

An *in vivo* metastasis screen of chromatin modifiers reveals a potential role for dopamine  
in adrenal and brain colonization.

A thesis

Submitted by

Benjamin T. Dake

In partial fulfillment of the requirements for the degree of

Doctor of Philosophy

In

Cell, Molecular and Developmental Biology

TUFTS UNIVERSITY

Sackler School of Graduate Biomedical Sciences

September 2015

Adviser: Charlotte Kuperwasser

## ABSTRACT

Metastasis is the ultimate and fatal stage of malignant progression responsible for the vast majority of cancer deaths. Experimentally we know that metastasis is a highly inefficient process with adaptation to the secondary site being the rate limiting step. Experimental systems able to recapitulate colonization of multiple organs are lacking. To better understand this process I established a novel *in vivo* model for the study of systemic metastasis by a breast cancer tumor cell line. Ultrasound guided intracardiac injection coupled with non-invasive imaging allowed me to seed all of the organs in the body and track the establishment and growth of the resulting brain, adrenal, ovarian and other tumors. This cell line was established from a metastasis to the brain and I describe the neurotropic properties it gained during adaptation to this environment including a previously unknown, and importantly clinically addressable, response to dopamine which increases tumor cell viability *in vitro*.

Additionally I leveraged this new model to globally address metastatic organ tropism. While cancer has long been acknowledged as a genetic disease, it is now recognized that the number of epigenetic abnormalities exceeds those of a genetic nature. Clinically, metastases in breast cancer may manifest years or decades after surgery implying a period of dormancy for the tumor seeds sowed by the primary tumor. Aberrant chromatin modification has been established as a driver of malignancy and I hypothesized that chromatin modification may be involved in the transition from a covert dormant cell or micro metastasis to a full blown secondary tumor. I performed a comprehensive screen knocking down all human chromatin modifiers singly to gauge the effect on metastatic organ tropism. Analysis of the hits in the nonadrenal tumors identified four genes EHMT1,

JMJD5, PRDM14 and SIRT3 that were targeted above the noise threshold and by more than one hairpin. Additional experimentation using the system described herein may elucidate metastatic mechanisms of these genes in colonization of the brain and other organs.

## ACKNOWLEDGMENTS

I owe a debt I may never be able to repay to my wife Tamara who enabled me to be a graduate student father for five years. I hope she recognizes my dedication to science was always secondary to that of our family. I am quite certain that like most parents we did not understand all the difficulties and joys that parenthood would bring to us at its onset six years ago. I am equally sure the former were exacerbated by the coincidence of graduate school and for this I apologize. Dankeschön mein Schatz.

My daughter Penny was there from the beginning and I thank her for her unwavering confidence in me and for being the most exquisite big sister to my son James who came along in year three. Jamesy you don't know how lucky you are to have Penny as an example of how to live life and as a protector constantly looking out for you. Thank you both for being who you are and giving my life such meaning. Please always embrace your inquisitive nature and live your lives with passion. Oh, and don't forget to call your mother and me every now and then while you are out being passionate.

I feel very fortunate to have had such a great group of lab mates over the years. To Dr. Jamz and Patty, thank you so much for helping me figure out what the hell was going on in the beginning. Maja, Adam, Sarah and Wen thank you for your comradery over the years. Adam thank you especially for your beer, and your physical might (you too Wen) in helping me move furniture from Cambridge to Brighton to the suburbs. Devin, Jerrica, Tom and Ania, thank you so much for your honest friendship and perspective as I slogged through the screen and found bearing with the genesis of my dopamine project. Tom I especially enjoyed our tissue culture and food truck lamentations. Lisa, you were the rock

for our lab steadying us and guiding us. I am amazed by your skill, productivity and generosity. You always made time for me when I needed help learning a technique, or some advice, or to trade parental anecdotes. To the new kids Andrew, Jessica and Kayla I wish you the best and thank you for putting up with my grumpy self at the end.

To my committee: Grace as my chair for both the qual and my TAC you have been such an outstanding mentor and you touch the lives of so many students at Sackler I don't think I can express how lucky we are. Thank you. Jim your prescient guidance from the first day I saw you as a lecturer through my rotation in your lab and in our committee meetings has enshrined you as my grad student Dad and I can think of no other better. I hope I made you proud. Phil, you were there for me every day I needed help preparing a talk or interpreting a result and your working knowledge and history of cancer biology was an invaluable resource. Not to mention your excessively kind demeanor, great sense of humor, excellent genes, dedication to knowledge and earning the trust of your students. Charlotte, I don't know how to thank you for what you have done for me over the years. The people and the resources you provided made graduate school as enjoyable an endeavor as I think is possible.

To my TBBC/TUNECC compatriots Andy, Hailing, Mike, Jen, Bina and Sohini thank you for opening my eyes to the world of science beyond servitude where people are actually compensated monetarily. Who would think such a thing exists!?! Not to take away from those called to academic service. You will forever have my admiration, and hopefully more of my tax dollars in the future.

Lastly, to my original team, the g Team, Mark, Bill, Jon, Pohl, Ringo, Jules, Mary, Lauri, Jurita, Swapan, Vishal, Keogh, Scott and the others you will forever be in my heart.

# TABLE OF CONTENTS

<b>ABSTRACT</b> .....	<b>II</b>
<b>ACKNOWLEDGMENTS</b> .....	<b>IV</b>
<b>TABLE OF CONTENTS</b> .....	<b>VI</b>
<b>FIGURES</b> .....	<b>IX</b>
<b>TABLES</b> .....	<b>XI</b>
<b>ABBREVIATIONS</b> .....	<b>XII</b>
CHAPTER I: INTRODUCTION.....	1
<i>1.1 Chromatin and cellular identity</i> .....	2
<i>1.2 Chromatin and cancer</i> .....	4
<i>1.3 Metastasis</i> .....	8
1.3.1 Breast cancer and the metastatic cascade .....	10
1.3.2 Tumor cell dormancy.....	14
1.3.3 Experimental models of metastasis .....	16
1.3.4 Metastasis to the adrenal gland.....	22
CHAPTER II: MATERIALS AND METHODS.....	27
<i>II.1 Cell lines</i> .....	28
<i>II.2 Cell survival and proliferation assays</i> .....	28
<i>II.3 Lentiviral libraries</i> .....	28
<i>II.4 Flow cytometry</i> .....	29
<i>II.5 Animal Experiments</i> .....	30
II.5.1 Orthotopic injections.....	30
II.5.2 Intracardiac injections .....	31
II.5.3 Whole animal imaging.....	31

<i>II.6 Immunohistochemistry and Immunocytochemistry</i> .....	31
<i>II.7 Migration and invasion assays</i> .....	32
<i>II.8 Tumor Spheres</i> .....	32
<i>II.9 Adrenal hormone viability screen</i> .....	33
II.9.1 Dopamine response curve.....	33
II.9.2 PKA activity.....	33
<i>II.10 Dopamine quantal release</i> .....	34
<i>II.11 Western blots</i> .....	35
<i>II.12 Hairpin recovery from tumors</i> .....	36
<i>II.13 Bioinformatics</i> .....	41
<i>II.14 qPCR</i> .....	43
CHAPTER III: ESTABLISHMENT OF A NOVEL <i>IN VIVO</i> BREAST CANCER METASTASIS MODEL.....	44
<i>Abstract</i> .....	45
<i>III.1 The MDA-MB-361 cell line</i> .....	46
<i>III.2 In vivo characterization using orthotopic injection of 361-luc</i> .....	49
III.2.1 Histopathology of orthotopic 361-luc tumors.....	51
III.2.2 Immunohistochemical analysis of orthotopic 361-luc tumors .....	53
<i>III.3 Systemic organ seeding of 361-luc via intracardiac injection</i> .....	55
III.3.1 Histopathology of 361-luc adrenal tumors.....	56
III.3.2 Immunohistochemical analysis of adrenal 361-luc tumors .....	59
III.3.3 Histopathological characteristics of brain metastases .....	61
III.3.4 Histopathological characterization of ovarian metastases.....	63
<i>III.4 Inherent adrenal tropism of the 361-luc line</i> .....	65
III.4.1 Investigation of dopamine as a contributor to 361-luc viability.....	69
III.4.2 Dopamine signaling in 361-luc cells.....	75
III.4.3 Functional dopamine electrophysiology .....	78
CHAPTER IV: AN <i>IN VIVO</i> SCREEN OF CHROMATIN MODIFIERS IN METASTATIC COLONIZATION .....	80
<i>Abstract</i> .....	81

<i>IV.1 A screen for chromatin modifier involvement in metastatic colonization</i> .....	82
<i>IV.2 Expected outcome</i> .....	86
<i>IV.3 Execution of the screen</i> .....	86
IV.3.1 Bioinformatic analysis .....	96
<i>IV.4 Nonadrenal tumor analysis reveals targeting of EHMT1, JMJD5, PRDM14 and SIRT3</i> .....	99
IV.4.1 Validation of the SIRT3 hit .....	106
CHAPTER V: DISCUSSION .....	115
<i>V.1 Conclusions and significance of findings</i> .....	116
V.1.1 361-luc system .....	116
V.1.2 361-luc cells exhibit organ specific colonization .....	118
V.1.3 361-luc cells have an inherent neurotropism .....	120
V.1.4 361-luc cells respond to and are capable of synthesis, storage and release of dopamine .....	121
V.1.5 Limitations and future directions .....	125
<i>V.2 Chromatin modification and metastasis of 361-luc cells</i> .....	128
V.2.1 Chromatin modifier knock down screen of systemic colonization .....	128
V.2.2 Significance of findings and future directions .....	133
CHAPTER VI: APPENDIX .....	136
<i>VI.1 The RasGAP gene, RASAL2, is a tumor and metastasis suppressor</i> .....	141
<i>Abstract</i> .....	141
<i>VI.2 Using defined finger-finger interfaces as units of assembly for constructing zinc-finger nucleases.</i> .....	156
<i>Abstract</i> .....	156
CHAPTER VII: REFERENCES .....	168

## FIGURES

Figure I.1 The metastatic cascade.	11
Figure I.2 The evolution of clonal populations.	13
Figure I.3 In vitro models of metastasis.	17
Figure I.4 Structure of the human and mouse mammary gland.	19
Figure I.5 Differential organ colonization from tail vein and intracardiac delivery.	21
Figure I.6 Anatomy of the adrenal gland.	24
Figure I.7 Capillary structure.	25
Figure II.1 Statistical analysis for Interspike interval.	35
Figure III.1 361-luc cell morphology.	47
Figure III.2 Proliferation and doubling time of 361-luc cells in full media.	48
Figure III.3 361-luc FACS staining of mammary progenitor markers.	48
Figure III.4 Orthotopic growth of 361-luc tumors in the mammary fat pad of immunocompromised NOD/SCID mice.	50
Figure III.5 Bioluminescence of the 361-luc and 361-PLN lines.	50
Figure III.6 Bioluminescent imaging of orthotopic 361-luc tumors.	51
Figure III.7 Histology of 361-luc mammary fat pad tumors.	52
Figure III.8 Immunohistochemical staining of orthotopic 361-luc tumors.	54
Figure III.9 361-luc tumors arising from intracardiac injection over time.	55
Figure III.10 Adrenal metastases from intracardiac injection of 361-luc.	58
Figure III.11 Higher magnification of adrenal boundaries.	59
Figure III.12 IHC of 361-luc adrenal metastases.	60
Figure III.13 Histopathological staining of 361-luc brain metastases.	62
Figure III.14 Histopathologic characterization of 361-luc ovarian metastases.	64
Figure III.15 In vitro viability of 361-luc cells at 48 hours treated with hormones synthesized by the adrenal glands.	67
Figure III.16 In vitro viability of 361-luc cells at 96 hours treated with hormones synthesized by the adrenal glands.	68
Figure III.17 Dopamine increases viability of 361-luc cells.	69
Figure III.18 Dopamine signal transduction.	70
Figure III.19 Dopamine receptor expression in 361-luc cells.	71
Figure III.20 Dopamine reuptake receptors and biosynthetic enzymes in 361-luc cells.	72
Figure III.21 Pramipexole stimulation of 361-luc.	73
Figure III.22 Dose response curve of 361-luc cells.	74
Figure III.23 Px stimulation and vesicular machinery.	75
Figure III.24 Bioluminescent signal due to PKA activity upon dopamine stimulation.	76

Figure III.25 361-luc cells do not express the dopamine signal transducer DARPP-32.	77
Figure III.26 Immunofluorescent staining of Ki67 in 361-luc cells treated with 1uM dopamine.	77
Figure III.27 Depolarization induced quantal release in MDA-MB-361-luc cells.	79
Figure IV.1 Experimental approach for chromatin modifier knock down screen of 361-luc cells and metastatic colonization.	84
Figure IV.2 Bioinformatics scheme for recovery and analysis of hairpins from tumors.	85
Figure IV.3 Expression contexts for the two chromatin modifier libraries.	88
Figure IV.4 MOI determination for pHAGE library transduction.	89
Figure IV.5 Representation of all hairpins in the pooled pLKO plasmids.	91
Figure IV.6 Metastatic outgrowth over time of the pHAGE library.	92
Figure IV.7 Metastatic outgrowth over time of the pLKO library.	93
Figure IV.8 Bioluminescent signal detected per organ.	95
Figure IV.9 Per base sequence content of library reads.	96
Figure IV.10: Distribution of pre-screen pellet hairpins.	97
Figure IV.11 Probability of relapse free survival of breast cancer patients.	103
Figure IV.12 Probability of relapse free survival of luminal B breast cancer patients.	104
Figure IV.13 SIRT3 expression and probability of clinical disease progression in breast cancer patients.	105
Figure IV.14 Knock down of SIRT3 in 361-luc cells.	107
Figure IV.15 FACS analysis of CD49f and EpCAM mammary progenitor markers on shSIRT3 361-luc cells.	108
Figure IV.16 Transwell invasion by shSIRT3 361-luc cells.	108
Figure IV.17 Viability and stemness of shSIRT3 361-luc.	109
Figure IV.18 ROS measurement of shSIRT3 knock down 361-luc cells.	110
Figure IV.19 Orthotopic injection of shSIRT3 361-luc cells.	112
Figure IV.20 Mean total flux per hairpin group.	113
Figure IV.21 Colonization by shSIRT3 361-luc cells following UGIC.	114
Figure V.1 Potential mechanisms for dopamine in mediating metastasis.	125

## TABLES

Table I.II Characteristics of breast cancer subtypes.	10
Table II.I pHAGE PCR oligos, bar codes and tumor cross reference.	37
Table II.II pLKO PCR oligos, bar codes and tumor cross reference.	39
Table II.III qPCR primers.	43
Table III.I Latency and penetrance of bioluminescent tumors after injection of 1 million 361-luc cells into the left ventricle of NOD/SCID mice.	56
Table III.II Reagents and cell lines used for in vitro screen of adrenal gland hormone viability.	66
Table IV.I Genes targeted by RNAi hairpin libraries.	87
Table IV.II Experimental groups used for chromatin modifier metastasis screen.	90
Table IV.III Latency and penetrance of bioluminescent tumors after injection of 1 million 361-luc cells into the left ventricle of NOD/SCID mice.	94
Table IV.IV Tumor and hit measurements for pLKO and pHAGE grouped by nonadrenal vs adrenal.	98
Table IV.V pHAGE hairpin enrichment analysis of non-adrenal vs adrenal tumors.	99
Table IV.VI pLKO hairpin enrichment analysis of non-adrenal vs adrenal tumors.	101
Table IV.VII Hairpins present in monoclonal tumors.	102
Table VI.I The histone code: chromatin marks, modifying enzymes and their functions.	137

## ABBREVIATIONS

2HG	2-hydroxyglutarate
aKG	a-ketoglutarate
AML	acute myeloid leukemia
cAMP	cyclic adenosine monophosphate
CIMP	CpG island methylator phenotype
CpG	cytosine-phosphate-guanine dinucleotide
CTCs	circulating tumor cells
DCIA	ductal carcinoma in situ
DNA	deoxyribonucleic acid
DNMT	DNA methyl transferase
DRD	dopamine receptor
EF1a	eukaryotic elongation factor 1 alpha
EMT	epithelial to mesenchymal transition
ER	estrogen receptor
GBM	Glioblastoma
H&E	Hematoxylin and eosin
H3K27me3	trimethylated lysine 27 of histone 3
H3K4me3	trimethylation of lysine 4 on histone 3
H3K9Ac	acetylated lysine 9 of histone 3
Her2	human epidermal growth factor receptor 2
HMEC	human mammary epithelial cell
IDH1	isocitrate dehydrogenase 1
KDM4C	lysine specific demethylase 4C
MAPK	mitogen activated protein kinase
MCRA	most common recent ancestor
Mir30	micro RNA 30
MITF	microphthalmia-associated transcription factor
MOI	multiplicity of infection
PKA	protein kinase A
PNAS	Proceedings of the National Academy of Sciences
Px	Pramipexole
qPCR	quantitative polymerase chain reaction
RFP	red fluorescent protein
RNA	ribonucleic acid
RNAi	RNA interference
SETDB1	Su(var)3-9/Enhancer of zeste domain, bifurcated 1
TCGA	the cancer genome atlas
TET2	Tet methyl cytosine dioxygenase 2

TH	tyrosine hydroxylase
UGIC	ultrasound guided intracardiac injection
VMAT	vesicular monoamine transporter
ZFN	zinc finger nuclease

# Chapter I:

# Introduction

## **I.1 Chromatin and cellular identity**

All the diploid cells within a Metazoan organism have identical genomes. A terminally differentiated neuron has the exact same genetic information as a germ cell. Thus it is control of the expression and amount of gene products that determines cellular identity and function. Epigenetic expression control is one mechanism important in cell fate. In whole organism development, tissues arise through a variety of mechanisms in a process including cellular specification and commitment (Gilbert, 2014). Generally, once a cell type is specified and the cell is committed, its fate is established. This cell now has a limited repertoire of proteins and those in other cellular programs are forbidden. The aforementioned neuron will never undergo meiosis like the germ cell may. The meiosis genes in this neuron are still present but can be thought of as silenced. Physically this is manifested by the state of both the deoxyribonucleic acid (DNA) that makes up these genes and the histone proteins with which DNA complexes. Silent genes are chemically modified by methylation of cytosine (Hotchkiss, 1948). In animals, the enzymes responsible for cytosine methylation are a family called DNA methyl transferases (DNMTs) (Yen et al., 1992). These enzymes act on cytosine immediately 5' of guanine within a strand of DNA. This dinucleotide arrangement is abbreviated CpG to denote the linking phosphate found between them within DNA strands and to distinguish it from hydrogen bonding that occurs between cytosine and guanine in base pairing. CpG dinucleotides are distributed in clusters, called CpG islands, which are more frequently found within promoters and gene control elements (Bird et al., 1985; Saxonov et al., 2006). When the CpG islands in a gene promoter are methylated the gene is less likely to be expressed (Bird, 2002).

