 50-56.
- Chen, C.S., Tan, J., and Tien, J. 2004. Mechanotransduction at cell-matrix and cell-cell contacts. *Annu Rev Biomed Eng* **6**: 275-302.
- Chen, Y.J., Wei, Y.Y., Chen, H.T., Fong, Y.C., Hsu, C.J., Tsai, C.H., Hsu, H.C., Liu, S.H., and Tang, C.H. 2009. Osteopontin increases migration and MMP-9 up-regulation via alphavbeta3 integrin, FAK, ERK, and NF-kappaB-dependent pathway in human chondrosarcoma cells. *J Cell Physiol* **221**(1): 98-108.
- Chen, Z., Wu, J., Yang, C., Fan, P., Balazs, L., Jiao, Y., Lu, M., Gu, W., Li, C., Pfeffer, L.M. et al. 2012. DiGeorge syndrome critical region 8(DGCR8) - mediated miRNA biogenesis is essential for vascular smooth muscle cell development in mice. *J Biol Chem*.
- Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T. et al. 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiology. *Cancer Cell* **10**(6): 529-541.
- Choi, C.L., Koski, K.J., Olson, A.C.K., and Alivisatos, A.P. 2010. Luminescent nanocrystal stress gauge. *Proc Natl Acad Sci USA* **107**(50): 21306-21310.
- Choi, C.L., Koski, K.J., Sivasankar, S., and Alivisatos, A.P. 2009. Strain-dependent photoluminescence behavior of CdSe/CdS nanocrystals with spherical, linear, and branched topologies. *Nano Lett* **9**(10): 3544-3549.
- Chou, J., Provot, S., and Werb, Z. 2010. GATA3 in development and cancer differentiation: cells GATA have it! *J Cell Physiol* **222**(1): 42-49.
- Chu, I.M., Michalowski, A.M., Hoenerhoff, M., Szauter, K.M., Luger, D., Sato, M., Flanders, K., Oshima, A., Csiszar, K., and Green, J.E. 2012. GATA3 inhibits lysyl oxidase-mediated metastases of human basal triple-negative breast cancer cells. *Oncogene* **31**: 2017-2027.

- Cimmino, L., Abdel-Wahab, O., Levine, R.L., and Aifantis, I. 2011. TET family proteins and their role in stem cell differentiation and transformation. *Cell Stem Cell* **9**(3): 193-204.
- Clayclomb, W.C., N. A. Lanson, J., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., and N. J. Izzo, J. 1998. HL-1 cells: A cardiac muscle line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci USA* **95**: 2979-2984.
- Claycomb, W.C., N. A. Lanson, J., Stallworth, B.S., Egeland, D.B., Delcarpio, J.B., Bahinski, A., and N. J. Izzo, J. 1998. HL-1 cells: A cardiac muscle line that contracts and retains phenotypic characteristics of the adult cardiomyocyte. *Proc Natl Acad Sci USA* **95**: 2979-2984.
- Clevers, H. 2006. Wnt/beta-catenin signaling in development and disease. *Cell* **127**(3): 469-480.
- Collins, C., Rommens, J.M., Kowbel, D., Godfrey, T., Tanner, M., Hwang, S.I., Polikoff, D., Nonet, G., Cochran, J., Myambo, K. et al. 1998. Positional cloning of ZNF217 and NABC1: genes amplified at 20q13.2 and overexpressed in breast carcinoma. *Proc Natl Acad Sci U S A* **95**(15): 8703-8708.
- Collins, C., Volik, S., Kowbel, D., Ginzinger, D., Ylstra, B., Cloutier, T., Hawkins, T., Predki, P., Martin, C., Wernick, M. et al. 2001. Comprehensive genome sequence analysis of a breast cancer amplicon. *Genome Res* **11**(6): 1034-1042.
- Colmone, A., Amorim, M., Pontier, A.L., Wang, S., Jablonski, E., and Sipkins, D.A. 2008. Leukemic cells create bone marrow niches that disrupt the behavior of normal hematopoietic progenitor cells. *Science* **322**(5909): 1861-1865.
- Coussens, L.M. and Werb, Z. 2002. Inflammation and cancer. *Nature* **420**(6917): 860-867.
- Cowger, J.J., Zhao, Q., Isovich, M., and Torchia, J. 2007. Biochemical characterization of the zinc-finger protein 217 transcriptional repressor complex: identification of a ZNF217 consensus recognition sequence. *Oncogene* **26**(23): 3378-3386.
- Cowin, P., Rowlands, T.M., and Hatsell, S.J. 2005. Cadherins and catenins in breast cancer. *Curr Opin Cell Biol* **17**(5): 499-508.
- Cross, J.C., Flannery, M.L., Blonar, M.A., Steingrimsson, E., Jenkins, N.A., Copeland, N.G., Rutter, W.J., and Werb, Z. 1995. Hxt encodes a basic helix-loop-helix transcription factor that regulates trophoblast cell development. *Development* **121**(8): 2513-2523.
- Cui, W., Li, Q., Feng, L., and Ding, W. 2011. MiR-126-3p regulates progesterone receptors and involves development and lactation of mouse mammary gland. *Mol Cell Biochem* **355**(1-2): 17-25.
- Cushing, L., Kuang, P.P., Qian, J., Shao, F., Wu, J., Little, F., Thannickal, V.J., Cardoso, W.V., and Lu, J. 2010. miR-29 is a major regulator of genes associated with pulmonary fibrosis. *Am J Respir Cell Mol Biol* **45**(2): 287-294.

- Daniel, J.M. and Reynolds, A.B. 1999. The catenin p120(ctn) interacts with Kaiso, a novel BTB/POZ domain zinc finger transcription factor. *Mol Cell Biol* **19**(5): 3614-3623.
- Daniel, J.M., Spring, C.M., Crawford, H.C., Reynolds, A.B., and Baig, A. 2002. The p120(ctn)-binding partner Kaiso is a bi-modal DNA-binding protein that recognizes both a sequence-specific consensus and methylated CpG dinucleotides. *Nucleic Acids Res* **30**(13): 2911-2919.
- Daniels, D.L. and Weis, W.I. 2005. Beta-catenin directly displaces Groucho/TLE repressors from Tcf/Lef in Wnt-mediated transcription activation. *Nat Struct Mol Biol* **12**(4): 364-371.
- Davis, M.A., Ireton, R.C., and Reynolds, A.B. 2003. A core function for p120-catenin in cadherin turnover. *J Cell Biol* **163**(3): 525-534.
- de Guzman Strong, C., Wertz, P.W., Wang, C., Yang, F., Meltzer, P.S., Andl, T., Millar, S.E., Ho, I.C., Pai, S.Y., and Segre, J.A. 2006. Lipid defect underlies selective skin barrier impairment of an epidermal-specific deletion of Gata-3. *J Cell Biol* **175**(4): 661-670.
- de Visser, K.E., Korets, L.V., and Coussens, L.M. 2005. De novo carcinogenesis promoted by chronic inflammation is B lymphocyte dependent. *Cancer Cell* **7**(5): 411-423.
- Debnath, J. and Brugge, J.S. 2005. Modelling glandular epithelial cancers in three-dimensional cultures. *Nat Rev Cancer* **5**(9): 675-688.
- Dedes, K.J., Natrajan, R., Lambros, M.B., Geyer, F.C., Lopez-Garcia, M.A., Savage, K., Jones, R.L., and Reis-Filho, J.S. 2011. Down-regulation of the miRNA master regulators Drosha and Dicer is associated with specific subgroups of breast cancer. *Eur J Cancer* **47**(1): 138-150.
- DeNardo, D.G., Barreto, J.B., Andreu, P., Vasquez, L., Tawfik, D., Kolhatkar, N., and Coussens, L.M. 2009. CD4(+) T cells regulate pulmonary metastasis of mammary carcinomas by enhancing protumor properties of macrophages. *Cancer Cell* **16**(2): 91-102.
- Deng, J., Liu, Y., Lee, H., Herrmann, A., Zhang, W., Zhang, C., Shen, S., Priceman, S.J., Kujawski, M., Pal, S.K. et al. 2012. S1PR1-STAT3 Signaling Is Crucial for Myeloid Cell Colonization at Future Metastatic Sites. *Cancer Cell* **21**(5): 642-654.
- 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.
- Desmedt, C., Piette, F., Loi, S., Wang, Y., Lallemand, F., Haibe-Kains, B., Viale, G., Delorenzi, M., Zhang, Y., d'Assignies, M.S. et al. 2007. Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series. *Clin Cancer Res* **13**(11): 3207-3214.
- DiMeo, T.A., Anderson, K., Phadke, P., Fan, C., Perou, C.M., Naber, S., and Kuperwasser, C. 2009. A novel lung metastasis signature links Wnt signaling with cancer cell self-renewal and epithelial-mesenchymal transition in basal-like breast cancer. *Cancer Res* **69**(13): 5364-5373.

- Dolled-Filhart, M., Ryden, L., Cregger, M., Jirstrom, K., Harigopal, M., Camp, R.L., and Rimm, D.L. 2006. Classification of breast cancer using genetic algorithms and tissue microarrays. *Clin Cancer Res* **12**(21): 6459-6468.
- Dore, L.C., Amigo, J.D., Dos Santos, C.O., Zhang, Z., Gai, X., Tobias, J.W., Yu, D., Klein, A.M., Dorman, C., Wu, W. et al. 2008. A GATA-1-regulated microRNA locus essential for erythropoiesis. *Proc Natl Acad Sci U S A* **105**(9): 3333-3338.
- Dorfman, R., Glazer, L., Weihe, U., Wernet, M.F., and Shilo, B.Z. 2002. Elbow and Noc define a family of zinc finger proteins controlling morphogenesis of specific tracheal branches. *Development* **129**(15): 3585-3596.
- Du, R., Lu, K.V., Petritsch, C., Liu, P., Ganss, R., Passegue, E., Song, H., Vandenberg, S., Johnson, R.S., Werb, Z. et al. 2008. HIF1alpha induces the recruitment of bone marrow-derived vascular modulatory cells to regulate tumor angiogenesis and invasion. *Cancer Cell* **13**(3): 206-220.
- Du, Z., Podsypanina, K., Huang, S., McGrath, A., Toneff, M.J., Bogoslovskaja, E., Zhang, X., Moraes, R.C., Fluck, M., Allred, D.C. et al. 2006. Introduction of oncogenes into mammary glands in vivo with an avian retroviral vector initiates and promotes carcinogenesis in mouse models. *Proc Natl Acad Sci U S A* **103**(46): 17396-17401.
- Dumont, N. and Arteaga, C.L. 2000. Transforming growth factor-beta and breast cancer: Tumor promoting effects of transforming growth factor-beta. *Breast Cancer Res* **2**(2): 125-132.
- Dydensborg, A.B., Rose, A.A., Wilson, B.J., Grote, D., Paquet, M., Giguere, V., Siegel, P.M., and Bouchard, M. 2009. GATA3 inhibits breast cancer growth and pulmonary breast cancer metastasis. *Oncogene* **28**(29): 2634-2642.
- Dyxhoorn, D.M., Wu, Y., Xie, H., Yu, F., Lal, A., Petrocca, F., Martinvalet, D., Song, E., Lim, B., and Lieberman, J. 2009. miR-200 enhances mouse breast cancer cell colonization to form distant metastases. *PLoS One* **4**(9): e7181.
- Egeblad, M., Nakasone, E.S., and Werb, Z. 2010. Tumors as organs: complex tissues that interface with the entire organism. *Dev Cell* **18**(6): 884-901.
- Elenbaas, B., Spirio, L., Koerner, F., Fleming, M.D., Zimonjic, D.B., Donaher, J.L., Popescu, N.C., Hahn, W.C., and Weinberg, R.A. 2001. Human breast cancer cells generated by oncogenic transformation of primary mammary epithelial cells. *Genes Dev* **15**(1): 50-65.
- Enerly, E., Steinfeld, I., Kleivi, K., Leivonen, S.K., Aure, M.R., Russnes, H.G., Ronneberg, J.A., Johnsen, H., Navon, R., Rodland, E. et al. 2011. miRNA-mRNA integrated analysis reveals roles for miRNAs in primary breast tumors. *PLoS One* **6**(2): e16915.
- Engelhardt, J.J., Boldajipour, B., Beemiller, P., Pandurangi, P., Sorensen, C., Werb, Z., Egeblad, M., and Krummel, M.F. 2011. Marginating dendritic cells of the tumor microenvironment cross-present tumor antigens and stably engage tumor-specific T cells. *Cancer Cell* **21**(3): 402-417.
- Engler, A.J., Griffin, M.A., Sen, S., Bonnemann, C.G., Sweeney, H.L., and Discher, D.E. 2004. Myotubes differentiate optimally on substrates with