DNA methylation is not the only phenomenon whereby gene access is physically restricted. Histones, known to be the scaffolds of DNA condensation, also play an important role in gene access. The basic unit of DNA structure is the nucleosome consisting of an octamer of histones wound nearly twice by a DNA double strand. In addition to serving as the core for nucleosome winding, histones also connect nucleosomes (2014a). To get an idea of how pliable DNA is consider that the longest human chromosome, numbered one, is nearly 250 million base pairs and when unwound would be 8.5 centimeters long. Clearly not all of this DNA can be accessible within a micrometer scale nucleus so DNA exists exclusively in a somewhat compacted state. The term chromatin is used to refer to the nucleic acid and histone complex. Although the reality is more complex, an open or closed dichotomy is often useful when referring to chromatin state as it relates to gene access. Less compacted, open chromatin, which can associate with RNA polymerases and other transcription factors is termed euchromatin, while closed, more compacted chromatin with tightly packed inactive genes is called heterochromatin.

Similar to how DNMTs chemically modify DNA there are a variety of enzymes that modify histones. The lysine residues are frequently modified with acetyl and methyl groups. Enzymes that add these modifications are termed histone writers and those that remove them erasers. The writing and erasing of these marks is analogous to a written language or histone code that can be read by other proteins and whose words have functional consequences (Jenuwein, 2001). In fact there are activating marks such as histone 3 trimethylation of lysine 4 (H3K4me3) or acetylation of lysine 9 (H3K9Ac) and repressive marks like histone 3 trimethylation of lysine 27 (H3K27me3) (Santos-Rosa et al., 2002; Schuettengruber et al., 2007; Vermeulen and Timmers, 2010; Karmodiya et al.,

2012). Histone lysines can be marked with one, two, or three methyl groups for which a host of reading enzymes exist that can distinguish the number of groups and differentially respond (Bannister et al., 2001; Strahl et al., 1999). Unlike methylation, acetylation of lysine eliminates positive charge and directly alters DNA binding. These marks are known to control transcriptional activation and repression, as well as DNA damage and DNA condensation (Marmorstein and Zhou, 2014). A gene is considered effectively shut off, or silenced, when a combination of repressive histone marks and DNA methylation is present. There has recently been a great deal of interest in chromatin post-translational modification with high sensitivity mass spectrometry identifying many other marks the consequences of which are only now being explored (Chen et al., 2007; Tan et al., 2011; Xie et al., 2012). Table VI.I provides a comprehensive list of currently identified histone marks and their associated functions.

## **I.2 Chromatin and cancer**

As cellular identity is determined epigenetically the question arises if abnormal chromatin state could allow a cell to sample an ordinarily forbidden program. Cancer can be described as a disease of lost cellular identity in which cells stop behaving as what they are and through a series of genetic insults begin to multiply uncontrollably leading eventually to disruption of their host tissue and ultimately to invasion and metastasis. Observations of abnormal chromatin in tumor cells were first made in 1983. It was discovered that human cancer cells had lower levels of CpG methylation than their corresponding normal tissues (Feinberg and Vogelstein, 1983). This hypomethylation implied that there are more genes accessible to a tumor cell.

Shortly thereafter it was discovered that in addition to global hypomethylation there were distinct regions that were hypermethylated in tumors. This work demonstrated in retinoblastoma to be a mechanism for silencing the tumor suppressor gene RB (Greger et al., 1989). This was the first instance where abnormal chromatin state was linked mechanistically to tumor progression. Pioneering work in colorectal cancer identified a pattern of promoter silencing with a CpG island methylator phenotype (CIMP) distinguishable from age related and individually cancer associated CpG island methylation (Toyota et al., 1999). The colorectal CIMP was found to include distinct tumor suppressors consistently hypermethylated in a subset of tumors. This subset had a marked lack of mutation in the tumor suppressor p53 and very high incidence of oncogenic mutation of either KRAS or BRAF (Toyota et al., 2000). With the advent of the cancer genome atlas (TCGA) CIMPs have been found in breast and endometrial tumors as well as in glioblastomas and acute myeloid leukemias (Weisenberger, 2014). The most recent genome-wide analyses using unsupervised clustering have described ten different malignancies with objectively defined CIMPs (Suzuki et al., 2014). Clustering of CIMP tumors implies that methylation is not random and therefore could be selected for within a neoplasm.

The question of whether these chromatin changes observed in malignancies are causative, contributing to the malignant progression of cells, or rather a consequence of the transformed state is a difficult one to answer. It requires demonstration of a chromatin modifier mediated event that experimentally drives malignant progression in a context that is present in actual human disease. Such a story recently emerged when the isocitrate dehydrogenase 1 (IDH1) gene was revealed to show the R132H mutation in 12% of

glioblastoma (GBM) tumors (Parsons et al., 2008). This mutation was found to correlate tightly with the existence of a GBM CIMP and importantly overexpression of the mutant can establish the CIMP experimentally in primary human astrocytes (Turcan et al., 2012). How exactly this occurred was revealed by looking at IDH1 mutant leukemias. The R132H mutation was found in CIMP positive acute myeloid leukemia (AML) where it was mutually exclusive to mutations in the mitochondrial homolog IDH2, implicating the isocitrate dehydrogenase activity in the establishment of the CIMP (Figueroa et al., 2010). IDH is an enzyme in the tricarboxylic acid cycle responsible for converting isocitrate to  $\alpha$ -ketoglutarate ( $\alpha$ KG). Surprisingly, the mutant IDH results in conversion to a rare metabolite 2-hydroxyglutarate (2HG) instead (Dang et al., 2009). Subsequent buildup of 2HG inhibits several chromatin modifiers including the DNA demethylase TET2 and the lysine demethylase KDM4C resulting in chromatin changes at both the nucleic acid and the histone level which have been proven experimentally to alter gene expression and block differentiation of non-transformed cells (Xu et al., 2011; Lu et al., 2012). In essence this was experimental validation that dysregulation of chromatin modifiers can alter cellular identity and drive malignant progression.

Melanoma is a disease in which the changes from a benign nevus to a pre-cancerous lesion and eventually a fully transformed and invasive cancer can be studied chronologically at the same site in the same individual. Progression is known to be driven by the V600E mutant kinase BRAF (Davies et al., 2002). This mutation results in hyperactive MAPK signaling but can be found early in benign moles, and itself is not sufficient for transformation at least in part because cells in the benign tumors become senescent. The amplification of specific regions can control the development of melanoma

and microphthalmia-associated transcription factor (MITF) is one gene frequently amplified which when coexpressed with mutant BRAF is able to transform primary human melanocytes (Garraway et al., 2005). Another region frequently amplified in melanoma is 1q21. An analysis performed on this region in zebrafish determined that the histone methyl transferase SETDB1 was amplified and could similarly drive melanoma progression (Ceol et al., 2011). In patients, SETDB1 amplification was found in melanomas, but not benign nevi or normal melanocytes. SETDB1 catalyzes the repressive mark H3K9me3 which in forced overexpression was found along with the SETDB1 protein itself to more readily occupy the promoters of HOX genes. Interestingly, this pattern of SETDB1 and H3K9me3 occupation of HOX genes was initially described, using a chromatin modifier knock down screen, in mouse embryonic stem cells to repress developmental regulators and maintain pluripotency (Bilodeau et al., 2009). This finding extends beyond melanoma as SETDB1 is known to be focally amplified in breast cancer and other tumors (Liu et al., 2015; Rodriguez-Paredes et al., 2014).

The subversion of the developmental epithelial to mesenchymal transition (EMT) program by cancer cells provides a survival advantage, increases stemness and facilitates invasion and metastasis by converting non-stem tumor cells to a more stem-like state. It has been shown that the pleiotropic chemokine TGF beta is capable of mediating this EMT induction in basal breast cancer cells in part by driving transcription of the ZEB1 EMT transcription factor to convert non-stem cancer cells to cancer stem cells. This switch in phenotype coincides with the chromatin state of the ZEB1 promoter which in untransformed basal cells is repressed, but found to contain repressive and active marks in basal cancer cells and to contain only active marks in basal cancer stem cells (Chaffer et

al., 2013). Cancer stem cells exhibit most if not all of the characteristics required for metastatic progression discussed in more detail later and a strong case can be made that cancer stem cells are the effectors of metastasis (Baccelli et al., 2013; Thiery et al., 2009).

The three above examples provide evidence that chromatin modifiers contribute to malignancy at multiple levels. TET2 and KDM4C can act to reprogram a cell to a cancer promoting phenotype. SETDB1 in melanoma can drive an already altered cell into a fully transformed cell. And, as yet unrevealed chromatin modifying machinery is involved in the transition from a non-stem to a cancer stem cell via EMT subversion. Perhaps the most compelling evidence that chromatin modifiers can drive malignant progression come from genome scale studies which have identified recurring mutations in specific cancers of DNA demethylases and methyl transferases as well as histone methyl transferases, demethylases and acetyl transferases (Garraway and Lander, 2013).

### **I.3 Metastasis**

In 1889 an assistant surgeon in a London hospital, Stephen Paget, was perplexed about the distribution of metastatic tumors in fatal cases of breast cancer. He noted that secondary growths occurred only in specific organs and for some reason the majority of the secondary tumors were found in the liver. Thus there was something about the liver which made it more hospitable to certain breast cancers. He posited that cancer cells which escaped the primary tumor were like seeds distributed throughout the body that needed fertile soil in which to grow if they were to establish an overt metastasis (Paget, 1889).

In the 125 years since Dr. Paget's observations, a great deal of time and treasure has been spent studying and learning about how and why cancers grow. We now

understand that cancer is a genetic disease which progresses stepwise as mutations accumulate driving cells through normal physiologic safeguards unleashing a destructive proliferative force. When cancers arise in non-vital organs like the breast, surgery can ordinarily eliminate the immediate threat but the costs in disfigurement and future uncertainty cannot be quantified. Many patients relapse because the primary tumor has planted its seeds elsewhere in the body in organs whose function is essential for life. It may take many years before these seeds germinate and eventually become overt metastases capable of disrupting organ function. Thus, metastasis is the ultimate stage of malignant progression and in 2012 breast cancer killed over half a million people worldwide (Torre et al., 2015). In the US it is estimated that approximately 40,000 women will die from the disease this year (Siegel et al., 2015).

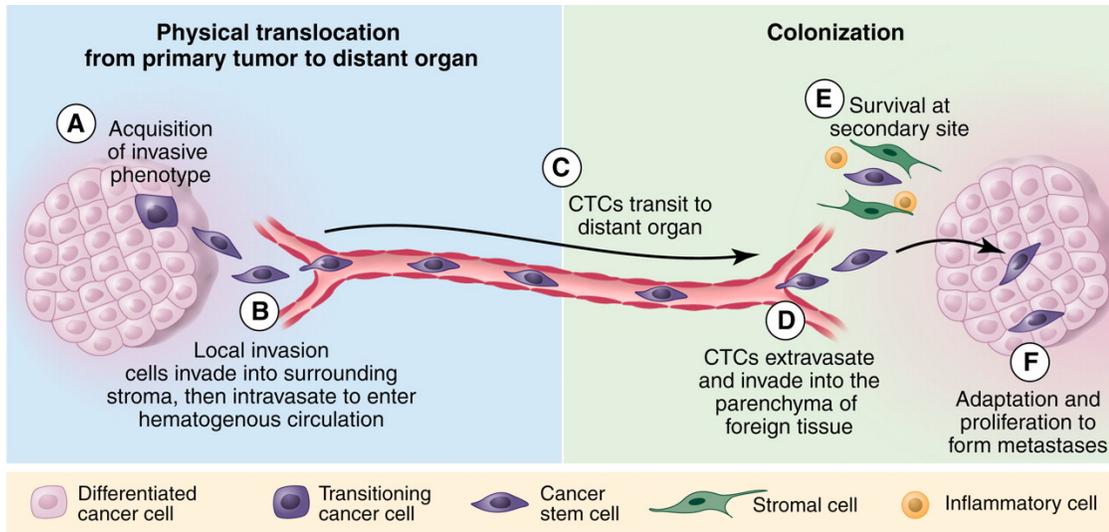
Paget observed most metastases in the liver but there were other patients with healthy livers who died from growths in the lungs, bone or brain and occasionally had tumors in multiple organs. Paget's described organ tropism did not fit the monolithic view of the disease at the time. It demanded a complexity not before appreciated. In fact, we now know that breast cancer is a group of heterogeneous diseases which can be stratified based on histopathologic characteristics and molecular patterns of expression into multiple subtypes (Perou et al., 2000; Reiner et al., 1988). Depending on the resolution of molecular stratification a current comprehensive taxonomy of human breast tumors can have seven or more subtypes, but for the purpose of this discussion it is helpful to consider the four groups listed in Table I.II determined by the expression of the estrogen hormone receptor (ER), the human epidermal growth factor receptor 2 (Her2) and the proliferative marker Ki67.

**Table I.II: Characteristics of breast cancer subtypes.** Compiled from (Kennecke et al., 2010)

Subtype	Molecular characteristics	Incidence of colonization				10 year survival	Mean survival with distant metastasis (years)
		Bone	Liver	Lung	Brain		
Luminal A (N=1.639)	ER+ Ki67 low	<b>19%</b>	8%	7%	2%	70%	2.2
Luminal B (1137)	ER+ Ki67 high	<b>30%</b>	15%	15%	6%	55%	1.6
Her2 enriched (266)	ER- Her2+	<b>30%</b>	23%	24%	14%	48%	0.7
Basal like (367)	ER- Her2-	17%	9%	<b>19%</b>	11%	53%	0.5

### I.3.1 Breast cancer and the metastatic cascade

Metastasis is the ultimate phase of malignant progression which can be conceptualized as a cascade of events (Figure I.1) including escape from the primary tumor by cells that penetrate into the local stroma, entry into efferent lymphatic or blood vessels, survival in the circulation to a distant site, deposition in a foreign milieu and adaptation to, possibly including subversion of, the new environment leading to proliferation and growth into an overt secondary tumor (Fidler et al., 1978; Chaffer and Weinberg, 2011).



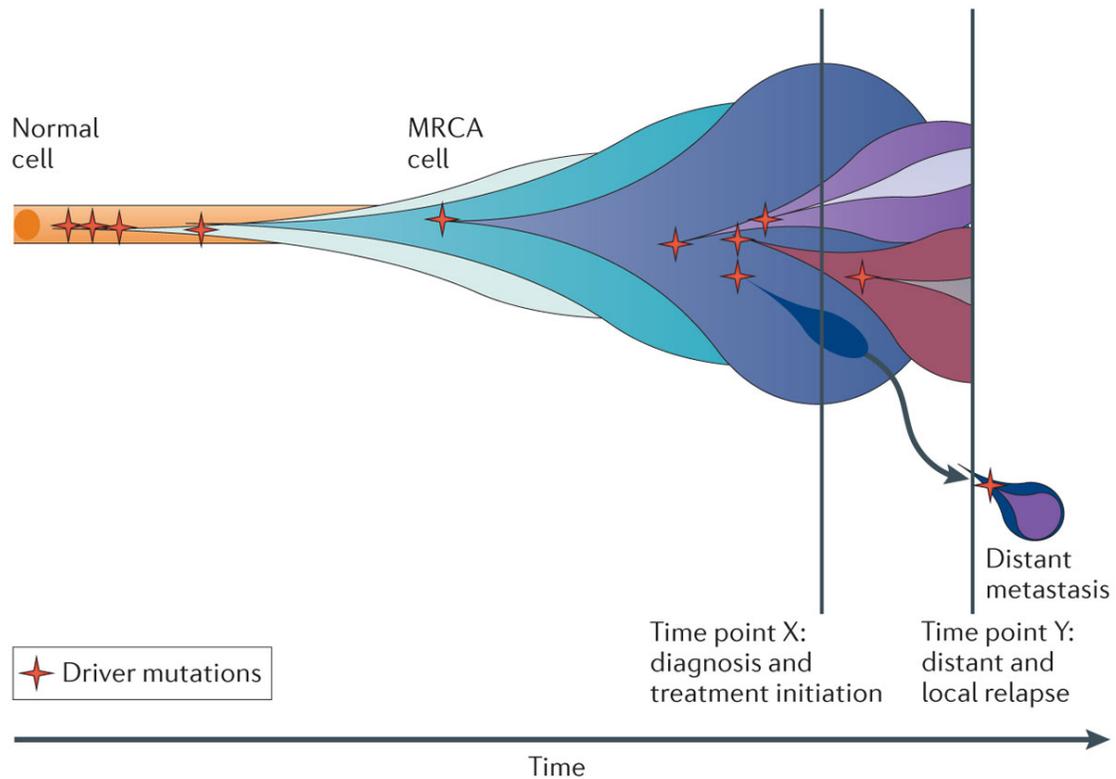
**Figure I.1: The metastatic cascade.** Metastasis can be envisioned as a process that occurs in two major phases: (i) physical translocation of cancer cells from the primary tumor to a distant organ and (ii) colonization of the translocated cells within that organ. (A) To begin the metastatic cascade, cancer cells within the primary tumor acquire an invasive phenotype. (B) Cancer cells can then invade into the surrounding matrix and toward blood vessels, where they intravasate to enter the circulation, which serves as their primary means of passage to distant organs. (C) Cancer cells traveling through the circulation are CTCs. They display properties of anchorage-independent survival. (D) At the distant organ, CTCs exit the circulation and invade into the microenvironment of the foreign tissue. (E) At that foreign site, cancer cells must be able to evade the innate immune response and also survive as a single cell (or as a small cluster of cells). (F) To develop into an active macrometastatic deposit, the cancer cell must be able to adapt to the microenvironment and initiate proliferation. From Christine L. Chaffer and Robert A. Weinberg. *Science* 25 March 2011: 331 (6024), 1559-1564. Reprinted with permission from AAAS.

Certain cancers like melanoma have a high rate of metastasis such that once the dermis is invaded, metastasis to distant organs is nearly absolute (Miller and Mihm, 2006), while others, like the more frequent skin malignancy basal cell carcinoma, metastasize less than 0.5% of the time (Weiss and Korn, 2012). In breast cancer, the four major subtypes have differing metastatic proclivities in that they differ in their metastatic aggression as well as the organs that they ultimately colonize. An ongoing study of 111 patients with breast cancer metastases diagnosed between 2002 and 2007 conducted in Sweden reported that these four subtypes are independently predictive of survival after relapse providing the

most recent confirmation that the molecular characteristics of the primary tumor are important in metastasis as well (Tobin et al., TEX Trialists Group, 2015). In general, Her2 and basal-like tumors are more prone to metastasis and do so earlier than luminal tumors with luminal A tumors having the best overall prognosis.

There is also variability in the organ tropism exhibited by different cancers. Her2 and basal-like tumors metastasize to the liver, brain and lung ordinarily less than five years after primary tumor removal (Soni et al., 2015), while luminal tumors metastasize at a more constant rate up to twenty years post-surgery with a preference for bone (Kennecke et al., 2010; Smid et al., 2008).

The clonal selection hypothesis illustrated in Figure I.2 holds that cancers are driven stepwise by key acquired mutations causing loss of tumor suppressor function and gain of oncogenic abilities. A natural extension of this theory is that the progression to the final phase, metastasis, occurs by one or more analogous mutations after escape from the primary tumor that provide a metastatic advantage.



Nature Reviews | **Genetics**

**Figure I.2 The evolution of clonal populations.** Cancers arise via acquisition of driver mutations creating a population of genetically distinct clones (represented by colored plumes) within a tumor. Here all of the clones at diagnosis (time point X) and relapse (Y) arise from a single most recent common ancestor (MRCA). It is notable that some clones which were dominant early have died out before diagnosis and other clones have emerged after the pictured metastasis seeding clone escaped. From (Yates and Campbell, 2012)

There is evidence that this can occur (Minn et al., 2005; Gupta et al., 2007), but an increasing number of genome scale studies analyzing primary and distant clones suggest that this is rare (Bozic et al., 2010; Campbell et al., 2010; Ding et al., 2010; Yachida et al., 2010). These comprehensive studies in which distant tumors were sequenced at high coverage and compared to the primary tumor have found that metastases arise from clones already present in the primary tumor without acquisition of additional driver mutations. One comprehensive study examined animal models of spontaneous metastasis comparing highly metastatic lung, breast and kidney cancer tumor line variants to the weakly

metastatic founding lines and found very little genetic divergence in the metastatic lines. They diligently focused in on known cancer drivers and did find mutations, but they did not correlate with changes in gene expression, and they were not enriched or depleted in the metastatic populations (Jacob et al., 2015). These new mutations are most likely the consequence of the hallmark genomic instability of transformed cells and are therefore merely passenger events. Instead reports of tumor cells reconfiguring existing oncogenic signals have contributed to the move away from the concept of metastasis specific mutations (Chen et al., 2011; Oskarsson et al., 2011; Sethi et al., 2011; Zhang et al., 2013).

### **I.3.2 Tumor cell dormancy**

Observational and experimental evidence abounds that the rate limiting step of the metastatic cascade is the final adaptation to the new environment where a single cell or micro metastasis acquires the ability to proliferate (Kang and Pantel, 2013). The escape from the primary tumor has been well defined mechanistically with cells gaining an invasive phenotype by expressing matrix remodeling proteins or inducing stromal players to do this for them and also by recruiting endothelial cells to establish a vascular supply for the tumor. The abnormal tumor vasculature, frequently termed leaky, provides an entry point for motile tumor cells such as those which have undergone oncogenic EMT (Thiery et al., 2009). Tumor cells can be found in the circulation in many solid tumors and the number of circulating tumor cells (CTCs) is directly correlated with risk of relapse. When tumors are eventually detected clinically, larger tumors also exhibit a greater number of CTCs compared to smaller tumors. The number of cells in circulation at any one time originating from a primary tumor is on the scale of thousands (Baccelli et al., 2013). As patients do not present with thousands of metastases, these cells are either unimportant for

metastasis, which seems unlikely, or the colonization phase of the metastatic cascade presents significant barriers.

Almost all of the cells that escape a primary tumor die (Luzzi et al., 1998). It is not understood why this is. Mouse experiments have been done that show normal mammary epithelial cells can survive lodged ectopically in the lungs for months. When forced oncogene overexpression is induced they awaken and become tumors (Podsypanina et al., 2008). But the relevance of this experiment to clinical disease is not entirely clear given rodent cells transform more easily than do human cells, and human metastases are believed to be founded by wholly transformed cells not normal epithelial cells. In a landmark study of over 500 patients with breast cancer, 36% of the patients and 1% of non-diseased individuals, were found to have epithelial cells present in their bone marrow at the time of surgery. Four years later 25% of the patients with epithelial cells in their bone marrow died of cancer, while only 6% of the breast cancer patients without them died (Braun et al., 2000). It is noteworthy that this statistic includes all cancer deaths, not specifically those due to bone metastasis. Interestingly, tumor cells can be found in the bone marrow of patients with ductal carcinoma *in situ* (DCIS) as well and are also of prognostic value for disease progression at this early stage (Sänger et al., 2011).

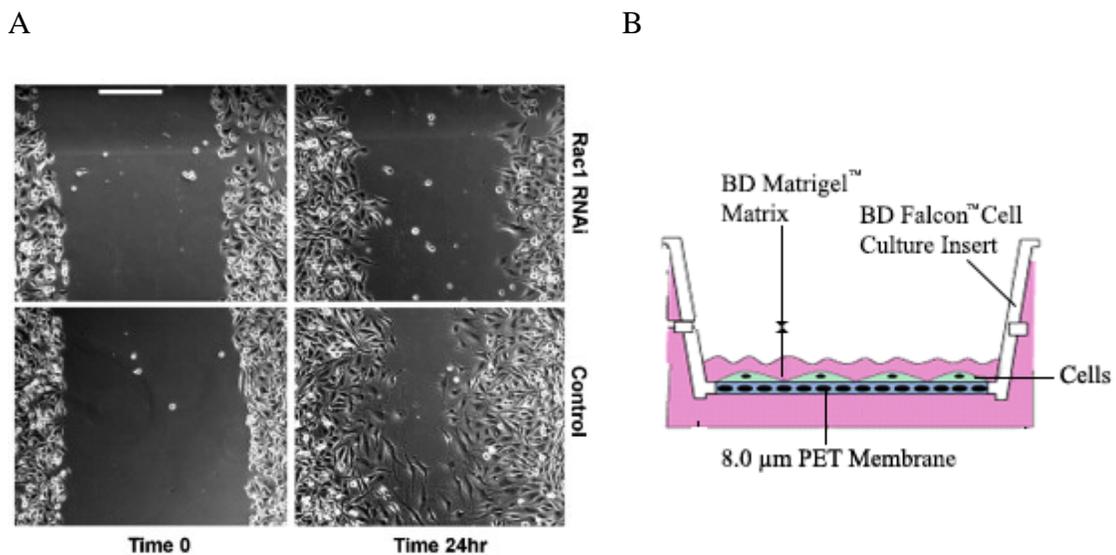
There is a great deal of heterogeneity in the onset of relapse across cancers as well as within breast cancers. Clinical data show two distinct risk peaks for breast cancer metastasis: one at ten months and a second around 30 months post-surgery (Demicheli et al., 2007). Mathematical modeling of clinical data suggest that disseminated tumor cells undergo an extended period of dormancy (Aguirre-Ghiso, 2007). In the case of individual cells this would be absent or severely compromised proliferation. In the case of micro

metastases the proliferation rate would be offset by an equal rate of death. As we know that these dormant, but fully transformed cells do not acquire new oncogenic mutations, tumor cell dormancy, and eventual reawakening must be an epigenetic phenomenon. Experimental evidence explaining dormancy is lacking, but several mechanisms can be envisioned including loss of effective angiogenesis in the new environment, the sudden absence of a stromal signal to which the tumor has become addicted, changes in survival and proliferation signaling and immune system mediation. Discovering how a tumor cell adapts to an environment like the brain or the adrenal gland could reveal important cancer biology mechanisms and has great clinical relevance. In this work I set out to contribute insight into such a non-genetic switch by experimentally modeling metastasis and focusing on organ specific colonization.

### **I.3.3 Experimental models of metastasis**

Useful modeling of the multi-organ nature of metastasis requires complicated experimental systems. Outside of the context of a whole organism it is not possible to model with any fidelity the entire cascade. However, systems of reduced complexity have been used successfully to provide insight into specific steps. For instance, motility can be studied at the cellular level by measuring the movement of cells in culture. The scratch test is a simple assay for motility where a confluent layer of cells is disrupted with a pipette tip and observed to see how quickly the neighboring cells move in (Figure I.3A). A slightly more complicated system which tracks displacement of beads by individual cells has been used successfully in a RNA interference (RNAi) screen to identify motility related genes which were then established to be associated with metastasis as well (van Roosmalen et al., 2015). However, motility alone is not sufficient for metastasis and to recapitulate an

entire step in the cascade requires the involvement of other cellular processes. The first step, local invasion, can be recapitulated *in vitro* by measuring penetration of cells through a simulated basement membrane. The transwell invasion assay accomplishes this by seeding cells on one side of an 8  $\mu$ m porous membrane coated with a basement membrane-like gelatinous sarcoma secretion, Matrigel, and measuring how many cells can penetrate to the other side (figure I.3B). This tool has also been used in a knock down screen that identified a metastasis suppressing role for the FOXO4 transcription factor (Su et al., 2014).



**Figure I.3 *In vitro* models of metastasis.** A) After scratching confluent cells with a pipet tip the rate at which cells move in from the periphery can be measured. From (Valster et al., 2005) B) In the transwell invasion assay cells are plated on top of a Matrigel simulated basement membrane situated above a porous membrane. Cells that traverse the Matrigel and membrane are counted and compared to a well without Matrigel to calculate invasion. From BD package insert.

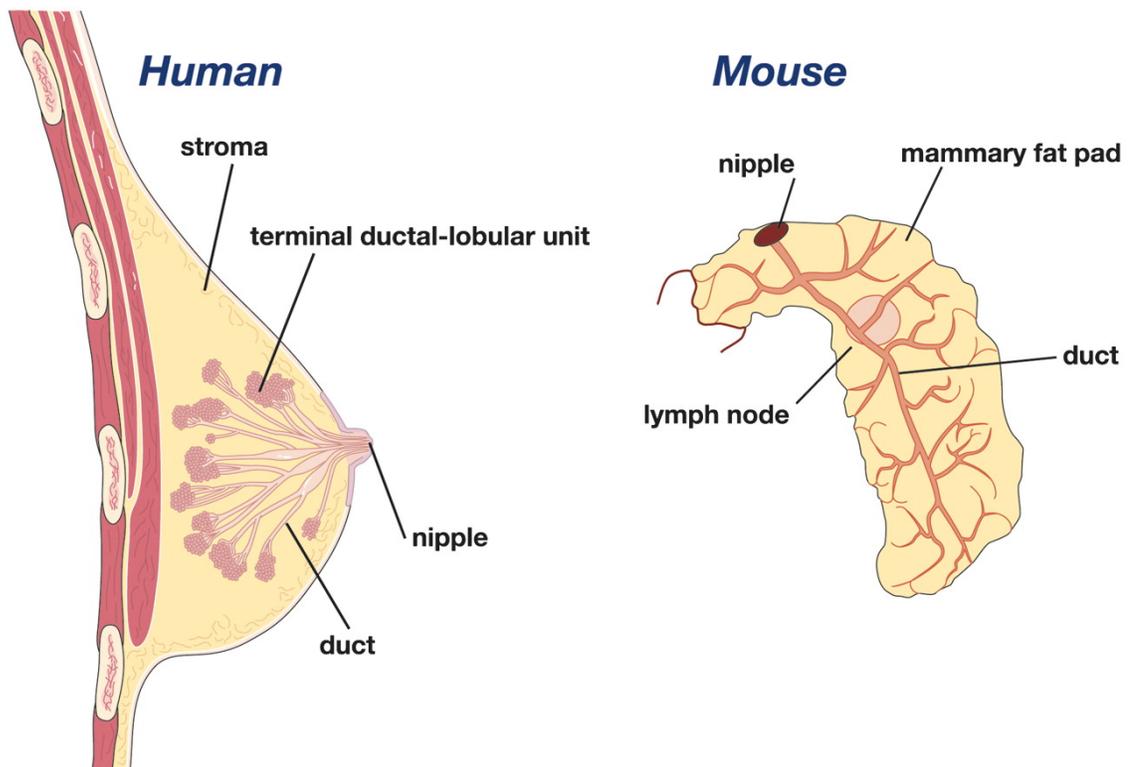
Moving beyond the first step of the metastatic cascade, it becomes increasingly difficult to model the next step *in vitro*. The most comprehensive way to study these processes is by using an entire organism. Many of the currently known mechanisms

underpinning metastasis have been discovered through use of the immunocompromised laboratory mouse. The mouse has the same organ systems and tissues as humans with analogous cell populations and cell biologic processes involved in all steps of the metastatic cascade including similar stromal compartments and players.

There are several ways to use the mouse to study metastasis, each with advantages and limitations. Purely transgenic models like the mouse mammary tumor virus promoter driven polyoma middle T (MMTV-PyMT) provide a system with spontaneous tumor growth in the mammary gland which reproducibly metastasizes to the lungs. This model uses a viral promoter that is restricted to mammary gland expression to drive a viral oncogene which results in the transformation of the mouse mammary epithelium. It has proven useful in studying each step of malignant transformation and progression and recapitulates many of the features of human disease including disease staging, associated biomarkers, loss of endocrine receptors, amplification of oncogenes and recruitment of immune cells (Lin et al., 2003). This model shares the limitations of all traditional transgenic mice that the effects of only one or few genes may be studied. The tumor cells in these models are also obviously not human and species differences such as the relative ease of rodent cell transformation, aging dynamics, evolutionarily divergent intracellular signaling and others are significant (Rangarajan and Weinberg, 2003).

Using tumor cells of human origin circumvents many of the aforementioned limitations. Another popular mouse model to study actual human disease at the patient level uses patient derived xenografts (PDX) in which a piece of a tumor is transplanted directly to an immunocompromised mouse. The location of the transplant is often subcutaneous, but where surgically tractable, orthotopic sites are increasingly used. These models have

the most promise in elucidating late stage mechanisms and therapeutic responses. Unfortunately many transplanted tumors simply do not grow out, and there is some question as to how the tumor which emerges in the immunocompromised mouse shares relevance to the tumor in a person where there are differences in selective pressures and local environments (Whittle et al., 2015). Consider for instance the anatomic differences between a mouse and a human mammary gland in Figure I.4. The ductal structure itself is different with humans exhibiting clusters of ducts that terminate in lobules whereas the mouse has alveolar buds that are formed during each estrous cycle.