- tissue-like stiffness: pathological implications for soft or stiff microenvironments. *J Cell Biol* **166**: 877-887.
- Erler, J.T., Bennewith, K.L., Cox, T.R., Lang, G., Bird, D., Koong, A., Le, Q.T., and Giaccia, A.J. 2009. Hypoxia-induced lysyl oxidase is a critical mediator of bone marrow cell recruitment to form the premetastatic niche. *Cancer Cell* **15**(1): 35-44.
- Erler, J.T., Bennewith, K.L., Nicolau, M., Dornhofer, N., Kong, C., Le, Q.T., Chi, J.T., Jeffrey, S.S., and Giaccia, A.J. 2006. Lysyl oxidase is essential for hypoxia-induced metastasis. *Nature* **440**(7088): 1222-1226.
- Ewald, A.J., Brenot, A., Duong, M., Chan, B.S., and Werb, Z. 2008. Collective epithelial migration and cell rearrangements drive mammary branching morphogenesis. *Dev Cell* **14**(4): 570-581.
- Fabbri, M., Garzon, R., Cimmino, A., Liu, Z., Zanesi, N., Callegari, E., Liu, S., Alder, H., Costinean, S., Fernandez-Cymering, C. et al. 2007. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. *Proc Natl Acad Sci U S A* **104**(40): 15805-15810.
- Faggad, A., Budczies, J., Tchernitsa, O., Darb-Esfahani, S., Sehouli, J., Muller, B.M., Wirtz, R., Chekerov, R., Weichert, W., Sinn, B. et al. 2010. Prognostic significance of Dicer expression in ovarian cancer-link to global microRNA changes and oestrogen receptor expression. *J Pathol* **220**(3): 382-391.
- Fang, T.C., Yashiro-Ohtani, Y., Del Bianco, C., Knoblock, D.M., Blacklow, S.C., and Pear, W.S. 2007. Notch directly regulates Gata3 expression during T helper 2 cell differentiation. *Immunity* **27**(1): 100-110.
- Farnham, P.J. 2009. Insights from genomic profiling of transcription factors. *Nat Rev Genet* **10**(9): 605-616.
- Feun, L.G., Savaraj, N., Bodey, G.P., Lu, K., Yap, B.S., Ajani, J.A., Burgess, M.A., Benjamin, R.S., McKelvey, E., and Krakoff, I. 1984. Phase I study of tricyclic nucleoside phosphate using a five-day continuous infusion schedule. *Cancer Res* **44**(8): 3608-3612.
- Ficz, G., Branco, M.R., Seisenberger, S., Santos, F., Krueger, F., Hore, T.A., Marques, C.J., Andrews, S., and Reik, W. 2011. Dynamic regulation of 5-hydroxymethylcytosine in mouse ES cells and during differentiation. *Nature* **473**(7347): 398-402.
- Fidler, I.J. 2003. The pathogenesis of cancer metastasis: the 'seed and soil' hypothesis revisited. *Nat Rev Cancer* **3**(6): 453-458.
- Fiore, A., Mastria, R., Lupo, M.G., Lanzani, G., Giannini, C., Carlino, E., Morello, G., Giorgi, M.D., Li, Y., Cingolani, R. et al. 2009. Tetrapod-shaped colloidal nanocrystals of II-VI semiconductors prepared by seeded growth. *J Am Chem Soc* **131**: 2274-2282.
- Foulkes, W.D., Smith, I.E., and Reis-Filho, J.S. 2010. Triple-negative breast cancer. *N Engl J Med* **363**(20): 1938-1948.
- Frame, M.C. 2002. Src in cancer: deregulation and consequences for cell behaviour. *Biochim Biophys Acta* **1602**(2): 114-130.

- Franco, O.E., Shaw, A.K., Strand, D.W., and Hayward, S.W. 2009. Cancer associated fibroblasts in cancer pathogenesis. *Semin Cell Dev Biol* **21**(1): 33-39.
- Fu, J., Tang, W., Du, P., Wang, G., Chen, W., Li, J., Zhu, Y., Gao, J., and Cui, L. 2012. Identifying MicroRNA-mRNA regulatory network in colorectal cancer by a combination of expression profile and bioinformatics analysis. *BMC Syst Biol* **6**(1): 68.
- Fuchs, E. 2007. Scratching the surface of skin development. *Nature* **445**(7130): 834-842.
- Fujiwara, T., O'Geen, H., Keles, S., Blahnik, K., Linnemann, A.K., Kang, Y.A., Choi, K., Farnham, P.J., and Bresnick, E.H. 2009. Discovering hematopoietic mechanisms through genome-wide analysis of GATA factor chromatin occupancy. *Mol Cell* **36**(4): 667-681.
- Garcia, M.J., Pole, J.C., Chin, S.F., Teschendorff, A., Naderi, A., Ozdag, H., Vias, M., Kranjac, T., Subkhankulova, T., Paish, C. et al. 2005. A 1 Mb minimal amplicon at 8p11-12 in breast cancer identifies new candidate oncogenes. *Oncogene* **24**(33): 5235-5245.
- Garrett, C.R., Coppola, D., Wenham, R.M., Cubitt, C.L., Neuger, A.M., Frost, T.J., Lush, R.M., Sullivan, D.M., Cheng, J.Q., and Sebt, S.M. 2011. Phase I pharmacokinetic and pharmacodynamic study of triciribine phosphate monohydrate, a small-molecule inhibitor of AKT phosphorylation, in adult subjects with solid tumors containing activated AKT. *Invest New Drugs* **29**(6): 1381-1389.
- Gasperowicz, M. and Otto, F. 2005. Mammalian Groucho homologs: redundancy or specificity? *J Cell Biochem* **95**(4): 670-687.
- Gebhardt, C., Nemeth, J., Angel, P., and Hess, J. 2006. S100A8 and S100A9 in inflammation and cancer. *Biochem Pharmacol* **72**(11): 1622-1631.
- Gelsi-Boyer, V., Orsetti, B., Cervera, N., Finetti, P., Sircoulomb, F., Rouge, C., Lasorsa, L., Letessier, A., Ginestier, C., Monville, F. et al. 2005. Comprehensive profiling of 8p11-12 amplification in breast cancer. *Mol Cancer Res* **3**(12): 655-667.
- Goldfarb, D.S., Corbett, A.H., Mason, D.A., Harreman, M.T., and Adam, S.A. 2004. Importin alpha: a multipurpose nuclear-transport receptor. *Trends Cell Biol* **14**(9): 505-514.
- Gough, P.J., Gomez, I.G., Wille, P.T., and Raines, E.W. 2006. Macrophage expression of active MMP-9 induces acute plaque disruption in apoE-deficient mice. *J Clin Invest* **116**(1): 59-69.
- Grande-Garcia, A., Echarri, A., de Rooij, J., Alderson, N.B., Waterman-Storer, C.M., Valdivielso, J.M., and del Pozo, M.A. 2007. Caveolin-1 regulates cell polarization and directional migration through Src kinase and Rho GTPases. *J Cell Biol* **177**(4): 683-694.
- Green, J.E., Shibata, M.A., Yoshidome, K., Liu, M.L., Jorcyk, C., Anver, M.R., Wigginton, J., Wiltrot, R., Shibata, E., Kaczmarczyk, S. et al. 2000. The C3(1)/SV40 T-antigen transgenic mouse model of mammary cancer: ductal epithelial cell targeting with multistage progression to carcinoma. *Oncogene* **19**(8): 1020-1027.

- Greene, S.B., Gunaratne, P.H., Hammond, S.M., and Rosen, J.M. 2010. A putative role for microRNA-205 in mammary epithelial cell progenitors. *J Cell Sci* **123**(Pt 4): 606-618.
- Grelier, G., Voirin, N., Ay, A.S., Cox, D.G., Chabaud, S., Treilleux, I., Leon-Goddard, S., Rimokh, R., Mikaelian, I., Venoux, C. et al. 2009. Prognostic value of Dicer expression in human breast cancers and association with the mesenchymal phenotype. *Br J Cancer* **101**(4): 673-683.
- Grogan, J.L. and Locksley, R.M. 2002. T helper cell differentiation: on again, off again. *Curr Opin Immunol* **14**(3): 366-372.
- Grote, D., Boualia, S.K., Souabni, A., Merkel, C., Chi, X., Costantini, F., Carroll, T., and Bouchard, M. 2008. Gata3 acts downstream of beta-catenin signaling to prevent ectopic metanephric kidney induction. *PLoS Genet* **4**(12): e1000316.
- Gudjonsson, T., Villadsen, R., Nielsen, H.L., Ronnov-Jessen, L., Bissell, M.J., and Petersen, O.W. 2002. Isolation, immortalization, and characterization of a human breast epithelial cell line with stem cell properties. *Genes Dev* **16**(6): 693-706.
- Gumbiner, B.M. 2005. Regulation of cadherin-mediated adhesion in morphogenesis. *Nat Rev Mol Cell Biol* **6**(8): 622-634.
- Guo, W., Keckesova, Z., Donaher, J.L., Shibue, T., Tischler, V., Reinhardt, F., Itzkovitz, S., Noske, A., Zurrer-Hardi, U., Bell, G. et al. 2012. Slug and Sox9 cooperatively determine the mammary stem cell state. *Cell* **148**(5): 1015-1028.
- Gupta, G.P. and Massague, J. 2006. Cancer metastasis: building a framework. *Cell* **127**(4): 679-695.
- Gupta, G.P., Perk, J., Acharyya, S., de Candia, P., Mittal, V., Todorova-Manova, K., Gerald, W.L., Brogi, E., Benezra, R., and Massague, J. 2007. ID genes mediate tumor reinitiation during breast cancer lung metastasis. *Proc Natl Acad Sci U S A* **104**(49): 19506-19511.
- Gupta, N. 1999. Acousto-optic tunable filter-based hyperspectral and polarization images for fluorescence and spectroscopic imaging. *Methods in Molecular Biology: Biosensors and Biodetection* **503**: 293-305.
- Gupta, P.B., Onder, T.T., Jiang, G., Tao, K., Kuperwasser, C., Weinberg, R.A., and Lander, E.S. 2009. Identification of selective inhibitors of cancer stem cells by high-throughput screening. *Cell* **138**(4): 645-659.
- Gupta, R.A., Shah, N., Wang, K.C., Kim, J., Horlings, H.M., Wong, D.J., Tsai, M.C., Hung, T., Argani, P., Rinn, J.L. et al. 2010. Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. *Nature* **464**(7291): 1071-1076.
- Guy, C.T., Cardiff, R.D., and Muller, W.J. 1992. Induction of mammary tumors by expression of polyomavirus middle T oncogene: a transgenic mouse model for metastatic disease. *Mol Cell Biol* **12**(3): 954-961.
- Halpern, J., Lynch, C.C., Fleming, J., Hamming, D., Martin, M.D., Schwartz, H.S., Matrisian, L.M., and Holt, G.E. 2006. The application of a murine bone bioreactor as a model of tumor: bone interaction. *Clin Exp Metastasis* **23**(7-8): 345-356.

- Hanahan, D. and Coussens, L.M. 2012. Accessories to the crime: functions of cells recruited to the tumor microenvironment. *Cancer Cell* **21**(3): 309-322.
- Hanahan, D. and Weinberg, R.A. 2000. The hallmarks of cancer. *Cell* **100**(1): 57-70.
- . 2011. Hallmarks of cancer: the next generation. *Cell* **144**(5): 646-674.
- Harburg, G.C. and Hinck, L. 2011. Navigating breast cancer: axon guidance molecules as breast cancer tumor suppressors and oncogenes. *J Mammary Gland Biol Neoplasia* **16**(3): 257-270.
- Hess, K.R., Anderson, K., Symmans, W.F., Valero, V., Ibrahim, N., Mejia, J.A., Booser, D., Theriault, R.L., Buzdar, A.U., Dempsey, P.J. et al. 2006. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24**(26): 4236-4244.
- Hiratsuka, S., Nakamura, K., Iwai, S., Murakami, M., Itoh, T., Kijima, H., Shipley, J.M., Senior, R.M., and Shibuya, M. 2002. MMP9 induction by vascular endothelial growth factor receptor-1 is involved in lung-specific metastasis. *Cancer Cell* **2**(4): 289-300.
- Hiratsuka, S., Watanabe, A., Aburatani, H., and Maru, Y. 2006. Tumour-mediated upregulation of chemoattractants and recruitment of myeloid cells predetermines lung metastasis. *Nat Cell Biol* **8**(12): 1369-1375.
- Hiratsuka, S., Watanabe, A., Sakurai, Y., Akashi-Takamura, S., Ishibashi, S., Miyake, K., Shibuya, M., Akira, S., Aburatani, H., and Maru, Y. 2008. The S100A8-serum amyloid A3-TLR4 paracrine cascade establishes a pre-metastatic phase. *Nat Cell Biol* **10**(11): 1349-1355.
- Ho, I.C. and Pai, S.Y. 2007. GATA-3 - not just for Th2 cells anymore. *Cell Mol Immunol* **4**(1): 15-29.
- Ho, I.C., Tai, T.S., and Pai, S.Y. 2009. GATA3 and the T-cell lineage: essential functions before and after T-helper-2-cell differentiation. *Nat Rev Immunol* **9**(2): 125-135.
- Hoch, R.V., Thompson, D.A., Baker, R.J., and Weigel, R.J. 1999. GATA-3 is expressed in association with estrogen receptor in breast cancer. *Int J Cancer* **84**(2): 122-128.
- Hoeflich, K.P., O'Brien, C., Boyd, Z., Cavet, G., Guerrero, S., Jung, K., Januario, T., Savage, H., Punnoose, E., Truong, T. et al. 2009. In vivo antitumor activity of MEK and phosphatidylinositol 3-kinase inhibitors in basal-like breast cancer models. *Clin Cancer Res* **15**(14): 4649-4664.
- Hoffman, K., Holmes, F.A., Frascini, G., Esparza, L., Frye, D., Raber, M.N., Newman, R.A., and Hortobagyi, G.N. 1996. Phase I-II study: tricyclic nucleoside phosphate for metastatic breast cancer. *Cancer Chemother Pharmacol* **37**(3): 254-258.
- Holland, D.G., Burleigh, A., Git, A., Goldgraben, M.A., Perez-Mancera, P.A., Chin, S.F., Hurtado, A., Bruna, A., Ali, H.R., Greenwood, W. et al. 2011. ZNF703 is a common Luminal B breast cancer oncogene that differentially regulates luminal and basal progenitors in human mammary epithelium. *EMBO Mol Med* **3**(3): 167-180.

- Hollestelle, A., Nagel, J.H., Smid, M., Lam, S., Elstrodt, F., Wasielewski, M., Ng, S.S., French, P.J., Peeters, J.K., Rozendaal, M.J. et al. 2010. Distinct gene mutation profiles among luminal-type and basal-type breast cancer cell lines. *Breast Cancer Res Treat* **121**(1): 53-64.
- Hong, S.J., Choi, H.J., Hong, S., Huh, Y., Chae, H., and Kim, K.S. 2008. Transcription factor GATA-3 regulates the transcriptional activity of dopamine beta-hydroxylase by interacting with Sp1 and AP4. *Neurochem Res* **33**(9): 1821-1831.
- Hoyle, J., Tang, Y.P., Wieltete, E.L., Wardle, F.C., and Sive, H. 2004. nlz gene family is required for hindbrain patterning in the zebrafish. *Dev Dyn* **229**(4): 835-846.
- Hsu, Y.C. and Fuchs, E. 2012. A family business: stem cell progeny join the niche to regulate homeostasis. *Nat Rev Mol Cell Biol* **13**(2): 103-114.
- Huang, G., Krig, S., Kowbel, D., Xu, H., Hyun, B., Volik, S., Feuerstein, B., Mills, G.B., Stokoe, D., Yaswen, P. et al. 2005. ZNF217 suppresses cell death associated with chemotherapy and telomere dysfunction. *Hum Mol Genet* **14**(21): 3219-3225.
- Huang, P., Duda, D.G., Jain, R.K., and Fukumura, D. 2008. Histopathologic findings and establishment of novel tumor lines from spontaneous tumors in FVB/N mice. *Comp Med* **58**(3): 253-263.
- Hurst, D.R., Edmonds, M.D., and Welch, D.R. 2009. Metastamir: the field of metastasis-regulatory microRNA is spreading. *Cancer Res* **69**(19): 7495-7498.
- Husemann, Y., Geigl, J.B., Schubert, F., Musiani, P., Meyer, M., Burghart, E., Forni, G., Eils, R., Fehm, T., Riethmuller, G. et al. 2008. Systemic spread is an early step in breast cancer. *Cancer Cell* **13**(1): 58-68.
- Hutvagner, G., McLachlan, J., Pasquinelli, A.E., Balint, E., Tuschl, T., and Zamore, P.D. 2001. A cellular function for the RNA-interference enzyme Dicer in the maturation of the let-7 small temporal RNA. *Science* **293**(5531): 834-838.
- Hwang, H.W., Wentzel, E.A., and Mendell, J.T. 2007. A hexanucleotide element directs microRNA nuclear import. *Science* **315**(5808): 97-100.
- Hynes, R.O. 2002. Integrins: bidirectional, allosteric signaling machines. *Cell* **110**(6): 673-687.
- 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.
- Iguchi, H., Urashima, Y., Inagaki, Y., Ikeda, Y., Okamura, M., Tanaka, T., Uchida, A., Yamamoto, T.T., Kodama, T., and Sakai, J. 2007. SOX6 suppresses cyclin D1 promoter activity by interacting with beta-catenin and histone deacetylase 1, and its down-regulation induces pancreatic beta-cell proliferation. *J Biol Chem* **282**(26): 19052-19061.
- Ishiyama, N., Lee, S.H., Liu, S., Li, G.Y., Smith, M.J., Reichardt, L.F., and Ikura, M. 2010. Dynamic and static interactions between p120 catenin and E-cadherin regulate the stability of cell-cell adhesion. *Cell* **141**(1): 117-128.

- Israel, A., Sharan, R., Ruppin, E., and Galun, E. 2009. Increased microRNA activity in human cancers. *PLoS One* **4**(6): e6045.
- Ito, S., D'Alessio, A.C., Taranova, O.V., Hong, K., Sowers, L.C., and Zhang, Y. 2010. Role of Tet proteins in 5mC to 5hmC conversion, ES-cell self-renewal and inner cell mass specification. *Nature* **466**(7310): 1129-1133.
- Ivshina, A.V., George, J., Senko, O., Mow, B., Putti, T.C., Smeds, J., Lindahl, T., Pawitan, Y., Hall, P., Nordgren, H. et al. 2006. Genetic reclassification of histologic grade delineates new clinical subtypes of breast cancer. *Cancer Res* **66**(21): 10292-10301.
- Jacquemier, J., Charafe-Jauffret, E., Monville, F., Esterni, B., Extra, J.M., Houvenaeghel, G., Xerri, L., Bertucci, F., and Birnbaum, D. 2009. Association of GATA3, P53, Ki67 status and vascular peritumoral invasion are strongly prognostic in luminal breast cancer. *Breast Cancer Res* **11**(2): R23.
- Jamora, C., DasGupta, R., Kocieniewski, P., and Fuchs, E. 2003. Links between signal transduction, transcription and adhesion in epithelial bud development. *Nature* **422**(6929): 317-322.
- Jenssen, T.K., Kuo, W.P., Stokke, T., and Hovig, E. 2002. Associations between gene expressions in breast cancer and patient survival. *Hum Genet* **111**(4-5): 411-420.
- Jones, J.M. and Warchol, M.E. 2009. Expression of the Gata3 transcription factor in the acoustic ganglion of the developing avian inner ear. *J Comp Neurol* **516**(6): 507-518.
- Jones, S.E. 2008. Metastatic breast cancer: the treatment challenge. *Clin Breast Cancer* **8**(3): 224-233.
- Joyce, J.A. and Pollard, J.W. 2009. Microenvironmental regulation of metastasis. *Nat Rev Cancer* **9**(4): 239-252.
- Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massague, J. 2003. A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* **3**(6): 537-549.
- Kaplan, R.N., Riba, R.D., Zacharoulis, S., Bramley, A.H., Vincent, L., Costa, C., MacDonald, D.D., Jin, D.K., Shido, K., Kerns, S.A. et al. 2005. VEGFR1-positive haematopoietic bone marrow progenitors initiate the pre-metastatic niche. *Nature* **438**(7069): 820-827.
- Kaufman, C.K., Zhou, P., Pasolli, H.A., Rendl, M., Bolotin, D., Lim, K.C., Dai, X., Alegre, M.L., and Fuchs, E. 2003. GATA-3: an unexpected regulator of cell lineage determination in skin. *Genes Dev* **17**(17): 2108-2122.
- Keirsebilck, A., Bonne, S., Staes, K., van Hengel, J., Nollet, F., Reynolds, A., and van Roy, F. 1998. Molecular cloning of the human p120ctn catenin gene (CTNND1): expression of multiple alternatively spliced isoforms. *Genomics* **50**(2): 129-146.
- Kessenbrock, K., Plaks, V., and Werb, Z. 2010. Matrix metalloproteinases: regulators of the tumor microenvironment. *Cell* **141**(1): 52-67.
- Keyse, S.M. 2008. Dual-specificity MAP kinase phosphatases (MKPs) and cancer. *Cancer Metastasis Rev* **27**(2): 253-261.

- Khokha, R. and Werb, Z. 2010. Mammary gland reprogramming: metalloproteinases couple form with function. *Cold Spring Harb Perspect Biol* **3**(4).
- Kim, S., Takahashi, H., Lin, W.W., Descargues, P., Grivennikov, S., Kim, Y., Luo, J.L., and Karin, M. 2009. Carcinoma-produced factors activate myeloid cells through TLR2 to stimulate metastasis. *Nature* **457**(7225): 102-106.
- Koh, K.P., Yabuuchi, A., Rao, S., Huang, Y., Cunniff, K., Nardone, J., Laiho, A., Tahiliani, M., Sommer, C.A., Mostoslavsky, G. et al. 2011. Tet1 and Tet2 regulate 5-hydroxymethylcytosine production and cell lineage specification in mouse embryonic stem cells. *Cell Stem Cell* **8**(2): 200-213.
- Korkaya, H., Paulson, A., Charafe-Jauffret, E., Ginestier, C., Brown, M., Dutcher, J., Clouthier, S.G., and Wicha, M.S. 2009. Regulation of mammary stem/progenitor cells by PTEN/Akt/beta-catenin signaling. *PLoS Biol* **7**(6): e1000121.
- Kota, J., Chivukula, R.R., O'Donnell, K.A., Wentzel, E.A., Montgomery, C.L., Hwang, H.W., Chang, T.C., Vivekanandan, P., Torbenson, M., Clark, K.R. et al. 2009. Therapeutic microRNA delivery suppresses tumorigenesis in a murine liver cancer model. *Cell* **137**(6): 1005-1017.
- Kouros-Mehr, H., Bechis, S.K., Slorach, E.M., Littlepage, L.E., Egeblad, M., Ewald, A.J., Pai, S.Y., Ho, I.C., and Werb, Z. 2008. GATA-3 links tumor differentiation and dissemination in a luminal breast cancer model. *Cancer Cell* **13**(2): 141-152.
- Kouros-Mehr, H., Slorach, E.M., Sternlicht, M.D., and Werb, Z. 2006. GATA-3 maintains the differentiation of the luminal cell fate in the mammary gland. *Cell* **127**(5): 1041-1055.
- Kouros-Mehr, H. and Werb, Z. 2006. Candidate regulators of mammary branching morphogenesis identified by genome-wide transcript analysis. *Dev Dyn* **235**(12): 3404-3412.
- Kowanetz, M., Wu, X., Lee, J., Tan, M., Hagenbeek, T., Qu, X., Yu, L., Ross, J., Korsisaari, N., Cao, T. et al. 2010. Granulocyte-colony stimulating factor promotes lung metastasis through mobilization of Ly6G+Ly6C+ granulocytes. *Proc Natl Acad Sci U S A* **107**(50): 21248-21255.
- Krek, A., Grun, D., Poy, M.N., Wolf, R., Rosenberg, L., Epstein, E.J., MacMenamin, P., da Piedade, I., Gunsalus, K.C., Stoffel, M. et al. 2005. Combinatorial microRNA target predictions. *Nat Genet* **37**(5): 495-500.
- Krig, S.R., Jin, V.X., Bieda, M.C., O'Geen, H., Yaswen, P., Green, R., and Farnham, P.J. 2007. Identification of genes directly regulated by the oncogene ZNF217 using chromatin immunoprecipitation (ChIP)-chip assays. *J Biol Chem* **282**(13): 9703-9712.
- Krig, S.R., Miller, J.K., Fietze, S., Beckett, L.A., Neve, R.M., Farnham, P.J., Yaswen, P.I., and Sweeney, C.A. 2010. ZNF217, a candidate breast cancer oncogene amplified at 20q13, regulates expression of the ErbB3 receptor tyrosine kinase in breast cancer cells. *Oncogene* **29**(40): 5500-5510.
- Kufe, D.W. and Bast, R.C. 2003. *Cancer medicine*. B C Decker, Hamilton, Ont. ; London.