**Figure 1.4: Structure of the human and mouse mammary gland.** Adapted from (Visvader, 2009)

Mechanical elements of signal transduction are likely to be different between the two. In the context of metastasis, the source material for the transplant is also important as in heterogeneous diseases like breast cancer, where it is known that the vast majority of tumor

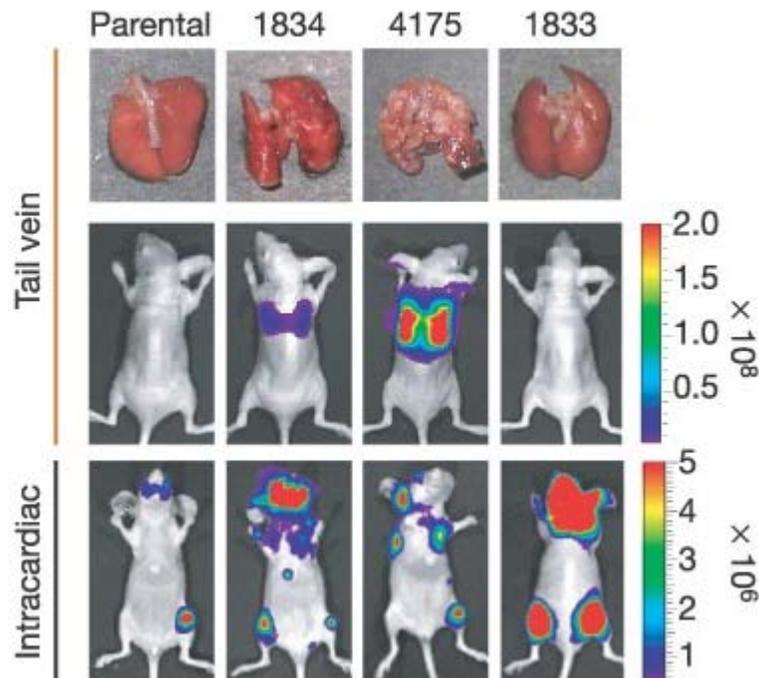
escapes die, there is no way to ensure the transplanted sample includes all relevant clones. In the patient, the metastatic founders are likely already disseminated before the primary tumor is detected.

A third *in vivo* model which encompasses many of the advantages of the previously discussed two models consists of the single cell suspension injection of established tumor lines. The lines may not completely recapitulate the original tumor, with genetic drift occurring during passage in culture, but they are known to grow well and preliminary findings from the mouse can be validated and probed in the same genetic background *in vitro*. A virtually unlimited number of cells can be generated and shared with other investigators to confirm findings. Additionally, genetic manipulation of the tumor line before injection can greatly expand the scope of research. A tag can allow tracing of the cells and its descendants. Genetic manipulations *in vitro* to overexpress or knock down genes allow focused investigation and screening.

An important aspect of tumor cell injection models is the route of delivery. Groundbreaking whole animal metastasis studies were conducted in the Massague lab using the breast carcinoma derived MDA-MB-231 cell line. This line was established by pleural effusion years after removal of the primary tumor in a patient with multi-organ metastasis (Cailleau et al., 1978). The lab was able to characterize a model system utilizing tail vein injection to elucidate a lung metastasis gene signature (Minn et al., 2005). Tail vein delivery is well suited for studying colonization of the lungs because after the right ventricle the venous return deposits the cells directly into the pulmonary capillaries.

To study systemic colonization it is necessary to seed all of the organs in the animal. The only way this can be done is to inject cells into the left ventricle. In the aforementioned

study with tail vein delivery when the cells were instead delivered via intracardiac injection, the lung tropism, relative to the parental line, was no longer observed (Figure I.5). This leads to the conclusion that although tail vein delivery did allow the discovery of a lung metastasis specific gene signature, the delivery primarily into the lungs in the system introduced some bias that would not have existed if the cells were dispersed to many organs. Simply put, if you inject into the lungs you are going to find metastases in the lungs. In other papers the lab used the same line via intracardiac delivery to discover genes involved in brain and bone metastasis as well (Bos et al., 2009; Kang et al., 2003).



**Figure I.5 Differential organ colonization from tail vein and intracardiac delivery.** In this study of MDA-MB-231 derived cell lines, sub-line 4175 (third column) is described as strongly metastatic to the lung. This seems evident in the lung tumor burden and imaging after tail vein delivery in the top two rows of images. But in the bottom row when the entire circulation is seeded notice the lack of growth in the 4175 lung and the drastic change in brain and bone metastases present in all four lines. Adapted from (Minn et al., 2005)

In summary, *in vitro* models are excellent tools to study the early stages of the metastatic cascade where they may work well for testing mechanisms of an individual step.

Among the mouse models of metastasis, purely transgenic models like PyMT have proven invaluable for learning about transitional mechanisms between steps but are limited in modeling human cell and stromal traits. PDX models are proving to be good tools for therapeutic modeling of engraftment and individual tumor responses. But to comprehensively model metastasis in multiple organs, intracardiac delivery of a tumor cell line, as described here, has greater discovery potential in the context of organ tropism and tumor cell adaptation.

### **I.3.4 Metastasis to the adrenal gland**

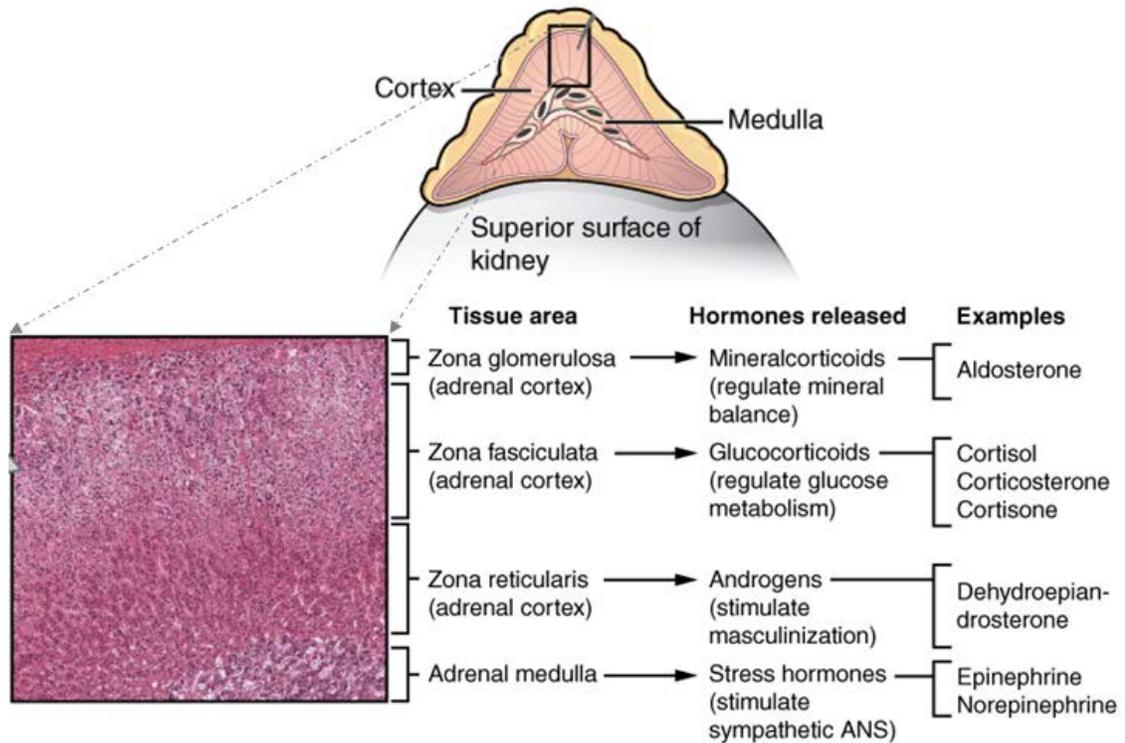
Metastasis to the adrenal gland in cancer is surprisingly common. A large autopsy study performed in Massachusetts on 3,827 cancer patients that died from 1914 to 1943, who received no chemotherapy or radiation treatment, found that the adrenal gland was the fourth most common site of secondary growth after the liver, lung and bone (diSibio and French, 2009).

The adrenal gland is an endocrine organ residing anterior to the kidney in which the synthesis and delivery to the circulation of mineralocorticoids, glucocorticoids, androgens and stress hormones occurs. Figure I.6 shows the gland consists of an outer capsule made up of connective tissue encasing a cortex and medulla. The adrenal cortex is functionally divided into three layers each responsible for the synthesis and release of the aforementioned hormone classes. The outermost zona granulosa is the source of mineralocorticoids. The middle zona fasciculata is the origin of glucocorticoids and the innermost zona reticularis can synthesize androgens. The phenylalanine-derivative catecholamines epinephrine, norepinephrine and dopamine are made by cells of the adrenal medulla. The medulla is of neuroendocrine origin and the chromaffin cells that synthesize

catecholamines are actually modified postganglionic neurons. Fibers from preganglionic neurons of the autonomic nervous system whose cell bodies reside in the spinal cord reach directly to the chromaffin cells (Hammond, 2015).

The adrenal location in proximity to the abdominal aorta and arterial blood supply from three separate arteries ensures the access to large volumes of blood necessary for adrenal function to efficiently deliver hormones to distal organs. Arterioles branch to provide separate flows to the cortex and the medulla, which are capable of independent regulation. Medullary blood flow is known to increase when catecholamines are released. Cortical hormone release is not thought to require such rapid changes in blood flow (Breslow, 1992). However, because there are more arterioles feeding the cortex, the overall pattern of flow is from the outer gland in through the cortex and then to the medulla which is drained by the medullary vein (Vinson et al., 1985).

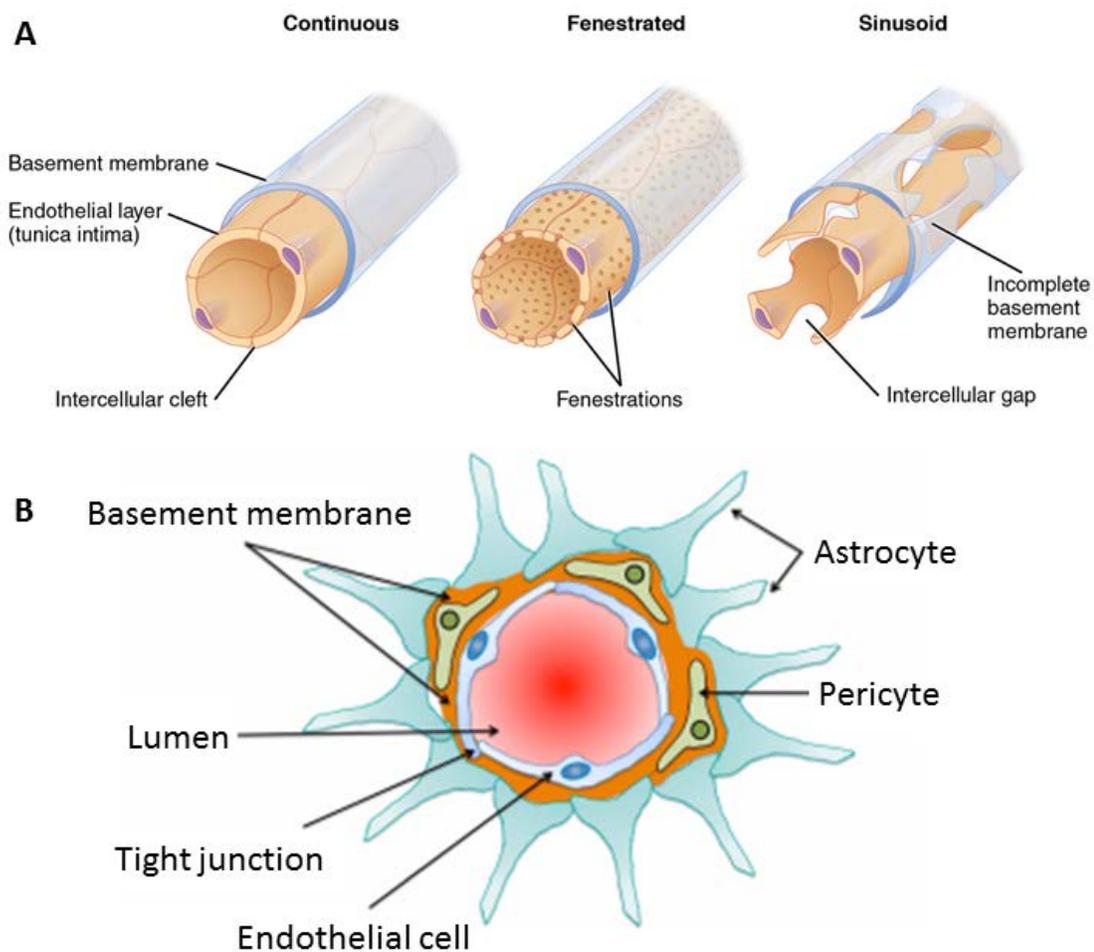
Although it is not a site frequently examined clinically, whole body scans routinely reveal adrenal metastases. Outgrowth here is not as lethal as the brain, liver or bone with indolent adrenal insufficiency rarely progressing to a full adrenal crisis (Seidenwurm et al., 1984).



**Figure I.6: Anatomy of the adrenal gland.** The adrenal gland consists of a capsule enclosing a cortex and an inner medulla. The cortex can be divided up into the three distinct zones shown here responsible for the production of the hormones listed at the right. The medulla is of neuroendocrine origin and preganglionic autonomic nervous system neuron processes terminate directly on the medullary chromaffin cells which produce the catecholamines epinephrine, norepinephrine and dopamine. Modified from The Adrenal Glands. OpenStax CNX. Jun 19, 2013 <http://cnx.org/contents/91a5f6cb-65d9-441e-b08d-eb2cb4c4cc91@3>

There exists a relative dearth of research on malignant adrenal colonization. Speculation as to why the gland is regularly colonized is not frequently addressed in clinical literature. It is assumed that the sinusoidal capillary structure of the adrenal gland vasculature facilitates invasion from the circulation (Lam and Lo, 2002). Indeed Figure I.7A illustrates large intercellular gaps and the lack of a basement membrane in sinusoidal capillaries assumedly lowering the difficulty of extravasation by a tumor cell in circulation. It has also been suggested that the local high concentration of corticosteroids may promote implantation of metastases (Kung et al., 1990; Williams 2011). However, this claim is

difficult to reconcile with the fact that corticosteroid treatment is frequently used in multiple cancers and can show clinical benefit in relapse free survival (McKay and Cidlowski, 2003). How and why cells seed the adrenal gland and adapt to growth there is still fertile scientific ground.



**Figure I.7: Capillary structure.** A) The three types of capillaries and their ultrastructural characteristics. Sinusoids (right) are found in the adrenal gland, liver, spleen and bone marrow. Fenestrated capillaries are present in intestines, pancreas and kidney. From Anatomy & Physiology. OpenStax CNX. Jul 30, 2014 <http://cnx.org/contents/14fb4ad7-39a1-4eee-ab6e-3ef2482e3e22@6.27> B) The special case of the blood-brain barrier which includes astrocyte foot processes coordinating with pericytes and the endothelium adapted from (Hawkes et al., 2014).

As chromatin structure can control cellular identity and has roles in carcinogenesis and malignant progression, the hypothesis that chromatin modification can drive

progression to metastasis was tested in this work. Specifically knock down of individual chromatin modifiers was examined in metastatic tropism using lentiviral hairpin libraries to probe metastatic tropism in a model of systemic colonization via intracardiac injection of tumor cells. While executing these experiments it was discovered that the tumor cell line used had an inherent tropism for the brain, ovary and adrenal gland. An *in vitro* screen of adrenal hormones demonstrated dopamine increased tumor cell viability. The significance of this finding was extended to reveal that breast cancer tumor cells may be able to adapt to the central nervous system by utilizing dopamine signaling to increase proliferation. Further in addition to responding to dopamine, tumor cells may even gain the ability to synthesize and store their own dopamine and participate in autocrine or paracrine dopamine signaling to acquire a metastatic advantage in the brain and adrenal glands.

# Chapter II:

## Materials and Methods

## **II.1 Cell lines**

MDA-MB-361 and MCF7 cells were purchased from ATCC and grown in DMEM supplemented with antibiotic/antimycotic and 10% FBS. SUM149, 159 and 225 cell lines were obtained from Dr. Stephen Ethier and are commercially available from Asterand. These cells were grown in F12 with 5ug/ml insulin, 1ug/ml hydrocortisone, antibiotic/antimycotic and 5% FBS. For signaling assays cells were starved overnight in phenol red free DMEM without supplementation. The 361-PLN line was generated by viral transduction with a lentiviral based luciferase construct (addgene.org #21471) using neomycin at 500ug/ml for eukaryotic selection. The 361-luc line was generated by viral transduction performed by Kai Tao using a luciferase expressing construct under neomycin selection.