- Kuilman, T., Michaloglou, C., Mooi, W.J., and Peeper, D.S. 2010. The essence of senescence. *Genes Dev* **24**(22): 2463-2479.
- Kumar, M.S., Pester, R.E., Chen, C.Y., Lane, K., Chin, C., Lu, J., Kirsch, D.G., Golub, T.R., and Jacks, T. 2009. Dicer1 functions as a haploinsufficient tumor suppressor. *Genes Dev* **23**(23): 2700-2704.
- Kurek, D., Garinis, G.A., van Doorninck, J.H., van der Wees, J., and Grosveld, F.G. 2007. Transcriptome and phenotypic analysis reveals Gata3-dependent signalling pathways in murine hair follicles. *Development* **134**(2): 261-272.
- Lammi, J., Huppunen, J., and Aarnisalo, P. 2004. Regulation of the osteopontin gene by the orphan nuclear receptor NURR1 in osteoblasts. *Mol Endocrinol* **18**(6): 1546-1557.
- Landthaler, M., Yalcin, A., and Tuschl, T. 2004. The human DiGeorge syndrome critical region gene 8 and its D. melanogaster homolog are required for miRNA biogenesis. *Curr Biol* **14**(23): 2162-2167.
- Lathia, J.D., Heddleston, J.M., Venere, M., and Rich, J.N. 2011. Deadly teamwork: neural cancer stem cells and the tumor microenvironment. *Cell Stem Cell* **8**(5): 482-485.
- Lee, G.R., Fields, P.E., and Flavell, R.A. 2001. Regulation of IL-4 gene expression by distal regulatory elements and GATA-3 at the chromatin level. *Immunity* **14**(4): 447-459.
- Lee, G.Y., Kenny, P.A., Lee, E.H., and Bissell, M.J. 2007. Three-dimensional culture models of normal and malignant breast epithelial cells. *Nat Methods* **4**(4): 359-365.
- Legant, W.R., Miller, J.S., Blakely, B.L., Cohen, D.M., Genin, G.M., and Chen, C.S. 2010. Measurement of mechanical tractions exerted by cells in three-dimensional matrices. *Nat Methods* **7**(12): 969-971.
- Levental, K.R., Yu, H., Kass, L., Lakins, J.N., Egeblad, M., Erler, J.T., Fong, S.F., Csiszar, K., Giaccia, A., Weninger, W. et al. 2009. Matrix crosslinking forces tumor progression by enhancing integrin signaling. *Cell* **139**(5): 891-906.
- Lewis, B.P., Burge, C.B., and Bartel, D.P. 2005. Conserved seed pairing, often flanked by adenosines, indicates that thousands of human genes are microRNA targets. *Cell* **120**(1): 15-20.
- Li, P., Maines-Bandiera, S., Kuo, W.L., Guan, Y., Sun, Y., Hills, M., Huang, G., Collins, C.C., Leung, P.C., Gray, J.W. et al. 2007. Multiple roles of the candidate oncogene ZNF217 in ovarian epithelial neoplastic progression. *Int J Cancer* **120**(9): 1863-1873.
- Li, X., Lewis, M.T., Huang, J., Gutierrez, C., Osborne, C.K., Wu, M.F., Hilsenbeck, S.G., Pavlick, A., Zhang, X., Chamness, G.C. et al. 2008. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. *J Natl Cancer Inst* **100**(9): 672-679.
- Li, Y., Welm, B., Podsypanina, K., Huang, S., Chamorro, M., Zhang, X., Rowlands, T., Egeblad, M., Cowin, P., Werb, Z. et al. 2003. Evidence that transgenes encoding components of the Wnt signaling pathway

- preferentially induce mammary cancers from progenitor cells. *Proc Natl Acad Sci U S A* **100**(26): 15853-15858.
- Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A. et al. 2009. Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat Med* **15**(8): 907-913.
- Lim, K.C., Lakshmanan, G., Crawford, S.E., Gu, Y., Grosveld, F., and Engel, J.D. 2000. Gata3 loss leads to embryonic lethality due to noradrenaline deficiency of the sympathetic nervous system. *Nat Genet* **25**(2): 209-212.
- Lin, E.Y., Jones, J.G., Li, P., Zhu, L., Whitney, K.D., Muller, W.J., and Pollard, J.W. 2003. Progression to malignancy in the polyoma middle T oncoprotein mouse breast cancer model provides a reliable model for human diseases. *Am J Pathol* **163**(5): 2113-2126.
- Lin, E.Y., Li, J.F., Gnatovskiy, L., Deng, Y., Zhu, L., Grzesik, D.A., Qian, H., Xue, X.N., and Pollard, J.W. 2006. Macrophages regulate the angiogenic switch in a mouse model of breast cancer. *Cancer Res* **66**(23): 11238-11246.
- Lin, E.Y., Nguyen, A.V., Russell, R.G., and Pollard, J.W. 2001. Colony-stimulating factor 1 promotes progression of mammary tumors to malignancy. *J Exp Med* **193**(6): 727-740.
- Lin, H. 2002. The stem-cell niche theory: lessons from flies. *Nat Rev Genet* **3**(12): 931-940.
- Lin, S.C., Chien, C.W., Lee, J.C., Yeh, Y.C., Hsu, K.F., Lai, Y.Y., and Tsai, S.J. 2011. Suppression of dual-specificity phosphatase-2 by hypoxia increases chemoresistance and malignancy in human cancer cells. *J Clin Invest* **121**(5): 1905-1916.
- Linnemann, A.K., O'Geen, H., Keles, S., Farnham, P.J., and Bresnick, E.H. 2011. Genetic framework for GATA factor function in vascular biology. *Proc Natl Acad Sci U S A* **108**(33): 13641-13646.
- Littlepage, L.E., Adler, A.S., Kouros-Mehr, H., Huang, G., Chou, J., Krig, S.R., Griffith, O.L., Korkola, J.E., Qu, K., Lawson, D.A. et al. 2012. The Transcription Factor ZNF217 Is a Prognostic Biomarker and Therapeutic Target during Breast Cancer Progression. *Cancer Discov* **2**(7): 638-651.
- Liu, S., Yamashita, H., Weidow, B., Weaver, A.M., and Quaranta, V. 2010. Laminin-332-beta1 integrin interactions negatively regulate invadopodia. *J Cell Physiol* **223**(1): 134-142.
- Loi, S., Haibe-Kains, B., Desmedt, C., Lallemand, F., Tutt, A.M., Gillet, C., Ellis, P., Harris, A., Bergh, J., Foekens, J.A. et al. 2007. Definition of clinically distinct molecular subtypes in estrogen receptor-positive breast carcinomas through genomic grade. *J Clin Oncol* **25**(10): 1239-1246.
- Losick, V.P., Morris, L.X., Fox, D.T., and Spradling, A. 2011. Drosophila stem cell niches: a decade of discovery suggests a unified view of stem cell regulation. *Dev Cell* **21**(1): 159-171.
- Lu, C.H., Wyszomierski, S.L., Tseng, L.M., Sun, M.H., Lan, K.H., Neal, C.L., Mills, G.B., Hortobagyi, G.N., Esteva, F.J., and Yu, D. 2007. Preclinical testing of clinically applicable strategies for overcoming trastuzumab

- resistance caused by PTEN deficiency. *Clin Cancer Res* **13**(19): 5883-5888.
- Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A. et al. 2005. MicroRNA expression profiles classify human cancers. *Nature* **435**(7043): 834-838.
- Lu, P., Ewald, A.J., Martin, G.R., and Werb, Z. 2008. Genetic mosaic analysis reveals FGF receptor 2 function in terminal end buds during mammary gland branching morphogenesis. *Dev Biol* **321**(1): 77-87.
- Lu, P., Weaver, V.M., and Werb, Z. 2012. The extracellular matrix: a dynamic niche in cancer progression. *J Cell Biol* **196**(4): 395-406.
- Lu, P. and Werb, Z. 2008. Patterning mechanisms of branched organs. *Science* **322**(5907): 1506-1509.
- Lujambio, A. and Lowe, S.W. 2012. The microcosmos of cancer. *Nature* **482**(7385): 347-355.
- Lukanidin, E. and Sleeman, J.P. 2012. Building the niche: The role of the S100 proteins in metastatic growth. *Semin Cancer Biol* **22**(3): 216-225.
- Lum, D.H., Tan, J., Rosen, S.D., and Werb, Z. 2007. Gene trap disruption of the mouse heparan sulfate 6-O-endosulfatase gene, *Sulf2*. *Mol Cell Biol* **27**(2): 678-688.
- Ma, L., Reinhardt, F., Pan, E., Soutschek, J., Bhat, B., Marcusson, E.G., Teruya-Feldstein, J., Bell, G.W., and Weinberg, R.A. 2010a. Therapeutic silencing of miR-10b inhibits metastasis in a mouse mammary tumor model. *Nat Biotechnol* **28**(4): 341-347.
- 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.
- Ma, L. and Weinberg, R.A. 2008. Micromanagers of malignancy: role of microRNAs in regulating metastasis. *Trends Genet* **24**(9): 448-456.
- Ma, L., Young, J., Prabhala, H., Pan, E., Mestdagh, P., Muth, D., Teruya-Feldstein, J., Reinhardt, F., Onder, T.T., Valastyan, S. et al. 2010b. miR-9, a MYC/MYCN-activated microRNA, regulates E-cadherin and cancer metastasis. *Nat Cell Biol* **12**(3): 247-256.
- MacDonald, B.T., Tamai, K., and He, X. 2009. Wnt/beta-catenin signaling: components, mechanisms, and diseases. *Dev Cell* **17**(1): 9-26.
- Macias, H., Moran, A., Samara, Y., Moreno, M., Compton, J.E., Harburg, G., Strickland, P., and Hinck, L. 2011. SLIT/ROBO1 signaling suppresses mammary branching morphogenesis by limiting basal cell number. *Dev Cell* **20**(6): 827-840.
- Maeda, A., Moriguchi, T., Hamada, M., Kusakabe, M., Fujioka, Y., Nakano, T., Yoh, K., Lim, K.C., Engel, J.D., and Takahashi, S. 2009. Transcription factor GATA-3 is essential for lens development. *Dev Dyn* **238**(9): 2280-2291.
- Malanchi, I., Santamaria-Martinez, A., Susanto, E., Peng, H., Lehr, H.A., Delaloye, J.F., and Huelsken, J. 2011. Interactions between cancer stem cells and their niche govern metastatic colonization. *Nature* **481**(7379): 85-89.

- Maneechotesuwan, K., Xin, Y., Ito, K., Jazrawi, E., Lee, K.Y., Usmani, O.S., Barnes, P.J., and Adcock, I.M. 2007. Regulation of Th2 cytokine genes by p38 MAPK-mediated phosphorylation of GATA-3. *J Immunol* **178**(4): 2491-2498.
- Maneechotesuwan, K., Yao, X., Ito, K., Jazrawi, E., Usmani, O.S., Adcock, I.M., and Barnes, P.J. 2009. Suppression of GATA-3 nuclear import and phosphorylation: a novel mechanism of corticosteroid action in allergic disease. *PLoS Med* **6**(5): e1000076.
- Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M. et al. 2008. The epithelial-mesenchymal transition generates cells with properties of stem cells. *Cell* **133**(4): 704-715.
- Martello, G., Rosato, A., Ferrari, F., Manfrin, A., Cordenonsi, M., Dupont, S., Enzo, E., Guzzardo, V., Rondina, M., Spruce, T. et al. 2010. A MicroRNA targeting dicer for metastasis control. *Cell* **141**(7): 1195-1207.
- Maschler, S., Wirl, G., Spring, H., Bredow, D.V., Sordat, I., Beug, H., and Reichmann, E. 2005. Tumor cell invasiveness correlates with changes in integrin expression and localization. *Oncogene* **24**(12): 2032-2041.
- Mauti, L.A., Le Bitoux, M.A., Baumer, K., Stehle, J.C., Golshayan, D., Provero, P., and Stamenkovic, I. 2011. Myeloid-derived suppressor cells are implicated in regulating permissiveness for tumor metastasis during mouse gestation. *J Clin Invest* **121**(7): 2794-2807.
- McAllister, S.S., Gifford, A.M., Greiner, A.L., Kelleher, S.P., Saelzler, M.P., Ince, T.A., Reinhardt, F., Harris, L.N., Hylander, B.L., Repasky, E.A. et al. 2008. Systemic endocrine instigation of indolent tumor growth requires osteopontin. *Cell* **133**(6): 994-1005.
- McCaffrey, L.M. and Macara, I.G. 2009. The Par3/aPKC interaction is essential for end bud remodeling and progenitor differentiation during mammary gland morphogenesis. *Genes Dev* **23**(12): 1450-1460.
- McNally, S. and Martin, F. 2011. Molecular regulators of pubertal mammary gland development. *Ann Med* **43**(3): 212-234.
- Medici, D., Hay, E.D., and Goodenough, D.A. 2006. Cooperation between snail and LEF-1 transcription factors is essential for TGF-beta1-induced epithelial-mesenchymal transition. *Mol Biol Cell* **17**(4): 1871-1879.
- Mehra, R., Varambally, S., Ding, L., Shen, R., Sabel, M.S., Ghosh, D., Chinnaiyan, A.M., and Kleer, C.G. 2005. Identification of GATA3 as a breast cancer prognostic marker by global gene expression meta-analysis. *Cancer Res* **65**(24): 11259-11264.
- Meili, R., Cron, P., Hemmings, B.A., and Ballmer-Hofer, K. 1998. Protein kinase B/Akt is activated by polyomavirus middle-T antigen via a phosphatidylinositol 3-kinase-dependent mechanism. *Oncogene* **16**(7): 903-907.
- Melchor, L., Garcia, M.J., Honrado, E., Pole, J.C., Alvarez, S., Edwards, P.A., Caldas, C., Brenton, J.D., and Benitez, J. 2007. Genomic analysis of the 8p11-12 amplicon in familial breast cancer. *Int J Cancer* **120**(3): 714-717.

- Melo, S.A., Moutinho, C., Ropero, S., Calin, G.A., Rossi, S., Spizzo, R., Fernandez, A.F., Davalos, V., Villanueva, A., Montoya, G. et al. 2010. A genetic defect in exportin-5 traps precursor microRNAs in the nucleus of cancer cells. *Cancer Cell* **18**(4): 303-315.
- Melo, S.A., Ropero, S., Moutinho, C., Aaltonen, L.A., Yamamoto, H., Calin, G.A., Rossi, S., Fernandez, A.F., Carneiro, F., Oliveira, C. et al. 2009. A TARBP2 mutation in human cancer impairs microRNA processing and DICER1 function. *Nat Genet* **41**(3): 365-370.
- Merritt, W.M., Lin, Y.G., Han, L.Y., Kamat, A.A., Spannuth, W.A., Schmandt, R., Urbauer, D., Pennacchio, L.A., Cheng, J.F., Nick, A.M. et al. 2008. Dicer, Drosha, and outcomes in patients with ovarian cancer. *N Engl J Med* **359**(25): 2641-2650.
- Miao, Y., Liu, N., Zhang, Y., Liu, Y., Yu, J.H., Dai, S.D., Xu, H.T., and Wang, E.H. 2009. p120ctn isoform 1 expression significantly correlates with abnormal expression of E-cadherin and poor survival of lung cancer patients. *Med Oncol*.
- Miller, L.D., Smeds, J., George, J., Vega, V.B., Vergara, L., Ploner, A., Pawitan, Y., Hall, P., Klaar, S., Liu, E.T. et al. 2005. An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival. *Proc Natl Acad Sci U S A* **102**(38): 13550-13555.
- Minn, A.J., Gupta, G.P., Siegel, P.M., Bos, P.D., Shu, W., Giri, D.D., Viale, A., Olshen, A.B., Gerald, W.L., and Massague, J. 2005. Genes that mediate breast cancer metastasis to lung. *Nature* **436**(7050): 518-524.
- Mishra, P.J. and Merlino, G. 2009. MicroRNA reexpression as differentiation therapy in cancer. *J Clin Invest* **119**(8): 2119-2123.
- Mo, Y.Y. and Reynolds, A.B. 1996. Identification of murine p120 isoforms and heterogeneous expression of p120cas isoforms in human tumor cell lines. *Cancer Res* **56**(11): 2633-2640.
- Moriguchi, T., Takako, N., Hamada, M., Maeda, A., Fujioka, Y., Kuroha, T., Huber, R.E., Hasegawa, S.L., Rao, A., Yamamoto, M. et al. 2006. Gata3 participates in a complex transcriptional feedback network to regulate sympathoadrenal differentiation. *Development* **133**(19): 3871-3881.
- Mott, J.L., Kurita, S., Cazanave, S.C., Bronk, S.F., Werneburg, N.W., and Fernandez-Zapico, M.E. 2010. Transcriptional suppression of mir-29b-1/mir-29a promoter by c-Myc, hedgehog, and NF-kappaB. *J Cell Biochem* **110**(5): 1155-1164.
- Mullen, A.C., Orlando, D.A., Newman, J.J., Loven, J., Kumar, R.M., Bilodeau, S., Reddy, J., Guenther, M.G., DeKoter, R.P., and Young, R.A. 2011. Master transcription factors determine cell-type-specific responses to TGF-beta signaling. *Cell* **147**(3): 565-576.
- Mundy, G.R. 2002. Metastasis to bone: causes, consequences and therapeutic opportunities. *Nat Rev Cancer* **2**(8): 584-593.
- Murdoch, C., Muthana, M., Coffelt, S.B., and Lewis, C.E. 2008. The role of myeloid cells in the promotion of tumour angiogenesis. *Nat Rev Cancer* **8**(8): 618-631.

- Nakamura, M., Choe, S.K., Runko, A.P., Gardner, P.D., and Sagerstrom, C.G. 2008. Nlz1/Znf703 acts as a repressor of transcription. *BMC Dev Biol* **8**: 108.
- Nakamura, M., Runko, A.P., and Sagerstrom, C.G. 2004. A novel subfamily of zinc finger genes involved in embryonic development. *J Cell Biochem* **93**(5): 887-895.
- Nakao, K., Mehta, K.R., Fridlyand, J., Moore, D.H., Jain, A.N., Lafuente, A., Wiencke, J.W., Terdiman, J.P., and Waldman, F.M. 2004. High-resolution analysis of DNA copy number alterations in colorectal cancer by array-based comparative genomic hybridization. *Carcinogenesis* **25**(8): 1345-1357.
- Nakasone, E.S., Askautrud, H.A., Kees, T., Park, J.H., Plaks, V., Ewald, A.J., Fein, M., Rasch, M.G., Tan, Y.X., Qiu, J. et al. 2011. Imaging tumor-stroma interactions during chemotherapy reveals contributions of the microenvironment to resistance. *Cancer Cell* **21**(4): 488-503.
- Nawshad, A., Medici, D., Liu, C.C., and Hay, E.D. 2007. TGFbeta3 inhibits E-cadherin gene expression in palate medial-edge epithelial cells through a Smad2-Smad4-LEF1 transcription complex. *J Cell Sci* **120**(Pt 9): 1646-1653.
- Nelson, W.J. 2008. Regulation of cell-cell adhesion by the cadherin-catenin complex. *Biochem Soc Trans* **36**(Pt 2): 149-155.
- Nesbit, M.A., Bowl, M.R., Harding, B., Ali, A., Ayala, A., Crowe, C., Dobbie, A., Hampson, G., Holdaway, I., Levine, M.A. et al. 2004. Characterization of GATA3 mutations in the hypoparathyroidism, deafness, and renal dysplasia (HDR) syndrome. *J Biol Chem* **279**(21): 22624-22634.
- Neufeld, G. and Kessler, O. 2008. The semaphorins: versatile regulators of tumour progression and tumour angiogenesis. *Nat Rev Cancer* **8**(8): 632-645.
- Neve, R.M., Chin, K., Fridlyand, J., Yeh, J., Baehner, F.L., Fevr, T., Clark, L., Bayani, N., Coppe, J.P., Tong, F. et al. 2006. A collection of breast cancer cell lines for the study of functionally distinct cancer subtypes. *Cancer Cell* **10**(6): 515-527.
- Nguyen, D.X., Bos, P.D., and Massague, J. 2009. Metastasis: from dissemination to organ-specific colonization. *Nat Rev Cancer* **9**(4): 274-284.
- Nguyen, T., Kuo, C., Nicholl, M.B., Sim, M.S., Turner, R.R., Morton, D.L., and Hoon, D.S. 2010. Downregulation of microRNA-29c is associated with hypermethylation of tumor-related genes and disease outcome in cutaneous melanoma. *Epigenetics* **6**(3): 388-394.
- Nonet, G.H., Stampfer, M.R., Chin, K., Gray, J.W., Collins, C.C., and Yaswen, P. 2001. The ZNF217 gene amplified in breast cancers promotes immortalization of human mammary epithelial cells. *Cancer Res* **61**(4): 1250-1254.
- Ohkubo, T. and Ozawa, M. 2004. The transcription factor Snail downregulates the tight junction components independently of E-cadherin downregulation. *J Cell Sci* **117**(Pt 9): 1675-1685.