## **II.2 Cell survival and proliferation assays**

300,000 361-luc cells were plated in 6 well plates in full media. For proliferation, three wells were counted at each time point using a Bio-Rad TC10 cell counter. For selection, puromycin was added at concentrations from zero to 3ug/ml in 0.5ug increments. Cells were counted at 24 and 48 hours. At 48 hours, the 1.5ug and higher concentrations had killed all of the non-transduced cells by 48 hours as confirmed by light microscopy.

## **II.3 Lentiviral libraries**

The pHAGE library was provided as pooled viral supernatants from the Steve Elledge lab at Harvard Medical School Department of Genetics. The pLKO library was provided in the pLKO vector by Rob Mathis from the Piyush Gupta lab at the Whitehead

Institute at MIT. All the lentiviruses used here were produced through transient transfection of HEK293T cells using TransIT-LT1 transfection reagent (Mirus Bio). Volume was scaled according to the manufacturer's parameters for a 6 well plate. 500,000 293T cells were plated overnight in 2.5 ml of 10% FBS DMEM and transfected the following day. 125ul of Opti-MEM serum free media was mixed with 125 ng of the pCMV-VSVG (addgene #8454), 1125 ng of pCMV-dR8.2-dvpr (addgene #8455) and 1250 ng of the lentiviral packaging construct. In a separate tube 125 ul of Opti-MEM was mixed with 7.5ul of TransIT-LT1. The DNA containing media was added dropwise to the transfection reagent tube, flicked gently and incubated for 15 minutes at room temperature before being gently dripped in a circular pattern onto the 293T cells. The plates were then moved immediately to an incubator specified for virus production and the media was changed approximately 16 hours later. Media containing virus was collected after an additional 24 and 72 hours and stored at 4C. Virus was concentrated by spinning out cellular debris and then transferring the chilled supernatant to a tube containing 1ml of sterile filtered 50% PEG-3350 (Sigma) per 4.5ml of supernatant. The tube was inverted 5 times to mix and then allowed to precipitate overnight at 4C. In this state the virus is stable for up to 2 weeks at 4C. The concentrated virus was pelleted at 2000 G for 20 minutes before being resuspended in whatever desired volume of sterile 1% BSA PBS with added pen/strep to be aliquoted and frozen at -80C.

#### **II.4 Flow cytometry**

Analyses were carried out using a CyAn ADP flow cytometer. Cells were trypsinized as briefly as possible to detach and resuspended to 500,000 cells per 300 of FACS buffer (1% FBS in PBS). When antibody staining was required, 3ul of antibody

was added for 10 minutes on ice. 500ul of FACS buffer was then added and the cells were pelleted for 3 minutes at 800g before being resuspended in 500 ul of FACS buffer, pelleted again and then transferred to FACS tubes in 300ul FACS buffer for reading on the machine. The antibodies used were FITC-CD49f (BD 55735) and APC-EpCAM (BD 347200). The stainings performed by Patty Keller are listed in the figure legend.

For measuring ROS via FACS negative control cells were treated in the plate for 1 hour with 1 mM N-acetyl cysteine before being treated along with the positive control cells with 250 uM tert-butyl hydro peroxide for 30 minutes to induce ROS. CellRox green reagent (Life Technologies C10492) was then added for 60 minutes at 500 nM in DMSO. Cells were trypsinized and washed three times with FACS buffer before addition of the included cell viability dye and FACS analysis.

## **II.5 Animal Experiments**

All animals were housed at the Division of Laboratory Animal Medicine at Tufts University according to IACUC standards on an approved animal protocol. Female NOD/SCID mice were purchased at 6 weeks of age from Jackson Laboratories and Charles River.

### **II.5.1 Orthotopic injections**

As detailed in the animal protocol, 1 million cells were injected in a total volume of 60 ul in 50% Matrigel into the 4<sup>th</sup> mammary fat pad. Mice were anesthetized and monitored with supplemental analgesic as required.

### **II.5.2 Intracardiac injections**

Mice were anesthetized with isoflurane and briefly held fixed onto a stage for injection. 1 million 361-luc cells were injected in 100 ul of PBS through a 26 gauge needle using a VEVO 200 ultrasound device to locate the needle into the left ventricle.

### **II.5.3 Whole animal imaging**

Animals were imaged at first weekly and later bi or tri weekly depending on signal saturation using a Perkin Elmer IVIS 200 biophotonic imager at the Tufts Small Animal Imaging Facility. 150 mg of luciferin was administered via intraperitoneal injection 5 minutes before imaging. A five minute exposure time was used until tumors were detectable and after that point the machine was set to auto expose and the longest exposure without saturation was determined on a per cage basis. Bioluminescence was measured in photons and total flux in photons per second using regions of interest defined by hand across the different time points using Perkin Elmer Living Image v4.4. Each cage was collected as a single image and each experiment was analyzed as a longitudinal series applying a common scale. A universal background region was selected that contained a mouse foot without any luminescent signal and this amount was subtracted from all measurements within an experiment.

### **II.6 Immunohistochemistry and Immunocytochemistry**

IHC was performed by the New England Medical Center Histology Laboratory on paraffin-embedded tissue sections on a Ventana automated slide stainer with the iVIEW DAB detection kit for visualization. Antibodies used were all from Ventana Medical Systems and prediluted for p53, ER, Her2, Ki67 and PR.

For ICC Cells were plated in 8 well chamber slides and fixed for 10 minutes at room temperature in 1% PFA before staining. Staining solution was 1% BSA in PBS with 0.01% tween-20. Antibodies were used 1:100 of Ki-67 (abcam ab15580) and DRD4 (EMD 324405) overnight at 4C with DAPI counterstain applied after washing. Five fields of view were counted per treatment. The percentage of positive cells was determined using DAPI positive nuclei as the denominator.

## **II.7 Migration and invasion assays**

8um porous transwell membranes were purchased uncoated or coated with growth factor reduced Matrigel (Corning 354483). 100,000 cells were plated in 400ul of DMEM and placed into wells with 700ul of DMEM 10% FBS. The wells were removed and the side where the cells were seeded was scrubbed with an applicator to remove the Matrigel and cells from the near side. The well were then rinsed once in PBS and then fixed for 2 minutes in 1% PFA and then stained for 10 minutes in 0.5% crystal violet. After staining each well was submerged briefly two times in 500 ml of water before drying membrane side up overnight. 3 fields of view were counted for each well taking the top center field, the middle and the bottom center field to ensure edge and middle cell distributions would be included. Invasion was calculated as the number of cells counted on the coated transwells divided by the number of cells on the uncoated transwells.

## **II.8 Tumor Spheres**

Two thousand 361-luc cells were plated in 1 ml of full media in ultra-low adhesion 24 well plates (Corning 3473) with 6 wells per hairpin. Cells were left undisturbed for 1 week before particle size was determined using a BD multisizer. Each

well was diluted into BD isoton II solution buffered with glycerol for three reads per well to ensure all particles were counted. Bins representing total counts per well and implementing cutoffs of 25 and 40 um were considered.

## **II.9 Adrenal hormone viability screen**

15,000 361-luc cells were plated on the inner 60 wells of a 96 well plate in 90ul of phenol red free DMEM. The outer wells were filled with 200ul of PBS. Each treatment group was made up of 6 wells and the dopamine and hydrocortisone validations were done in biologic triplicate. Dopamine (1155645) and epinephrine (1187741) were purchased from Henry Schein. Adrenotropic hormone (A6303), aldosterone (A9477), deoxycorticosterone (D7000), didehydroepiandrosterone (D4000), hydrocortisone (H0888), norepinephrine (A7257) were from Sigma. Each drug was diluted in PBS from light protected stocks fresh for treatment at 1uM and 10 nM. Cell viability was measured according to the manufacturer's protocol using the Promega MTS assay (G3580).

### **II.9.1 Dopamine response curve**

The dopamine response curves were determined using the same plating strategy as the viability screen. Cell titer glo reagent was added 1:1 for bioluminescent assay per the manufacturer's protocol (Promega G7570). A noncolorimetric method was used here to ensure no colorimetric interference was possible with the high concentrations of dopamine. Luminescence was measured with a Promega GloMax Multi plate reader.

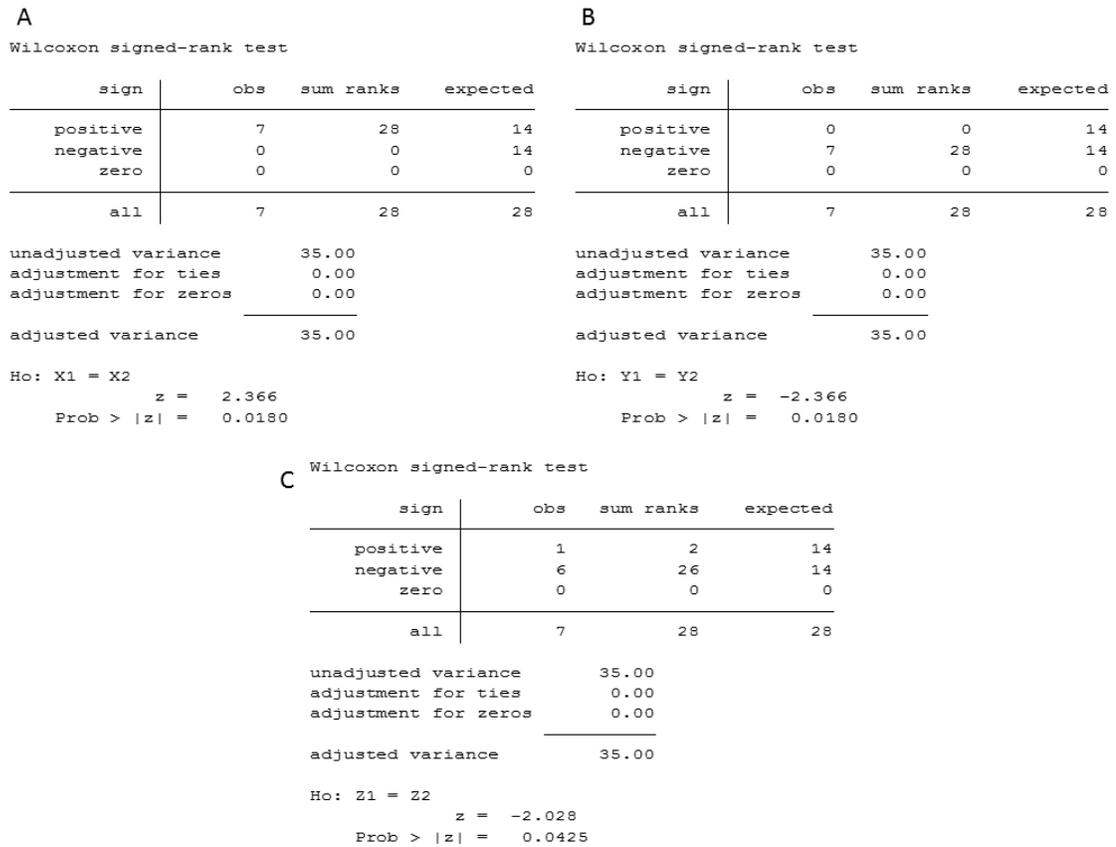
### **II.9.2 PKA activity**

PKA activity was measured with same plating and reader as above starving cells overnight in phenol red free DMEM before treating with 1uM and 10uM dopamine

beginning with the longest time course of four hours down to 0.5 hours with the entire plate quenched with the cAMP-glo lysis buffer and shaken for 15 minutes according to the manufacturer's protocol (Promega V1501). Relative luciferase activity was determined by dividing the relative luciferase units of each sample by that of the untreated cells. Statistical significance was determined using student's t-test assuming equal sample size and variance.

## **II.10 Dopamine quantal release**

361-luc cells were plated for 48 hours and treated with L-DOPA for 30 minutes. Carbon fiber amperometry was used to record catecholamine quantal size, peak amplitude and interspike interval. Analysis of variance was conducted to determine significance of the differences in the means.



**Figure II.1 Statistical analysis for Interspike interval. (A) Peak Amplitude (B) and Mean Quantal Size (C)**

## II.11 Western blots

After treatment and a PBS rinse, cells were lysed directly in the plate in RIPA buffer and frozen at -80C. Lysates were thawed and homogenized by pipette and vortex before protein concentration was measured using the Bio-Rad DC protein micro assay as instructed (5000112) using BSA as the protein standard. 20ug of total protein was loaded per lane and blots were detected using a 1:1000 primary antibody and 1:10,000 secondary with Pierce SuperSignal West Femto reagent (34095). Primary antibodies used were Cell Signaling Technology #2627 anti-SirT3, #2302 anti-DARPP-32, #2301 phospho-DARPP-32 and #5274 COX-IV HRP. Secondary detection used Pierce HRP conjugated

Donkey anti-Rabbit SA1-200. SIRT3 expression was quantified by pixel density per band on a Bio-Rad chemidoc XRS+ and was normalized to the in lane mitochondrial loading control COX IV signal cut from the same blot.

## **II.12 Hairpin recovery from tumors**

Samples were prepared from genomic DNA extracted from tumors using the Qiagen Dneasy blood and tissue kit paired with a TissueLyser LT using a 90 second pulverization with a single 7 mm stainless steel bead. Disposable 2 ml tubes were used with the twelve tube adapter to avoid cross-sample contamination. Hairpins were recovered from the gDNA by PCR with PrimeSTAR polymerase (Takara HSR010A) using 100ng gDNA initially denatured at 98C for 2 minutes. First round primers

pHAGEf3 ACTTTACAGAATCGTTGCCTGCAC

pHAGE-r GCCAGAGGCCCACTTGTGTAG

were used for the pHAGE library and

plko.bd.f2 ACAGGCCCGAAGGAATAGAAGAAG

plko.bd.r2 GCCAAAGTGGATCTCTGCTGTCC

were used for the pLKO primary amplicon. 15 cycles of 68C for 60 seconds followed by 98C for 10 seconds were performed and then the amplicon was purified away from the oligos and genomic DNA using paramagnetic beads (Axygen AXY-MAG-PCR-CL-50). 1/50<sup>th</sup> of this was used as starting material for the second amplification. For the second round the flow cell binding adapters and index codes were added to enable next generation sequencing and individual determination of each tumor. The following two

tables list the oligo plate well, the bar code name, the oligo sequence used for the forward primer, the tumor description, and the index for each library. An exhaustive list of six nucleotide bar codes with a hamming distance greater than three such that one base may be lost and the index is still likely readable is available from the Genomic Sequencing and Analysis Facility User Support Wiki at the University of Texas at Austin

<https://wikis.utexas.edu/display/GSAF/Illumina+--+all+flavors>

<b>Table II.I pHAGE PCR oligos, bar codes and tumor cross reference.</b>				
Oligo well	Oligo code	Oligo sequence	Sample description	Bar code
A01	BC_01	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	4 Heart	ATCACG
A02	BC_02	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	6 R adrenal	CGATGT
A03	BC_03	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	6 L adrenal	TTAGGC
A04	BC_04	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	7 R adrenal	TGACCA
A05	BC_05	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	7 brain	ACAGTG
A06	BC_06	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	8 R adrenal	GCCAAT
A07	BC_07	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	8 L adrenal	CAGATC
A08	BC_08	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	8 pituitary	ACTTGA
A09	BC_09	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	9 rib	GATCAG
A10	BC_10	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	40 R adrenal	TAGCTT
A11	BC_11	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	40 L adrenal	GGCTAC
A12	BC_12	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	40 L ovary	CTTGTA
B01	BC_13	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	44 R adrenal	AGTCAA
B02	BC_14	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	44 L adrenal	AGTTCC
B03	BC_15	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	44 brain	ATGTCA
B04	BC_16	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	45 R adrenal	CCGTCC
B05	BC_17	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	45 L adrenal	GTAGAG
B06	BC_18	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	45 pituitary	GTCCGC
B07	BC_19	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	45 brain	GTGAAA
B08	BC_20	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	46 R adrenal	GTGGCC