- Olson, L.E., Tollkuhn, J., Scafoglio, C., Kronen, A., Zhang, J., Ohgi, K.A., Wu, W., Taketo, M.M., Kemler, R., Grosschedl, R. et al. 2006. Homeodomain-mediated beta-catenin-dependent switching events dictate cell-lineage determination. *Cell* **125**(3): 593-605.
- 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.
- Oshimori, N. and Fuchs, E. 2012. Paracrine TGF-beta signaling counterbalances BMP-mediated repression in hair follicle stem cell activation. *Cell Stem Cell* **10**(1): 63-75.
- Oskarsson, T., Acharyya, S., Zhang, X.H., Vanharanta, S., Tavazoie, S.F., Morris, P.G., Downey, R.J., Manova-Todorova, K., Brogi, E., and Massague, J. 2011. Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nat Med* **17**(7): 867-874.
- Ouyang, W., Ranganath, S.H., Weindel, K., Bhattacharya, D., Murphy, T.L., Sha, W.C., and Murphy, K.M. 1998. Inhibition of Th1 development mediated by GATA-3 through an IL-4-independent mechanism. *Immunity* **9**(5): 745-755.
- Padua, D., Zhang, X.H., Wang, Q., Nadal, C., Gerald, W.L., Gomis, R.R., and Massague, J. 2008. TGFbeta primes breast tumors for lung metastasis seeding through angiopoietin-like 4. *Cell* **133**(1): 66-77.
- Paget, S. 1889. The distribution of secondary growths in cancer of the breast. *Lancet* **1**: 571-573.
- Pai, S.Y., Truitt, M.L., Ting, C.N., Leiden, J.M., Glimcher, L.H., and Ho, I.C. 2003. Critical roles for transcription factor GATA-3 in thymocyte development. *Immunity* **19**(6): 863-875.
- Pandolfi, P.P., Roth, M.E., Karis, A., Leonard, M.W., Dzierzak, E., Grosveld, F.G., Engel, J.D., and Lindenbaum, M.H. 1995. Targeted disruption of the GATA3 gene causes severe abnormalities in the nervous system and in fetal liver haematopoiesis. *Nat Genet* **11**(1): 40-44.
- Pase, L., Layton, J.E., Kloosterman, W.P., Carradice, D., Waterhouse, P.M., and Lieschke, G.J. 2009. miR-451 regulates zebrafish erythroid maturation in vivo via its target gata2. *Blood* **113**(8): 1794-1804.
- Passon, N., Gerometta, A., Puppini, C., Lavarone, E., Puglisi, F., Tell, G., Di Loreto, C., and Damante, G. 2012. Expression of Dicer and Drosha in triple-negative breast cancer. *J Clin Pathol* **65**(4): 320-326.
- Pei, X.H., Bai, F., Smith, M.D., Usary, J., Fan, C., Pai, S.Y., Ho, I.C., Perou, C.M., and Xiong, Y. 2009. CDK inhibitor p18(INK4c) is a downstream target of GATA3 and restrains mammary luminal progenitor cell proliferation and tumorigenesis. *Cancer Cell* **15**(5): 389-401.
- Peinado, H., Olmeda, D., and Cano, A. 2007. Snail, Zeb and bHLH factors in tumour progression: an alliance against the epithelial phenotype? *Nat Rev Cancer* **7**(6): 415-428.

- Pellegrino, T., Manna, L., Kudera, S., Liedl, T., Koktysh, D., Rogach, A.L., Keller, S., Rädler, J., Natile, G., and Parak, W.J. 2004. Hydrophobic nanocrystals coated with an amphiphilic polymer shell: A general route to water soluble nanocrystals. *Nano Lett* **4**(4): 703-707.
- Peric, D., Chvalova, K., and Rousselet, G. 2012. Identification of microprocessor-dependent cancer cells allows screening for growth-sustaining micro-RNAs. *Oncogene* **31**(16): 2039-2048.
- Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A. et al. 2000. Molecular portraits of human breast tumours. *Nature* **406**(6797): 747-752.
- Piao, H.L. and Ma, L. 2012. Non-coding RNAs as regulators of mammary development and breast cancer. *J Mammary Gland Biol Neoplasia* **17**(1): 33-42.
- Pietras, K., Pahler, J., Bergers, G., and Hanahan, D. 2008. Functions of paracrine PDGF signaling in the proangiogenic tumor stroma revealed by pharmacological targeting. *PLoS Med* **5**(1): e19.
- Pole, J.C., Courtoy-Cahen, C., Garcia, M.J., Blood, K.A., Cooke, S.L., Alsop, A.E., Tse, D.M., Caldas, C., and Edwards, P.A. 2006. High-resolution analysis of chromosome rearrangements on 8p in breast, colon and pancreatic cancer reveals a complex pattern of loss, gain and translocation. *Oncogene* **25**: 5693-5706.
- Pollard, J.W. 2009. Trophic macrophages in development and disease. *Nat Rev Immunol* **9**(4): 259-270.
- Polyak, K. and Weinberg, R.A. 2009. Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. *Nat Rev Cancer* **9**(4): 265-273.
- Prentice, L.M., Shadeo, A., Lestou, V.S., Miller, M.A., Deleeuw, R.J., Makretsov, N., Turbin, D., Brown, L.A., Macpherson, N., Yorida, E. et al. 2005. NRG1 gene rearrangements in clinical breast cancer: identification of an adjacent novel amplicon associated with poor prognosis. *Oncogene* **24**: 7281-7289.
- Psaila, B. and Lyden, D. 2009. The metastatic niche: adapting the foreign soil. *Nat Rev Cancer* **9**(4): 285-293.
- Qian, X., Karpova, T., Sheppard, A.M., McNally, J., and Lowy, D.R. 2004. E-cadherin-mediated adhesion inhibits ligand-dependent activation of diverse receptor tyrosine kinases. *Embo J* **23**(8): 1739-1748.
- Quinlan, K.G., Nardini, M., Verger, A., Francescato, P., Yaswen, P., Corda, D., Bolognesi, M., and Crossley, M. 2006. Specific recognition of ZNF217 and other zinc finger proteins at a surface groove of C-terminal binding proteins. *Mol Cell Biol* **26**(21): 8159-8172.
- Ranganath, S. and Murphy, K.M. 2001. Structure and specificity of GATA proteins in Th2 development. *Mol Cell Biol* **21**(8): 2716-2725.
- Rangaswami, H., Bulbule, A., and Kundu, G.C. 2006. Osteopontin: role in cell signaling and cancer progression. *Trends Cell Biol* **16**(2): 79-87.
- Ravi, A., Gurtan, A.M., Kumar, M.S., Bhutkar, A., Chin, C., Lu, V., Lees, J.A., Jacks, T., and Sharp, P.A. 2012. Proliferation and Tumorigenesis of a

- Murine Sarcoma Cell Line in the Absence of DICER1. *Cancer Cell* **21**(6): 848-855.
- Reagan-Shaw, S., Nihal, M., and Ahmad, N. 2008. Dose translation from animal to human studies revisited. *Faseb J* **22**(3): 659-661.
- Reynolds, A.B. and Roczniak-Ferguson, A. 2004. Emerging roles for p120-catenin in cell adhesion and cancer. *Oncogene* **23**(48): 7947-7956.
- Roderburg, C., Urban, G.W., Bettermann, K., Vucur, M., Zimmermann, H., Schmidt, S., Janssen, J., Koppe, C., Knolle, P., Castoldi, M. et al. 2010. Micro-RNA profiling reveals a role for miR-29 in human and murine liver fibrosis. *Hepatology* **53**(1): 209-218.
- Rosenthal, S.J., Chang, J.C., Kovtun, O., McBride, J.R., and Tomlinson, I.D. 2011. Biocompatible quantum dots for biological applications. *Chem Biol* **18**(1): 10-24.
- Runko, A.P. and Sagerstrom, C.G. 2003. Nlz belongs to a family of zinc-finger-containing repressors and controls segmental gene expression in the zebrafish hindbrain. *Dev Biol* **262**(2): 254-267.
- . 2004. Isolation of nlz2 and characterization of essential domains in Nlz family proteins. *J Biol Chem* **279**(12): 11917-11925.
- Salama, I., Malone, P.S., Mihaimeed, F., and Jones, J.L. 2008. A review of the S100 proteins in cancer. *Eur J Surg Oncol* **34**(4): 357-364.
- Sanchez-Elsner, T., Botella, L.M., Velasco, B., Corbi, A., Attisano, L., and Bernabeu, C. 2001. Synergistic cooperation between hypoxia and transforming growth factor-beta pathways on human vascular endothelial growth factor gene expression. *J Biol Chem* **276**(42): 38527-38535.
- Santini, D., Perrone, G., Roato, I., Godio, L., Pantano, F., Grasso, D., Russo, A., Vincenzi, B., Fratto, M.E., Sabbatini, R. et al. 2011a. Expression pattern of receptor activator of NFkappaB (RANK) in a series of primary solid tumors and related bone metastases. *J Cell Physiol* **226**(3): 780-784.
- Santini, D., Schiavon, G., Vincenzi, B., Gaeta, L., Pantano, F., Russo, A., Ortega, C., Porta, C., Galluzzo, S., Armento, G. et al. 2011b. Receptor activator of NF-kB (RANK) expression in primary tumors associates with bone metastasis occurrence in breast cancer patients. *PLoS One* **6**(4): e19234.
- Sarrio, D., Perez-Mies, B., Hardisson, D., Moreno-Bueno, G., Suarez, A., Cano, A., Martin-Perez, J., Gamallo, C., and Palacios, J. 2004. Cytoplasmic localization of p120ctn and E-cadherin loss characterize lobular breast carcinoma from preinvasive to metastatic lesions. *Oncogene* **23**(19): 3272-3283.
- Schaffhausen, B.S. and Roberts, T.M. 2009. Lessons from polyoma middle T antigen on signaling and transformation: A DNA tumor virus contribution to the war on cancer. *Virology* **384**(2): 304-316.
- Schedin, P. 2006. Pregnancy-associated breast cancer and metastasis. *Nat Rev Cancer* **6**(4): 281-291.
- Scheel, C., Eaton, E.N., Li, S.H., Chaffer, C.L., Reinhardt, F., Kah, K.J., Bell, G., Guo, W., Rubin, J., Richardson, A.L. et al. 2011. Paracrine and autocrine signals induce and maintain mesenchymal and stem cell states in the breast. *Cell* **145**(6): 926-940.

- Schilcher, R.B., Haas, C.D., Samson, M.K., Young, J.D., and Baker, L.H. 1986. Phase I evaluation and clinical pharmacology of tricyclic nucleoside 5'-phosphate using a weekly intravenous regimen. *Cancer Res* **46**(6): 3147-3151.
- Schmidt, M., Bohm, D., von Torne, C., Steiner, E., Puhl, A., Pilch, H., Lehr, H.A., Hengstler, J.G., Kolbl, H., and Gehrman, M. 2008. The humoral immune system has a key prognostic impact in node-negative breast cancer. *Cancer Res* **68**(13): 5405-5413.
- Schrier, J., Lee, B., and Wang, L.W. 2008. Mechanical and electronic-structure properties of compressed CdSe tetrapod nanocrystals. *J Nanosci Nanotechnol* **8**: 1994-1998.
- Sethi, A., Mao, W., Wordinger, R.J., and Clark, A.F. 2011. Transforming growth factor-beta induces extracellular matrix protein cross-linking lysyl oxidase (LOX) genes in human trabecular meshwork cells. *Invest Ophthalmol Vis Sci* **52**(8): 5240-5250.
- 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.
- She, Q.B., Chandrapaty, S., Ye, Q., Lobo, J., Haskell, K.M., Leander, K.R., DeFeo-Jones, D., Huber, H.E., and Rosen, N. 2008. Breast tumor cells with PI3K mutation or HER2 amplification are selectively addicted to Akt signaling. *PLoS One* **3**(8): e3065.
- She, Q.B., Halilovic, E., Ye, Q., Zhen, W., Shirasawa, S., Sasazuki, T., Solit, D.B., and Rosen, N. 2010. 4E-BP1 is a key effector of the oncogenic activation of the AKT and ERK signaling pathways that integrates their function in tumors. *Cancer Cell* **18**(1): 39-51.
- Sherr, C.J. and DePinho, R.A. 2000. Cellular senescence: mitotic clock or culture shock? *Cell* **102**(4): 407-410.
- Shi, Y., Sawada, J., Sui, G., Affar el, B., Whetstine, J.R., Lan, F., Ogawa, H., Luke, M.P., and Nakatani, Y. 2003. Coordinated histone modifications mediated by a CtBP co-repressor complex. *Nature* **422**(6933): 735-738.
- Sircoulomb, F., Nicolas, N., Ferrari, A., Finetti, P., Bekhouche, I., Rousselet, E., Lonigro, A., Adelaide, J., Baudalet, E., Esteyries, S. et al. 2011. ZNF703 gene amplification at 8p12 specifies luminal B breast cancer. *EMBO Mol Med* **3**(3): 153-166.
- Slorach, E.M., Chou, J., and Werb, Z. 2011. Zeppo1 is a novel metastasis promoter that represses E-cadherin expression and regulates p120-catenin isoform expression and localization. *Genes Dev* **25**(5): 471-484.
- Smalley, M. and Ashworth, A. 2003. Stem cells and breast cancer: A field in transit. *Nat Rev Cancer* **3**(11): 832-844.
- Smith, S.C. and Theodorescu, D. 2009. Learning therapeutic lessons from metastasis suppressor proteins. *Nat Rev Cancer* **9**(4): 253-264.
- Sneddon, J.B. and Werb, Z. 2007. Location, location, location: the cancer stem cell niche. *Cell Stem Cell* **1**(6): 607-611.