B09	BC_21	CAAGCAGAAGACGGCATAACGACGAAACTAGTGAAGCCACAGATGTA	46 L adrenal	GTTTCG
B10	BC_22	CAAGCAGAAGACGGCATAACGACGTACGTAGTGAAGCCACAGATGTA	46 brain TR	CGTACG
B11	BC_23	CAAGCAGAAGACGGCATAACGACCACTCTAGTGAAGCCACAGATGTA	46 brain TL	GAGTGG
B12	BC_24	CAAGCAGAAGACGGCATAACGAGCTACCTAGTGAAGCCACAGATGTA	46 pituitary	GGTAGC
C01	BC_25	CAAGCAGAAGACGGCATAACGAATCAGTTAGTGAAGCCACAGATGTA	46 inferior to pituitary	ACTGAT
C02	BC_26	CAAGCAGAAGACGGCATAACGAGCTCATTAGTGAAGCCACAGATGTA	46 pit bone	ATGAGC
C03	BC_27	CAAGCAGAAGACGGCATAACGAAGGAATTAGTGAAGCCACAGATGTA	46 lung	ATTCTT
C04	BC_28	CAAGCAGAAGACGGCATAACGACTTTTGTAGTGAAGCCACAGATGTA	47 R adrenal	CAAAAG
C05	BC_29	CAAGCAGAAGACGGCATAACGATAGTTGTAGTGAAGCCACAGATGTA	47 L adrenal	CAACTA
C06	BC_30	CAAGCAGAAGACGGCATAACGACCGGTGTAGTGAAGCCACAGATGTA	47 pituitary	CACCGG
C07	BC_31	CAAGCAGAAGACGGCATAACGAATCGTGTAGTGAAGCCACAGATGTA	50 R adrenal	CACGAT
C08	BC_32	CAAGCAGAAGACGGCATAACGATGAGTGTAGTGAAGCCACAGATGTA	50 L adrenal	CACTCA
C09	BC_33	CAAGCAGAAGACGGCATAACGACGCTGTAGTGAAGCCACAGATGTA	50 between kidneys	CAGGCG
C10	BC_34	CAAGCAGAAGACGGCATAACGAGCCATGTAGTGAAGCCACAGATGTA	50 R brain	CATGGC
C11	BC_35	CAAGCAGAAGACGGCATAACGAAAAATGTAGTGAAGCCACAGATGTA	50 F brain	CATTTT
C12	BC_36	CAAGCAGAAGACGGCATAACGATGTTGGTAGTGAAGCCACAGATGTA	4 L adrenal no luc	CCAACA
D01	BC_37	CAAGCAGAAGACGGCATAACGAATCCGTAGTGAAGCCACAGATGTA	4 R adrenal no luc	CGGAAT
D02	BC_38	CAAGCAGAAGACGGCATAACGAAGCTAGTGTAGTGAAGCCACAGATGTA	40 brain no luc	CTAGCT
D03	BC_39	CAAGCAGAAGACGGCATAACGAGTATAGTGTAGTGAAGCCACAGATGTA	40 normal ovary no luc	CTATAC
D04	BC_40	CAAGCAGAAGACGGCATAACGATCTGAGTAGTGAAGCCACAGATGTA	pre-screen pellet	CTCAGA
D05	BC_41	CAAGCAGAAGACGGCATAACGAGTCGTCTAGTGAAGCCACAGATGTA	90 L lung	GACGAC
E01	BC_49	CAAGCAGAAGACGGCATAACGATGCCGATAGTGAAGCCACAGATGTA	90 R lung	TCGGCA
D07	BC_43	CAAGCAGAAGACGGCATAACGACGATTATAGTGAAGCCACAGATGTA	94 heart T1	TAATCG
D08	BC_44	CAAGCAGAAGACGGCATAACGAGCTGTATAGTGAAGCCACAGATGTA	94 heart T2	TACAGC
D09	BC_45	CAAGCAGAAGACGGCATAACGAATTATATAGTGAAGCCACAGATGTA	95 brain	TATAAT
D10	BC_46	CAAGCAGAAGACGGCATAACGAGAATGATAGTGAAGCCACAGATGTA	95 R adrenal	TCATTC
D11	BC_47	CAAGCAGAAGACGGCATAACGATCGGGATAGTGAAGCCACAGATGTA	95 L adrenal	TCCCGA
D12	BC_48	CAAGCAGAAGACGGCATAACGACTTCGATAGTGAAGCCACAGATGTA	NOD/SCID gDNA	TCGAAG

<b>Table II.II pLKO PCR oligos, bar codes and tumor cross reference.</b>				
Oligo well	Oligo code	Oligo sequence	Sample description	Bar code
A01	BC_01	CAAGCAGAAGACGGCATAACGACGTGATTAGTGAAGCCACAGATGTA	5 BR T2	ATCACG
A02	BC_02	CAAGCAGAAGACGGCATAACGACATCGTAGTGAAGCCACAGATGTA	9 BR T	CGATGT
A03	BC_03	CAAGCAGAAGACGGCATAACGACCTAATAGTGAAGCCACAGATGTA	40 BR F	TTAGGC
A04	BC_04	CAAGCAGAAGACGGCATAACGATGGTCATAGTGAAGCCACAGATGTA	46 BR R	TGACCA
A05	BC_05	CAAGCAGAAGACGGCATAACGACTGTTAGTGAAGCCACAGATGTA	48 Eye	ACAGTG
A06	BC_06	CAAGCAGAAGACGGCATAACGAATTGGCTAGTGAAGCCACAGATGTA	7 peri	GCCAAT
A07	BC_07	CAAGCAGAAGACGGCATAACGAGATCTGTAGTGAAGCCACAGATGTA	49 chest 2	CAGATC
A08	BC_08	CAAGCAGAAGACGGCATAACGATCAAGTTAGTGAAGCCACAGATGTA	8 BR T	ACTTGA
B01	BC_13	CAAGCAGAAGACGGCATAACGATTGACTTAGTGAAGCCACAGATGTA	44 BR T	AGTCAA
B02	BC_14	CAAGCAGAAGACGGCATAACGAGGAAGTTAGTGAAGCCACAGATGTA	7 rib	AGTTCC
B03	BC_15	CAAGCAGAAGACGGCATAACGATGACATTAGTGAAGCCACAGATGTA	40 BR mid	ATGTCA
B04	BC_16	CAAGCAGAAGACGGCATAACGAGGACGGTAGTGAAGCCACAGATGTA	40 BR pit	CCGTCC
B05	BC_17	CAAGCAGAAGACGGCATAACGACTCTACTAGTGAAGCCACAGATGTA	47 BR 1	GTAGAG
B06	BC_18	CAAGCAGAAGACGGCATAACGAGCGGACTAGTGAAGCCACAGATGTA	8 OV R	GTCCGC
B07	BC_19	CAAGCAGAAGACGGCATAACGATTTCACTAGTGAAGCCACAGATGTA	49 chest 1	GTGAAA
B08	BC_20	CAAGCAGAAGACGGCATAACGAGGCCACTAGTGAAGCCACAGATGTA	47 BR 2	GTGGCC
C01	BC_25	CAAGCAGAAGACGGCATAACGAATCAGTTAGTGAAGCCACAGATGTA	49 chest 3	ACTGAT
C02	BC_26	CAAGCAGAAGACGGCATAACGAGCTCATTAGTGAAGCCACAGATGTA	6 BR T1	ATGAGC
C03	BC_27	CAAGCAGAAGACGGCATAACGAAGGAATTAGTGAAGCCACAGATGTA	N	ATTCTT
C04	BC_28	CAAGCAGAAGACGGCATAACGACTTTTGTAGTGAAGCCACAGATGTA	47L ADR	CAAAAG
C05	BC_29	CAAGCAGAAGACGGCATAACGATAGTTGTAGTGAAGCCACAGATGTA	47R ADR	CAACTA
C06	BC_30	CAAGCAGAAGACGGCATAACGACCGGTGTAGTGAAGCCACAGATGTA	48L ADR	CACCGG
C07	BC_31	CAAGCAGAAGACGGCATAACGAATCGTGTAGTGAAGCCACAGATGTA	48R ADR	CACGAT
C08	BC_32	CAAGCAGAAGACGGCATAACGATGAGTGTAGTGAAGCCACAGATGTA	67R ADR	CACTCA
D01	BC_37	CAAGCAGAAGACGGCATAACGAATCCGTAGTGAAGCCACAGATGTA	OR ADR	CGGAAT

D02	BC_38	CAAGCAGAAGACGGCATAACGAAGCTAGTAGTGAAGCCACAGATGTA	5L ADR	CTAGCT
D03	BC_39	CAAGCAGAAGACGGCATAACGAGTATAGTAGTGAAGCCACAGATGTA	5R ADR	CTATAC
D04	BC_40	CAAGCAGAAGACGGCATAACGATCTGAGTAGTGAAGCCACAGATGTA	6L ADR	CTCAGA
D05	BC_41	CAAGCAGAAGACGGCATAACGAGTCGTCTAGTGAAGCCACAGATGTA	6R ADR	GACGAC
D06	BC_42	CAAGCAGAAGACGGCATAACGATAGCGCTAGTGAAGCCACAGATGTA	7L ADR	GCGCTA
D07	BC_43	CAAGCAGAAGACGGCATAACGACGATTATAGTGAAGCCACAGATGTA	8L ADR	TAATCG
D08	BC_44	CAAGCAGAAGACGGCATAACGAGCTGTATAGTGAAGCCACAGATGTA	8R ADR	TACAGC
E01	BC_49	CAAGCAGAAGACGGCATAACGATGCCGATAGTGAAGCCACAGATGTA	9L ADR	TCGGCA
E02	BC_50	CAAGCAGAAGACGGCATAACGAGTGTITTTAGTGAAGCCACAGATGTA	9R ADR	AAACAC
E03	BC_51	CAAGCAGAAGACGGCATAACGACCTTCATAGTGAAGCCACAGATGTA	40L ADR	TGAAGG
E04	BC_52	CAAGCAGAAGACGGCATAACGATATGTTTATAGTGAAGCCACAGATGTA	40R ADR	AACATA
E05	BC_53	CAAGCAGAAGACGGCATAACGAGACGCGTAGTGAAGCCACAGATGTA	44L ADR	CGCGTC
E06	BC_54	CAAGCAGAAGACGGCATAACGATGTATCTAGTGAAGCCACAGATGTA	44R ADR	GATACA
E07	BC_55	CAAGCAGAAGACGGCATAACGACACCTAGTGAAGCCACAGATGTA	45R ADR	GGTGTG
E08	BC_56	CAAGCAGAAGACGGCATAACGATTCTTATAGTGAAGCCACAGATGTA	46L ADR	TAAGAA
A12	BC_12	CAAGCAGAAGACGGCATAACGATACAAGTAGTGAAGCCACAGATGTA	pre	CTTGTA

The reverse primer for the secondary pHAGE amplicon was p5PHAGE-miR#5

AATGATACGGCGACCACCGAGGAGATCACCCCTGAAAATACAAAT and for the pLKO it was p5pLKO-rob

AATGATACGGCGACCACCGATTCTTTCCCTGCACTGTAC

The secondary amplicons were again purified with paramagnetic beads and the concentration was determined by nanodrop. 44ng of each pHAGE amplicon and 23ng of each pLKO amplicon were pooled separately and diluted to provide a 20ul sample at 20nM for the Tufts sequencing core. Each sample was run in one lane on an Illumina Hi-Seq 2500 using the sequencing primers I designed

pLKO-robSeq CGAGGTCGAGAATTCCGAGGCAGTAGGCA

pHAGE-trunc-seq GTAGCCCCTTGAATTCCGAGGCAGTAGGCA

## II.13 Bioinformatics

The data from the sequencing runs were received as compressed FASTQ files. I wrote a bash shell script to use the FASTX clipper ([http://hannonlab.cshl.edu/fastx\\_toolkit/](http://hannonlab.cshl.edu/fastx_toolkit/) FASTX-Toolkit Hannon lab, Cold Springs Harbor Laboratory) to trim each read to the appropriate size and select only reads with a fully intact ATACATCTGTGGCTTCACTA miR-30 hairpin loop then the FASTX barcode splitter was used to break the reads out into groups based on their bar code. These bar code files were then aligned to the reference hairpin databases for each library using Bowtie2 (Langmead et al., 2009) to append the hairpin gene target before sorting on the hairpin and counting each file after stripping the metadata. The scripting can be executed like so in the parent directory with the command line argument \$1 being the compressed fastq file name

```
gunzip -c $1 | fastx_clipper -a ATACATCTGTGGCTTCACTA -d 6 -c -l 44 \  
-v -Q33 | fastx_barcode_splitter.pl -bcfile barcodes -eol \  
--mismatches 0 -prefix bc/pg_ --suffix ".fq"
```

And then in the bc subdirectory:

```
# -c keeps only clipped lines
# -l 18 drops all reads less than 18 bases. This will eliminate
#     hairpins shorter than 18
# -I input file from command line
# -o output file from command line
# -Q33 sets the ASCII offset for the quality scores in the fastq file
# -v verbose output
# fastx_clipper -a ATACATCTGTGGCTTCACTA -c -l 18 -v -I $1 -o $2 -Q33
#
# -x index
# -S SAM output filename
# -U input files (comma separated)
# -N number of mismatches allowed
# -G race are you actually reading this
# -T his is the best part of my thesis
# --local allows soft matches by trimming the ends of the reads
# --un-gz write out a gzipped file with the unaligned reads
# --al-gz write out a gzipped file with the aligned reads
# bowtie2 -x bt2/hairpins -U $1 -S $2 -N 2 -local
#
# -f 3 grabs the 3rd column only
# cut -f 3 $1 > $2
#
# -v inverted grep, we are excluding the lines with * or : in them
# grep -v -E '[*]|[:]' $1 > $2
#
# sort the lines first then count the unique lines
# sort $1 | uniq -c > $2
#
#
# Pipe them all together:

for I in *.fq
do
  f="${i%.*}"
  # f is the filename without an extension
  echo $f
  fastx_clipper -a ATACATCTGTGGCTTCACTA -c -l 18 -I $i \
    -o $f.clip.fq -Q33 -v
  bowtie2 -x bt2/hairpins -U $f.clip.fq -S $f.sam \
    --un-gz .bt2.un.fq.gz -al-gz $f.bt2.al.fq.gz
  cut -f 3 $f.sam | grep -v -E '[*]|[:]' | sort | uniq -c \
    | sort -rn > $f.uniq.txt
  echo ---
done
```

The result was one file per tumor sorted by the number of reads for each gene target. Post-processing was done by hand using TextPad and Excel to divide tumors into adrenal and non-adrenal groups per library.

The hairpin reference database tables are thousands of lines long and available online at <https://goo.gl/6tf8QQ>

## II.14 qPCR

cDNA was synthesized using the Bio-Rad iScript cDNA synthesis kit (1708891) and detection was performed in a Bio-Rad CFX96 PCR machine using Bioline's SensiFAST SYBR No-ROX Kit (BIO-98002) the listed oligos. Each assay was performed in three wells and repeated in three biologic replicates. Statistical significance was measured using Student's t-test with equal sample sizes and equal variance comparing Ct values. Samples were normalized to GAPDH.

<b>Table II.III qPCR primers.</b>		
Gene	Primer1	Primer2
DRD1	AGAAGTCCCTCTCCACCACC	TTTCTGGTGCCCAAGACAGT
DRD2	ATTCAGTGGATCCATCAGGG	CTGGAAGCCTCAAGCAGC
DRD3	GTGCCATGCTGGTAGTGAAC	ACCCTGGATGTCATGATGTG
DRD4	GAAGATGGAGGCGGTGC	GCCGCTCTTCGTCTACTCC
DRD5	CACACACACACAACCTGCTG	TATGCTCTCCCCTCCCTTTT
PAH	TCCAAATGGGTGAAAAATTCAT	GGCAAAGTATTGCGCTTAT
TH	CACGAAGTACTCCAGGTGGG	CGAGCTGTGAAGGTGTTTGA
DDC	ATTGTCAAAGGAGCAGCATGT	AGGAAGCCCTGGAGAGAGAC
SLC6A3	CAGGCAGGCTGTGAGCTG	CACCTGCTGCCGAGTACTTT
SLC29A4	CCGTGTACCCGTAGAAGCTG	GTGGCTGCAGCTCTTCTCTC
SLC6A2	GACGACCATCAGACAGAGCA	CGGCAGCCGAGTTTTATG
SLC6A4	TCTGTTGGTGTCTTCTGGGGT	CATCTTTTCAATTCATCCCC
DBH	TTCTCCCAGTCAGGTGTGTG	GCCTTCATCCTCACTGGCTA
SIRT3	AACACAATGTCGGGCTTCAC	CAGTCTGCCAAAGACCCTTC
GAPDH	CCATGGGGAAGGTGAAGGTC	TAAAAGCAGCCCTGGTGACC

Chapter III:  
Establishment of A  
Novel *in vivo* Breast  
Cancer Metastasis Model