- Sorlie, T., Perou, C.M., Fan, C., Geisler, S., Aas, T., Nobel, A., Anker, G., Akslen, L.A., Botstein, D., Borresen-Dale, A.L. et al. 2006. Gene expression profiles do not consistently predict the clinical treatment response in locally advanced breast cancer. *Mol Cancer Ther* **5**(11): 2914-2918.
- 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. et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proc Natl Acad Sci U S A* **98**(19): 10869-10874.
- Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S. et al. 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**(14): 8418-8423.
- Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B. et al. 2006. Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. *Journal of the National Cancer Institute* **98**(4): 262-272.
- Stefani, G. and Slack, F.J. 2008. Small non-coding RNAs in animal development. *Nat Rev Mol Cell Biol* **9**(3): 219-230.
- Steiner, D.F., Thomas, M.F., Hu, J.K., Yang, Z., Babiarz, J.E., Allen, C.D., Matloubian, M., Belloch, R., and Ansel, K.M. 2011. MicroRNA-29 regulates T-box transcription factors and interferon-gamma production in helper T cells. *Immunity* **35**(2): 169-181.
- Sternlicht, M.D., Kouros-Mehr, H., Lu, P., and Werb, Z. 2006. Hormonal and local control of mammary branching morphogenesis. *Differentiation* **74**(7): 365-381.
- 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.
- Stone, J.W., Sisco, P.N., Goldsmith, E.C., Baxter, S.C., and Murphy, C.J. 2007. Using gold nanorods to probe cell-induced collagen deformation. *Nano Lett* **7**: 116-119.
- Streichfuss, M., Erbs, F., Uhrig, K., Kurre, R., Clemen, A.E., Böhm, C.H., Haraszti, T., and Spatz, J.P. 2011. Measuring forces between two single actin filaments during bundle formation. *Nano Lett* **11**(9): 3676-3680.
- Symmans, W.F., Hatzis, C., Sotiriou, C., Andre, F., Peintinger, F., Regitnig, P., Daxenbichler, G., Desmedt, C., Domont, J., Marth, C. et al. 2010. Genomic index of sensitivity to endocrine therapy for breast cancer. *J Clin Oncol* **28**(27): 4111-4119.
- Taghon, T., Yui, M.A., and Rothenberg, E.V. 2007. Mast cell lineage diversion of T lineage precursors by the essential T cell transcription factor GATA-3. *Nat Immunol* **8**(8): 845-855.
- Takahashi, S., Shimizu, R., Suwabe, N., Kuroha, T., Yoh, K., Ohta, J., Nishimura, S., Lim, K.C., Engel, J.D., and Yamamoto, M. 2000. GATA factor transgenes under GATA-1 locus control rescue germline GATA-1 mutant deficiencies. *Blood* **96**(3): 910-916.

- Takemoto, N., Arai, K., and Miyatake, S. 2002. Cutting edge: the differential involvement of the N-finger of GATA-3 in chromatin remodeling and transactivation during Th2 development. *J Immunol* **169**(8): 4103-4107.
- Taketo, M., Schroeder, A.C., Mobraaten, L.E., Gunning, K.B., Hanten, G., Fox, R.R., Roderick, T.H., Stewart, C.L., Lilly, F., Hansen, C.T. et al. 1991. FVB/N: an inbred mouse strain preferable for transgenic analyses. *Proc Natl Acad Sci U S A* **88**(6): 2065-2069.
- Takeuchi, J.K. and Bruneau, B.G. 2009. Directed transdifferentiation of mouse mesoderm to heart tissue by defined factors. *Nature* **459**(7247): 708-711.
- Talvinen, K., Tuikkala, J., Nykanen, M., Nieminen, A., Anttinen, J., Nevalainen, O.S., Hurme, S., Kuopio, T., and Kronqvist, P. 2010. Altered expression of p120catenin predicts poor outcome in invasive breast cancer. *J Cancer Res Clin Oncol* **136**: 1377-1387.
- Tan, J.L., Tien, J., Pirone, D., Gray, D.S., and Chen, C.S. 2003. Cells lying on a bed of microneedles: An approach to isolate mechanical force. *Proc Natl Acad Sci USA* **100**: 1484-1489.
- Tan, W., Zhang, W., Strasner, A., Grivennikov, S., Cheng, J.Q., Hoffman, R.M., and Karin, M. 2011. Tumour-infiltrating regulatory T cells stimulate mammary cancer metastasis through RANKL-RANK signalling. *Nature* **470**(7335): 548-553.
- Tarin, D., Thompson, E.W., and Newgreen, D.F. 2005. The fallacy of epithelial mesenchymal transition in neoplasia. *Cancer Res* **65**(14): 5996-6000; discussion 6000-5991.
- Taulli, R., Bersani, F., Foglizzo, V., Linari, A., Vigna, E., Ladanyi, M., Tuschl, T., and Ponzetto, C. 2009. The muscle-specific microRNA miR-206 blocks human rhabdomyosarcoma growth in xenotransplanted mice by promoting myogenic differentiation. *J Clin Invest* **119**(8): 2366-2378.
- Tavazoie, S.F., Alarcon, C., Oskarsson, T., Padua, D., Wang, Q., Bos, P.D., Gerald, W.L., and Massague, J. 2008. Endogenous human microRNAs that suppress breast cancer metastasis. *Nature* **451**(7175): 147-152.
- Theisen, H., Syed, A., Nguyen, B.T., Lukacsovich, T., Purcell, J., Srivastava, G.P., Iron, D., Gaudenz, K., Nie, Q., Wan, F.Y. et al. 2007. Wingless directly represses DPP morphogen expression via an armadillo/TCF/Brinker complex. *PLoS One* **2**(1): e142.
- Thiery, J.P., Acloque, H., Huang, R.Y., and Nieto, M.A. 2009. Epithelial-mesenchymal transitions in development and disease. *Cell* **139**(5): 871-890.
- Thillainadesan, G., Isovich, M., Loney, E., Andrews, J., Tini, M., and Torchia, J. 2008. Genome analysis identifies the p15ink4b tumor suppressor as a direct target of the ZNF217/CoREST complex. *Mol Cell Biol* **28**(19): 6066-6077.
- Timpson, P., Jones, G.E., Frame, M.C., and Brunton, V.G. 2001. Coordination of cell polarization and migration by the Rho family GTPases requires Src tyrosine kinase activity. *Curr Biol* **11**(23): 1836-1846.

- Tong, Q., Dalgin, G., Xu, H., Ting, C.N., Leiden, J.M., and Hotamisligil, G.S. 2000. Function of GATA transcription factors in preadipocyte-adipocyte transition. *Science* **290**(5489): 134-138.
- Trompouki, E., Bowman, T.V., Lawton, L.N., Fan, Z.P., Wu, D.C., DiBiase, A., Martin, C.S., Cech, J.N., Sessa, A.K., Leblanc, J.L. et al. 2011. Lineage regulators direct BMP and Wnt pathways to cell-specific programs during differentiation and regeneration. *Cell* **147**(3): 577-589.
- Tsai, F.Y., Browne, C.P., and Orkin, S.H. 1998. Knock-in mutation of transcription factor GATA-3 into the GATA-1 locus: partial rescue of GATA-1 loss of function in erythroid cells. *Dev Biol* **196**(2): 218-227.
- Ucar, A., Vafaizadeh, V., Jarry, H., Fiedler, J., Klemmt, P.A., Thum, T., Groner, B., and Chowdhury, K. 2010. miR-212 and miR-132 are required for epithelial stromal interactions necessary for mouse mammary gland development. *Nat Genet* **42**(12): 1101-1108.
- Usary, J., Llaca, V., Karaca, G., Presswala, S., Karaca, M., He, X., Langerod, A., Karsen, R., Oh, D.S., Dressler, L.G. et al. 2004. Mutation of GATA3 in human breast tumors. *Oncogene* **23**(46): 7669-7678.
- 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.
- Valastyan, S. and Weinberg, R.A. 2011. Tumor metastasis: molecular insights and evolving paradigms. *Cell* **147**(2): 275-292.
- Van Esch, H., Groenen, P., Nesbit, M.A., Schuffenhauer, S., Lichtner, P., Vanderlinden, G., Harding, B., Beetz, R., Bilous, R.W., Holdaway, I. et al. 2000. GATA3 haplo-insufficiency causes human HDR syndrome. *Nature* **406**(6794): 419-422.
- van Hamburg, J.P., de Bruijn, M.J., Dingjan, G.M., Beverloo, H.B., Diepstraten, H., Ling, K.W., and Hendriks, R.W. 2008. Cooperation of Gata3, c-Myc and Notch in malignant transformation of double positive thymocytes. *Mol Immunol* **45**(11): 3085-3095.
- van Hengel, J. and van Roy, F. 2007. Diverse functions of p120ctn in tumors. *Biochim Biophys Acta* **1773**(1): 78-88.
- van Rooij, E., Sutherland, L.B., Qi, X., Richardson, J.A., Hill, J., and Olson, E.N. 2007. Control of stress-dependent cardiac growth and gene expression by a microRNA. *Science* **316**(5824): 575-579.
- van Rooij, E., Sutherland, L.B., Thatcher, J.E., DiMaio, J.M., Naseem, R.H., Marshall, W.S., Hill, J.A., and Olson, E.N. 2008. Dysregulation of microRNAs after myocardial infarction reveals a role of miR-29 in cardiac fibrosis. *Proc Natl Acad Sci U S A* **105**(35): 13027-13032.
- Veronesi, U., Boyle, P., Goldhirsch, A., Orecchia, R., and Viale, G. 2005. Breast cancer. *Lancet* **365**(9472): 1727-1741.
- Villadsen, R., Fridriksdottir, A.J., Ronnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M.A., Bissell, M.J., and Petersen, O.W. 2007. Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol* **177**(1): 87-101.

- Visvader, J.E. 2009. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev* **23**(22): 2563-2577.
- Visvader, J.E. and Lindeman, G.J. 2008. Cancer stem cells in solid tumours: accumulating evidence and unresolved questions. *Nat Rev Cancer* **8**(10): 755-768.
- Voduc, D., Cheang, M., and Nielsen, T. 2008. GATA-3 expression in breast cancer has a strong association with estrogen receptor but lacks independent prognostic value. *Cancer Epidemiol Biomarkers Prev* **17**(2): 365-373.
- Vogel, V. and Sheetz, M.P. 2009. Cell fate regulation by coupling mechanical cycles to biochemical signaling pathways. *Curr Opin Cell Biol* **21**: 38-46.
- Wachman, E.S., Niu, W., and Farkas, D.L. 1997. AOTF microscope for imaging with increased speed and spectral versatility. *Biophys J* **73**(3): 1215-1222.
- Wagner, K.U., McAllister, K., Ward, T., Davis, B., Wiseman, R., and Hennighausen, L. 2001. Spatial and temporal expression of the Cre gene under the control of the MMTV-LTR in different lines of transgenic mice. *Transgenic Res* **10**(6): 545-553.
- Wang, B., Komers, R., Carew, R., Winbanks, C.E., Xu, B., Herman-Edelstein, M., Koh, P., Thomas, M., Jandeleit-Dahm, K., Gregorevic, P. et al. 2011. Suppression of microRNA-29 Expression by TGF-beta1 Promotes Collagen Expression and Renal Fibrosis. *J Am Soc Nephrol* **2**: 252-265.
- Wang, H., Garzon, R., Sun, H., Ladner, K.J., Singh, R., Dahlman, J., Cheng, A., Hall, B.M., Qualman, S.J., Chandler, D.S. et al. 2008a. NF-kappaB-YY1-miR-29 regulatory circuitry in skeletal myogenesis and rhabdomyosarcoma. *Cancer Cell* **14**(5): 369-381.
- Wang, X.Y., Yin, Y., Yuan, H., Sakamaki, T., Okano, H., and Glazer, R.I. 2008b. Musashi1 modulates mammary progenitor cell expansion through proliferin-mediated activation of the Wnt and Notch pathways. *Mol Cell Biol* **28**(11): 3589-3599.
- Wang, Y., Baskerville, S., Shenoy, A., Babiarz, J.E., Baehner, L., and Blelloch, R. 2008c. Embryonic stem cell-specific microRNAs regulate the G1-S transition and promote rapid proliferation. *Nat Genet* **40**(12): 1478-1483.
- Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J. et al. 2005. Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. *Lancet* **365**(9460): 671-679.
- 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.
- Wang, Z., Sandiford, S., Wu, C., and Li, S.S. 2009. Numb regulates cell-cell adhesion and polarity in response to tyrosine kinase signalling. *Embo J* **28**(16): 2360-2373.
- Watnick, R.S., Cheng, Y.N., Rangarajan, A., Ince, T.A., and Weinberg, R.A. 2003. Ras modulates Myc activity to repress thrombospondin-1 expression and increase tumor angiogenesis. *Cancer Cell* **3**(3): 219-231.