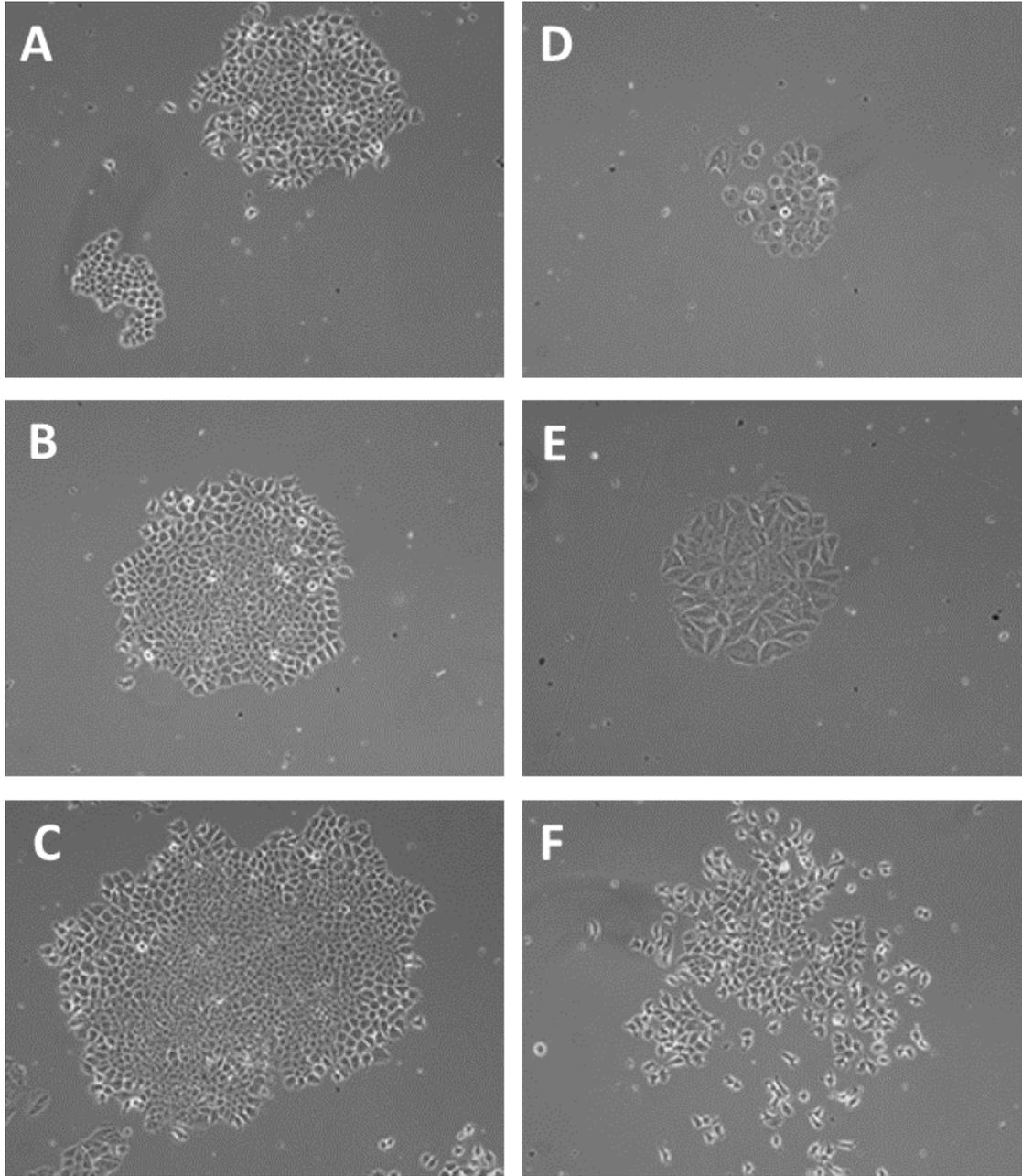
## **Abstract**

Metastasis is a difficult process to model due to its progression across the three dimensions of local tissue boundaries, distant organs and time. Single primary tumors have a limited circulatory route determined by lymphatic drainage and access to local arteries. But many solid tumors show patterns of specific organ growth beyond this circulation. Thus tumor cells must play an active role and there are host organ characteristics making some organs more amenable secondary sites. To address these characteristics I used a breast tumor line derived from a brain metastasis paired with ultrasound guided intracardiac injection to track colonization in vivo and gauge its full metastatic potential. Tumors grew out in many organs, primarily the adrenal glands, brain and ovary. The adrenal gland displayed the highest penetrance and I conducted an *in vitro* screen of adrenal hormones on cell viability which revealed that dopamine provides a proliferative advantage to 361-luc cells. Here I present evidence this response involves dopamine receptors, signal transduction, biosynthesis, storage and functional electrophysiology. As there are many dopamine related drugs available, these results may provide a pharmacologically targetable process important in metastasis.

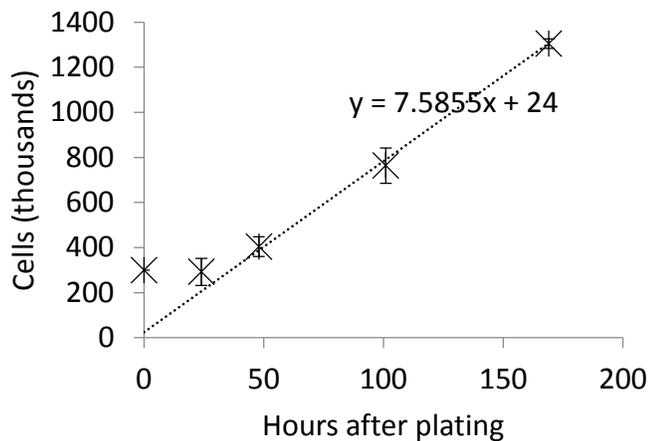
### **III.1 The MDA-MB-361 cell line**

The cell line used for the metastatic modeling and the chromatin screen in my work was established at the MD Anderson Cancer Center in 1975 from a brain metastasis recurring after surgical removal of the primary breast cancer from a 40 year old Caucasian female (Cailleau et al., 1978). The original histological examination from tumors arising after transplant into nude mice confirmed a breast epithelial origin for the line and karyotyping showed a modal chromosome count of 52. Our lab's source for the line was the American Type Culture Collection and their current characterization of the line now lists the modal number at 56 with a range of 54 to 61 (HTB-27 atcc.org 2015). As expected there has been some genetic drift in the line.

Figure III.1 shows the cuboidal cobblestone like colonies common to epithelial cells grown on plastic exhibited by the line with irregular shaped colonies showing variation in cell density. The line is not strongly adherent and a small percentage of floating cells are always present which when isolated and cultured alone resulted in colonies that were both tightly and more loosely packed as well as some colonies made exclusively of uncharacteristically large cells. The line grows relatively slowly for a tumor line with a measured doubling time of 74 hours in Dulbecco's modified Eagle's medium supplemented with 10% FBS (Figure III.2). The line is of the Luminal B molecular subtype and stably expresses both the estrogen and progesterone receptors and displays amplification of the Her2 human epidermal growth factor receptor 2 (Keller et al., 2010). Notable oncogenic mutations include a truncating mutation of the p53 tumor suppressor and an activating E545K mutation of the lipid receptor tyrosine kinase PIK3CA.



**Figure III.1 361-luc cell morphology.** 361-luc cells are loosely adherent cells with a typical epithelial cuboidal shape. During culture a small percentage of cells are found detached floating in the media. A-C) Typical colonies of the 361-luc line increasing in size and colony compaction. D-F) Atypical colony shapes after replating the floating population.

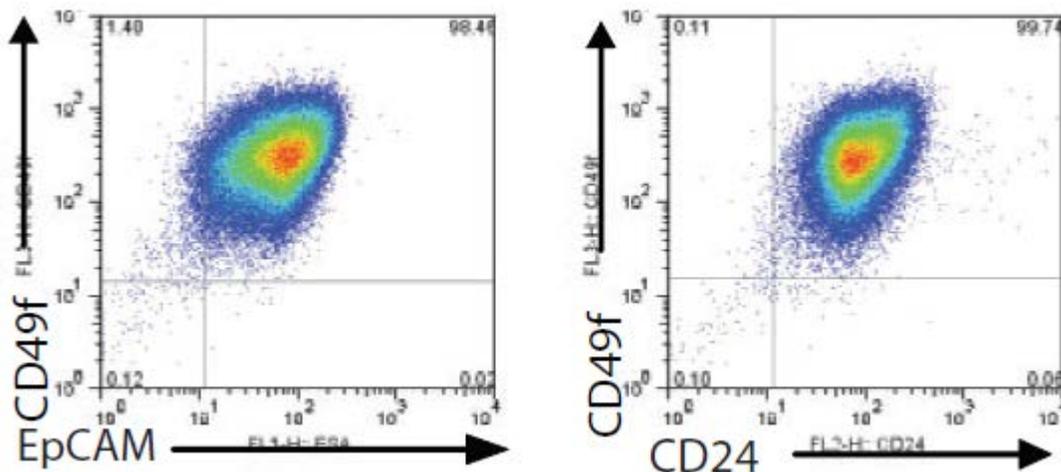


**Figure III.2 Proliferation and doubling time of 361-luc cells in full media.**

These two genes are the most frequently mutated genes in breast cancer and the top two mutations in luminal B tumors appearing 29% of the time (Network, 2012).

Flow cytometry studies conducted by Dr.

Patricia Keller on the MDA-MB-361 line showed that it stains strongly for the luminal markers EpCAM and CD24 (Figure III.3). Mature luminal human epithelial cells show low expression of CD49f, so this line could be described as a less differentiated, more progenitor-like line. Basal and pluripotent mammary epithelial cells express low levels of CD24 and EpCAM so this line does not represent these cells.



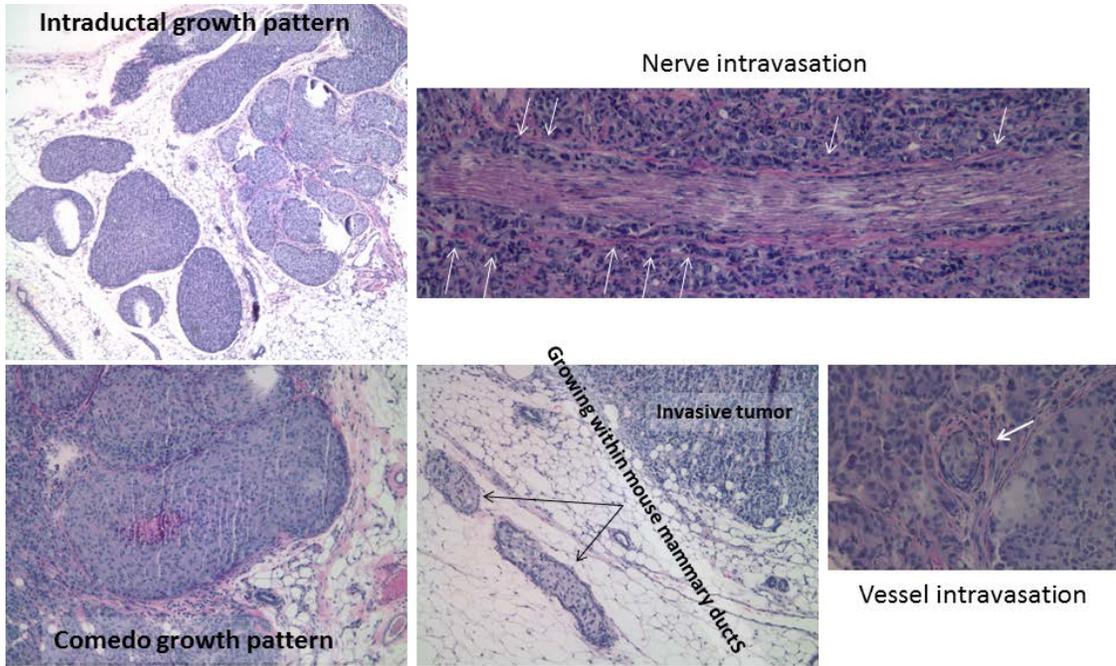
**Figure III.3 361-luc FACS staining of mammary progenitor markers.** Like normal luminal human mammary epithelial cells, 361-luc express EpCAM and CD24. However, co-staining of EpCAM/CD49f and CD24/CD49f marks them to be a luminal progenitor-like line as opposed to the differentiated mature luminal cells which express low levels of CD49f. Adapted from (Keller et al., 2010)

### **III.2 *In vivo* characterization using orthotopic injection of 361-luc**

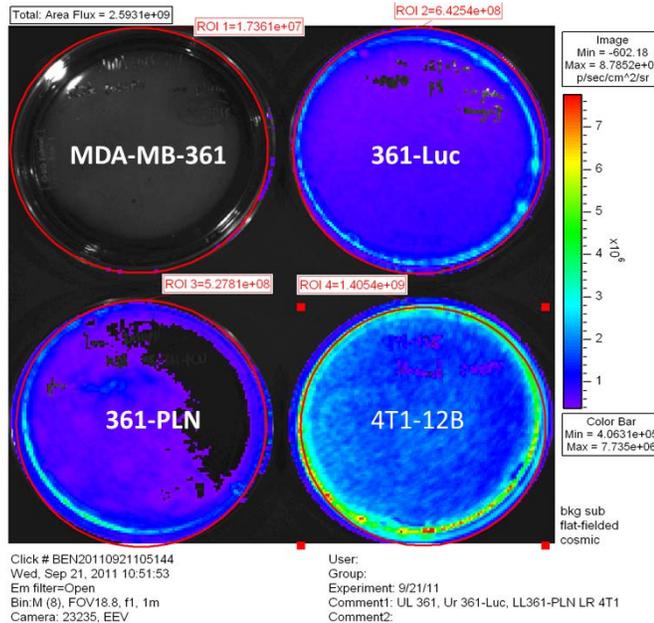
Historically, MDA-MB-361 cells have been investigated in the literature for metastatic behavior via orthotopic fat pad injection, tail vein seeding of the lungs and intra-carotid artery delivery to the brain (Zhang et al., 1991). The line was found to be poorly metastatic compared to three other tumor lines. It was however noted to perform better at colonizing the brain than the other two breast cancer lines tested, but still less so than a melanoma tumor line. This 1991 work is still the only major reference of MDA-MB-361 in regard to metastatic potential in pubmed although its sister-line the basal-like MDA-MB-231 which was also tested in this paper has gone on to great pubmed fame and glory as described earlier. In our lab, histopathologic analysis by Dr. Stephen Naber and Dr. Charlotte Kuperwasser of orthotopic injections performed by Ina Klebba demonstrated that orthotopic tumors do display metastatic characteristics including invasion into the local stroma, tumor cell presence in lymphatic and blood vessels, intraductal growth and growth along nerve bundles (Figure III.4).

All of the above described analyses were done through *ex vivo* histologic examination of suspected target organs. If there were metastatic colonies beyond the organs examined directly one would never know. In order to perform a systemic characterization of the line's metastatic propensity, MDA-MB-361 was transduced with a retrovirus encoding luciferase by Dr. Kai Tao in our lab to yield the 361-luc line. The transduction and selection were done *en masse* to maintain the heterogeneity inherent in the cell line. I performed my own luciferase transduction of MDA-MB-361 with a PGK promoter driven lentiviral luciferase to compare the luminescent activity as seen in Figure III.5 and found

that of my 361-PLN line to be slightly lower than 361-luc *in vitro*. Consequently the 361-luc line was used for all *in vivo* experimentation.



**Figure III.4** Orthotopic growth of 361-luc tumors in the mammary fat pad of immunocompromised NOD/SCID mice. Unpublished work from Dr. Charlotte Kuperwasser and Ina Klebba.

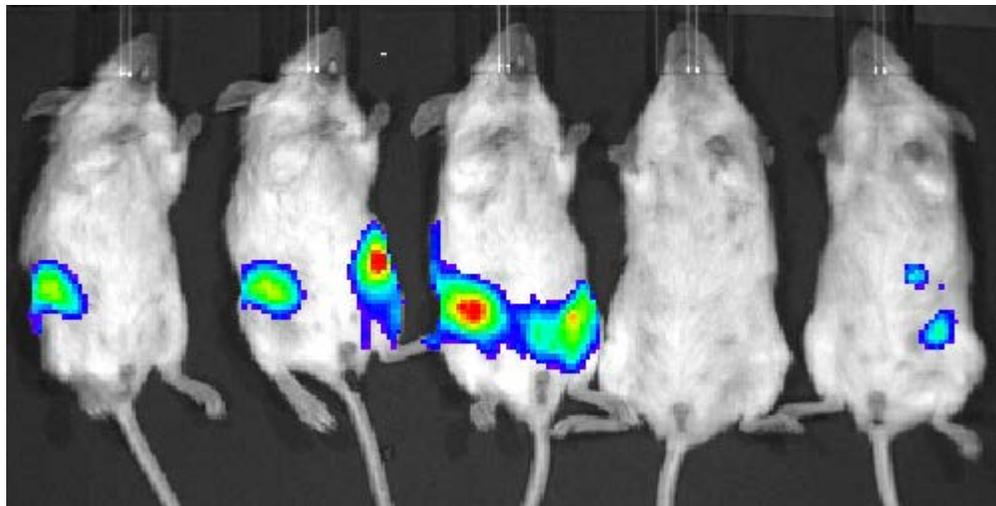


**Figure III.5** Bioluminescence of the 361-luc and 361-PLN lines. The 4T1-12B line was known to be detectable *in vivo* and served as a positive control.

The line was thus characterized as a relatively homogenous line based on FACS markers that maintained a small loosely adherent population which could exhibit heterogeneous colony morphology. Orthotopic tumors showed signs of local invasion with an interesting pattern of growth along nerve fibers. With the line adapted for *in vivo* imaging I moved on to recapitulate the original findings and expand them to the systemic context.