- Webster, M.A., Hutchinson, J.N., Rauh, M.J., Muthuswamy, S.K., Anton, M., Tortorice, C.G., Cardiff, R.D., Graham, F.L., Hassell, J.A., and Muller, W.J. 1998. Requirement for both Shc and phosphatidylinositol 3' kinase signaling pathways in polyomavirus middle T-mediated mammary tumorigenesis. *Mol Cell Biol* **18**(4): 2344-2359.
- Wei, G., Abraham, B.J., Yagi, R., Jothi, R., Cui, K., Sharma, S., Narlikar, L., Northrup, D.L., Tang, Q., Paul, W.E. et al. 2011. Genome-wide analyses of transcription factor GATA3-mediated gene regulation in distinct T cell types. *Immunity* **35**(2): 299-311.
- Welm, B.E., Dijkgraaf, G.J., Bledau, A.S., Welm, A.L., and Werb, Z. 2008. Lentiviral transduction of mammary stem cells for analysis of gene function during development and cancer. *Cell Stem Cell* **2**(1): 90-102.
- Winbanks, C.E., Wang, B., Beyer, C., Koh, P., White, L., Kantharidis, P., and Gregorevic, P. 2011. TGF-beta regulates miR-206 and miR-29 to control myogenic differentiation through regulation of HDAC4. *J Biol Chem* **286**(16): 13805-13814.
- Wiseman, B.S. and Werb, Z. 2002. Stromal effects on mammary gland development and breast cancer. *Science* **296**(5570): 1046-1049.
- Wong, D.J., Liu, H., Ridky, T.W., Cassarino, D., Segal, E., and Chang, H.Y. 2008. Module map of stem cell genes guides creation of epithelial cancer stem cells. *Cell Stem Cell* **2**(4): 333-344.
- Woodward, W.A., Chen, M.S., Behbod, F., and Rosen, J.M. 2005. On mammary stem cells. *J Cell Sci* **118**(Pt 16): 3585-3594.
- Wrana, J.L., Attisano, L., Carcamo, J., Zentella, A., Doody, J., Laiho, M., Wang, X.F., and Massague, J. 1992. TGF beta signals through a heteromeric protein kinase receptor complex. *Cell* **71**(6): 1003-1014.
- Wright, J.A., Richer, J.K., and Goodall, G.J. 2010. microRNAs and EMT in mammary cells and breast cancer. *J Mammary Gland Biol Neoplasia* **15**(2): 213-223.
- Xiong, Y., Fang, J.H., Yun, J.P., Yang, J., Zhang, Y., Jia, W.H., and Zhuang, S.M. 2009. Effects of microRNA-29 on apoptosis, tumorigenicity, and prognosis of hepatocellular carcinoma. *Hepatology* **51**(3): 836-845.
- Yamana, N., Arakawa, Y., Nishino, T., Kurokawa, K., Tanji, M., Itoh, R.E., Monypenny, J., Ishizaki, T., Bito, H., Nozaki, K. et al. 2006. The Rho-mDia1 pathway regulates cell polarity and focal adhesion turnover in migrating cells through mobilizing Apc and c-Src. *Mol Cell Biol* **26**(18): 6844-6858.
- Yamashita, M., Shinnakasu, R., Asou, H., Kimura, M., Hasegawa, A., Hashimoto, K., Hatano, N., Ogata, M., and Nakayama, T. 2005. Ras-ERK MAPK cascade regulates GATA3 stability and Th2 differentiation through ubiquitin-proteasome pathway. *J Biol Chem* **280**(33): 29409-29419.
- Yamashita, M., Ukai-Tadenuma, M., Miyamoto, T., Sugaya, K., Hosokawa, H., Hasegawa, A., Kimura, M., Taniguchi, M., DeGregori, J., and Nakayama, T. 2004. Essential role of GATA3 for the maintenance of type 2 helper T (Th2) cytokine production and chromatin remodeling at the Th2 cytokine gene loci. *J Biol Chem* **279**(26): 26983-26990.

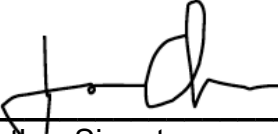
- Yan, H.H., Pickup, M., Pang, Y., Gorska, A.E., Li, Z., Chytil, A., Geng, Y., Gray, J.W., Moses, H.L., and Yang, L. 2010a. Gr-1+CD11b+ myeloid cells tip the balance of immune protection to tumor promotion in the premetastatic lung. *Cancer Res* **70**(15): 6139-6149.
- Yan, M., Huang, H.Y., Wang, T., Wan, Y., Cui, S.D., Liu, Z.Z., and Fan, Q.X. 2012. Dysregulated expression of dicer and drosha in breast cancer. *Pathol Oncol Res* **18**(2): 343-348.
- Yan, W., Cao, Q.J., Arenas, R.B., Bentley, B., and Shao, R. 2010b. GATA3 inhibits breast cancer metastasis through the reversal of epithelial-mesenchymal transition. *J Biol Chem* **285**(18): 14042-14051.
- Yanagisawa, M., Huvelde, D., Kreinest, P., Lohse, C.M., Cheville, J.C., Parker, A.S., Copland, J.A., and Anastasiadis, P.Z. 2008. A p120 catenin isoform switch affects Rho activity, induces tumor cell invasion, and predicts metastatic disease. *J Biol Chem* **283**(26): 18344-18354.
- 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. 2004a. Twist, a master regulator of morphogenesis, plays an essential role in tumor metastasis. *Cell* **117**(7): 927-939.
- Yang, J. and Weinberg, R.A. 2008. Epithelial-mesenchymal transition: at the crossroads of development and tumor metastasis. *Dev Cell* **14**(6): 818-829.
- Yang, L., Dan, H.C., Sun, M., Liu, Q., Sun, X.M., Feldman, R.I., Hamilton, A.D., Polokoff, M., Nicosia, S.V., Herlyn, M. et al. 2004b. Akt/protein kinase B signaling inhibitor-2, a selective small molecule inhibitor of Akt signaling with antitumor activity in cancer cells overexpressing Akt. *Cancer Res* **64**(13): 4394-4399.
- Yang, X.J. and Seto, E. 2008. The Rpd3/Hda1 family of lysine deacetylases: from bacteria and yeast to mice and men. *Nat Rev Mol Cell Biol* **9**(3): 206-218.
- Yang, Z., Gu, L., Romeo, P.H., Bories, D., Motohashi, H., Yamamoto, M., and Engel, J.D. 1994. Human GATA-3 trans-activation, DNA-binding, and nuclear localization activities are organized into distinct structural domains. *Mol Cell Biol* **14**(3): 2201-2212.
- Yang, Z.Q., Streicher, K.L., Ray, M.E., Abrams, J., and Ethier, S.P. 2006. Multiple interacting oncogenes on the 8p11-p12 amplicon in human breast cancer. *Cancer Res* **66**(24): 11632-11643.
- Yap, A.S., Niessen, C.M., and Gumbiner, B.M. 1998. The juxtamembrane region of the cadherin cytoplasmic tail supports lateral clustering, adhesive strengthening, and interaction with p120ctn. *J Cell Biol* **141**(3): 779-789.
- Yi, R., Qin, Y., Macara, I.G., and Cullen, B.R. 2003. Exportin-5 mediates the nuclear export of pre-microRNAs and short hairpin RNAs. *Genes Dev* **17**(24): 3011-3016.
- Yook, J.I., Li, X.Y., Ota, I., Fearon, E.R., and Weiss, S.J. 2005. Wnt-dependent regulation of the E-cadherin repressor snail. *J Biol Chem* **280**(12): 11740-11748.

- Yoon, N.K., Maresh, E.L., Shen, D., Elshimali, Y., Apple, S., Horvath, S., Mah, V., Bose, S., Chia, D., Chang, H.R. et al. 2010. Higher levels of GATA3 predict better survival in women with breast cancer. *Hum Pathol* **41**(12): 1794-1801.
- You, A., Tong, J.K., Grozinger, C.M., and Schreiber, S.L. 2001. CoREST is an integral component of the CoREST- human histone deacetylase complex. *Proc Natl Acad Sci U S A* **98**(4): 1454-1458.
- Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J. et al. 2007. let-7 regulates self renewal and tumorigenicity of breast cancer cells. *Cell* **131**(6): 1109-1123.
- Zawel, L., Dai, J.L., Buckhaults, P., Zhou, S., Kinzler, K.W., Vogelstein, B., and Kern, S.E. 1998. Human Smad3 and Smad4 are sequence-specific transcription activators. *Mol Cell* **1**(4): 611-617.
- Zeng, Y.A. and Nusse, R. 2010. Wnt proteins are self-renewal factors for mammary stem cells and promote their long-term expansion in culture. *Cell Stem Cell* **6**(6): 568-577.
- Zhang, Y., Sieuwerts, A.M., McGreevy, M., Casey, G., Cufer, T., Paradiso, A., Harbeck, N., Span, P.N., Hicks, D.G., Crowe, J. et al. 2009. The 76-gene signature defines high-risk patients that benefit from adjuvant tamoxifen therapy. *Breast Cancer Res Treat* **116**(2): 303-309.
- Zhao, G.Y., Li, Z.Y., Zou, H.L., Hu, Z.L., Song, N.N., Zheng, M.H., Su, C.J., and Ding, Y.Q. 2008. Expression of the transcription factor GATA3 in the postnatal mouse central nervous system. *Neurosci Res* **61**(4): 420-428.
- Zhao, X., Yang, Y., Fitch, D.H., and Herman, M.A. 2002. TLP-1 is an asymmetric cell fate determinant that responds to Wnt signals and controls male tail tip morphogenesis in *C. elegans*. *Development* **129**(6): 1497-1508.
- Zhao, Y., Ransom, J.F., Li, A., Vedantham, V., von Drehle, M., Muth, A.N., Tsuchihashi, T., McManus, M.T., Schwartz, R.J., and Srivastava, D. 2007. Dysregulation of cardiogenesis, cardiac conduction, and cell cycle in mice lacking miRNA-1-2. *Cell* **129**(2): 303-317.
- Zhou, M. and Ouyang, W. 2003. The function role of GATA-3 in Th1 and Th2 differentiation. *Immunol Res* **28**(1): 25-37.
- Zhou, M., Ouyang, W., Gong, Q., Katz, S.G., White, J.M., Orkin, S.H., and Murphy, K.M. 2001. Friend of GATA-1 represses GATA-3-dependent activity in CD4+ T cells. *J Exp Med* **194**(10): 1461-1471.
- Zhu, M., Yi, M., Kim, C.H., Deng, C., Li, Y., Medina, D., Stephens, R.M., and Green, J.E. 2011. Integrated miRNA and mRNA expression profiling of mouse mammary tumor models identifies miRNA signatures associated with mammary tumor lineage. *Genome Biol* **12**(8): R77.

Publishing Agreement

It is the policy of the University to encourage the distribution of all theses, dissertations, and manuscripts. Copies of all UCSF theses, dissertations, and manuscripts will be routed to the library via the Graduate Division. The library will make all theses, dissertations, and manuscripts accessible to the public and will preserve these to the best of their abilities, in perpetuity.

I hereby grant permission to the Graduate Division of the University of California, San Francisco to release copies of my thesis, dissertation, or manuscript to the Campus Library to provide access and preservation, in whole or in part, in perpetuity.



Author Signature

8/15/12
Date