### III.2.1 Histopathology of orthotopic 361-luc tumors

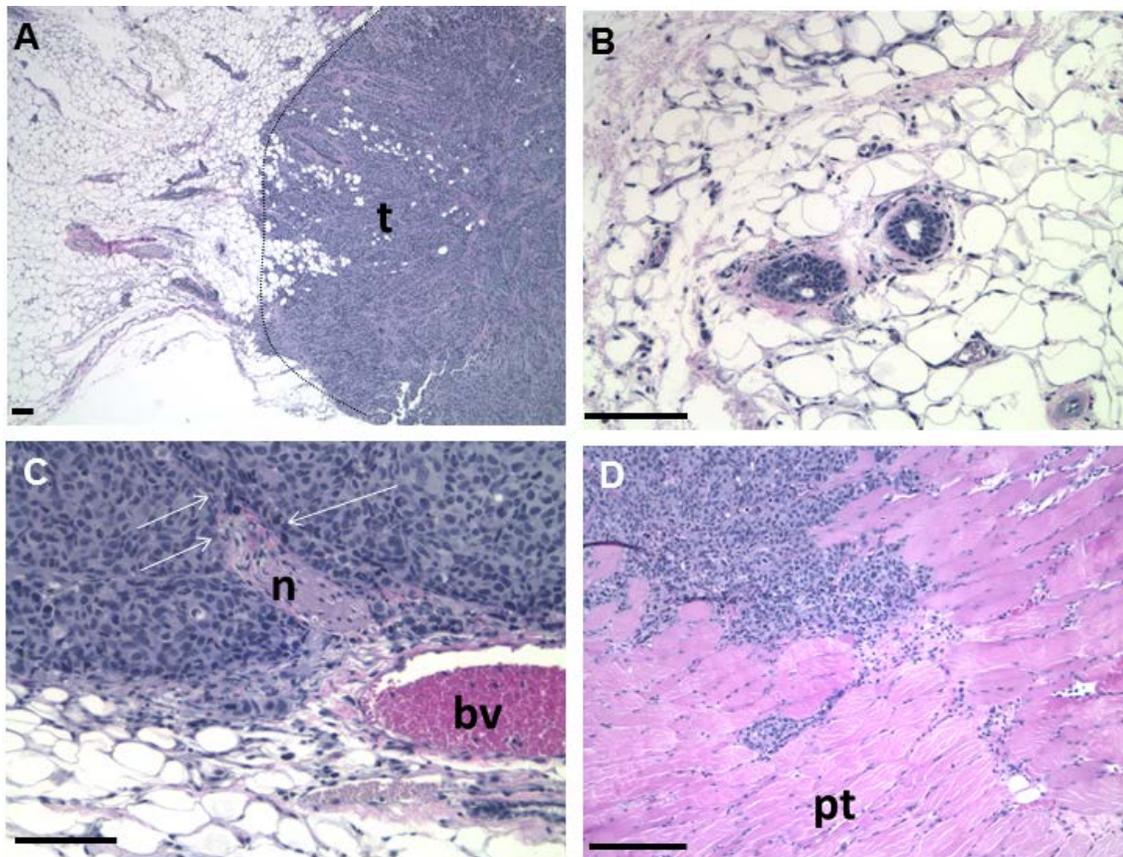
When the fourth mammary fat pads of NOD/SCID mice were injected with one million 361-luc cells harboring a control scramble hairpin, in 50% Matrigel, stable bioluminescent signal was measurable. Animals tolerated the growths well displaying normal behavior and showing no signs of distress.



**Figure III.6 Bioluminescent imaging of orthotopic 361-luc tumors.** The three mice on the left were injected with cells in 50% Matrigel. The two on the right received only cells.

Four months after the injections no distant metastases were detectable and the tumors were fixed and sectioned for histologic examination. Hematoxylin and eosin (H&E) staining (Fig. III.7) revealed tumors with long stretches of pushing border invading into the

mammary stroma. At some points overt invasion of the stroma and the peritoneum were observed with a more jagged-edged tumor intrusion. Growths were found inside the lumen of glands similar to the early stage ductal carcinoma *in situ* in humans. Tumor cells were also found inside both lymphatic and blood vessels. Tumor cells could be seen growing along and in close proximity to nerves and their sheaths. The tumors showed high vascularity and lobules could be clearly defined in the tumor ultrastructure but did not show distinct glandular structures aside from the tubular growth observed in ductal lumens.

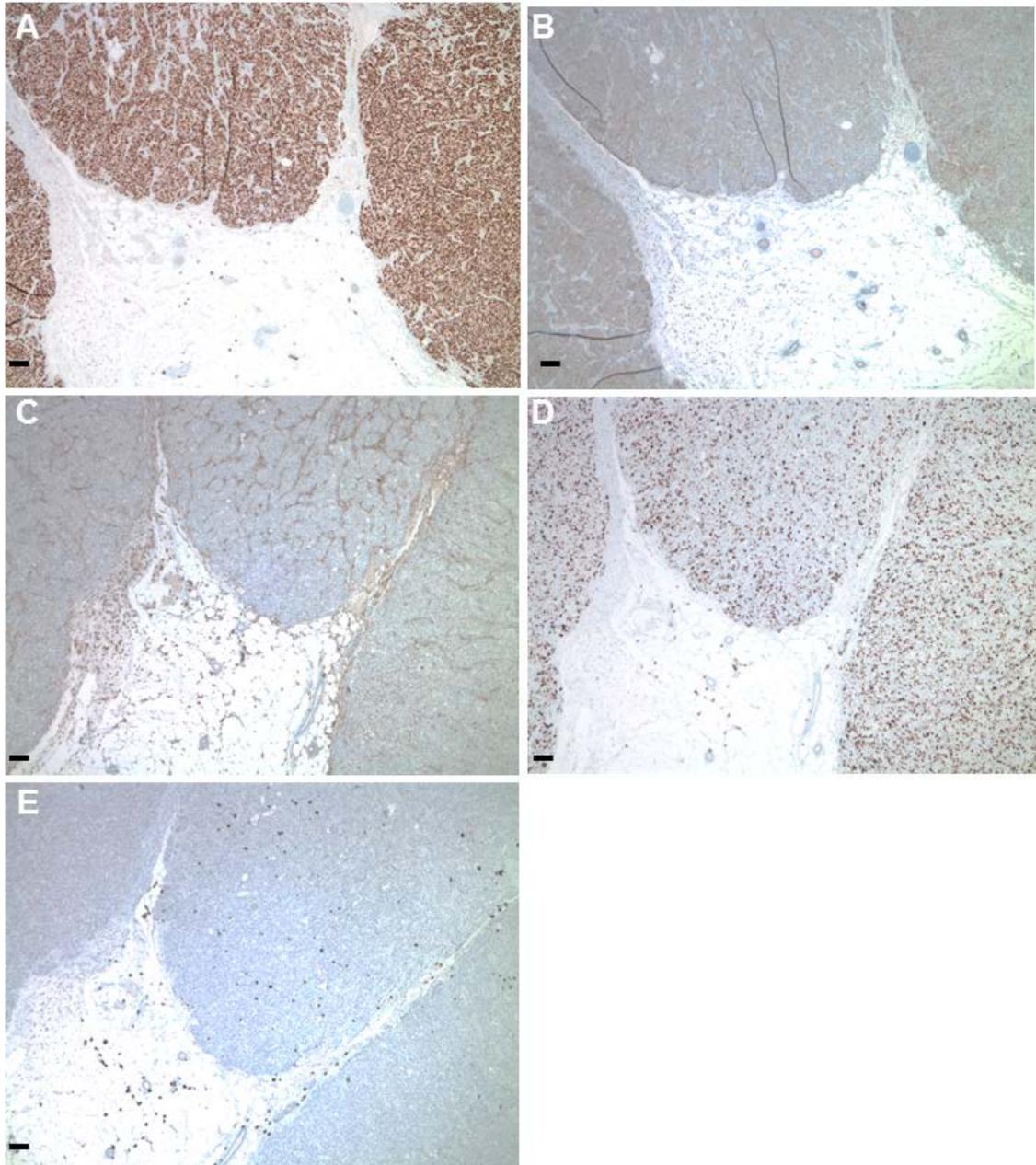


**Figure III.7 Histology of 361-luc mammary fat pad tumors.** A) Low magnification showing tumor (t) invading into stroma. B) Tumor cells growing inside the ductal lumen. C) Growth of tumor cells (white arrows) into the sheath of a cross section of nerve (n) and in proximity to a blood vessel (bv). D) Invasion into the peritoneum (pt). Bar = 100um

Grossly most of the tumors were non-cystic oblong spheroids of white or pinkish color with a roughened surface of tiny knobby dimples assumedly from the lobular growth pattern. They were firm and appeared to maintain a higher internal hydrostatic pressure as exudate quickly seeped out when tumors were bifurcated.

### **III.2.2 Immunohistochemical analysis of orthotopic 361-luc tumors**

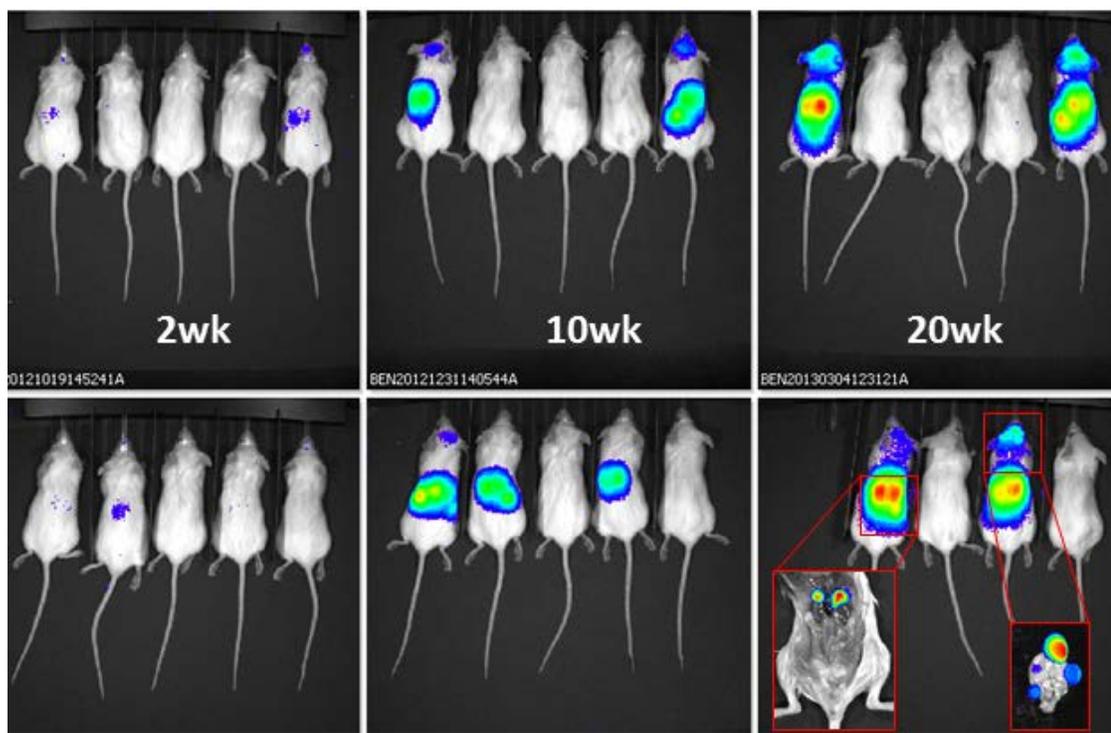
Immunohistochemical analysis (Fig. III.8) was performed on these tumors to see if they maintained expression of ER and Her2 as in culture. The proliferative marker Ki67 was also included as was the tumor suppressor p53. The mammary growths were strongly positive for ER indicating persistent estrogen signaling. Her2 amplification was also evident along the surface of the tumor cells. The tumors clearly exceeded the clinicopathological threshold of 20% Ki67 positivity for classification as highly proliferative. Only diffuse punctate p53 staining of non-tumor cells, most likely penetrating immune cells, were observed confirming the lack of p53 protein expression in the tumor cells.



**Figure III.8 Immunohistochemical staining of orthotopic 361-luc tumors.** All sections were from the same mammary tumor. A) Strong nuclear ER staining throughout the entire tumor. B) Distinct Her2 staining was localized to the tumor cell membrane. C) Progesterone receptor staining. Staining here is along the external cell interface. As PR is a nuclear antigen this staining is most likely nonspecific. D) Ki67 nuclear staining showing a high rate of proliferation in the tumor cells. E) The tumor cells do not express p53.

### III.3 Systemic organ seeding of 361-luc via intracardiac injection

To build upon my findings from the orthotopic injections, I proceeded to test if there were other organs that 361-luc could colonize if delivered directly rather than through spontaneous metastasis from the mammary fat pad. To this end I injected one million 361-luc cells into the left ventricle of ten NOD/SCID mice. As Figure III.9 shows, by two weeks luciferase signal could be detected in the dorsal abdominal region in two mice with a third mouse showing a similar signal as well as one coming from the skull. By twenty weeks post injection half of the mice had signal in multiple organs. Again, the mice tolerated the tumors well with no obvious mass detectable by palpation. They exhibited normal behavior and maintained a healthy mass over time with the exception of one animal that experienced a rapid weight loss followed by laborious respiration and was found, upon sacrifice, to have a thymic lymphoma which is common in the NOD/SCID genetic background.



**Figure III.9** 361-luc tumors arising from intracardiac injection over time. Insets show *ex vivo* imaging identifying adrenal and multi-focal brain sources.

*Ex vivo* imaging of the freshly dissected organs (Figure III.9 insets) revealed that the most common site of colonization was the adrenal gland in all five of the mice with signal. Four of them additionally had one or more tumors in the brain and two had ovarian growths. Table III.I shows the penetrance and latency of the tumors. The earliest adrenal and brain metastases were detectable at two weeks and ovarian tumors were detected in the sixth week post injection. Adrenal, brain and ovarian tumor mean latency was 24, 67 and 38 days respectively.

To determine if the 361-luc model could be used for the study of systemic metastasis I used UGIC to seed all organs. I found that multiple organs could be colonized, and that the tumors were well tolerated. Thus the 361-luc system is viable for studying systemic colonization and as the latency period varies, there may be different mechanisms involved in awakening from dormancy in different organs.

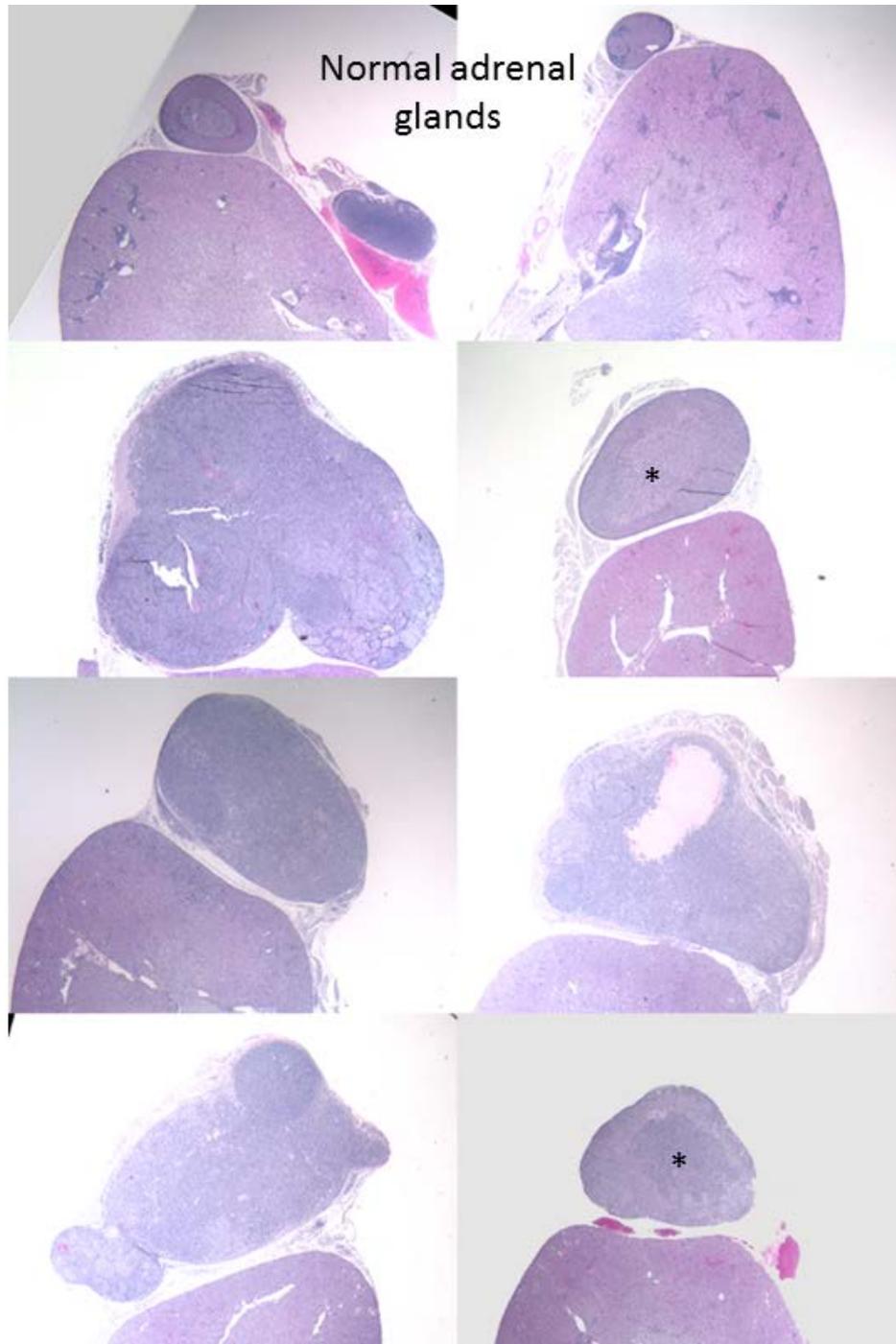
**Table III.I Latency and penetrance of bioluminescent tumors after injection of 1 million 361-luc cells into the left ventricle of NOD/SCID mice.** The organs colonized and the number of days post-injection (PI) when the signal was first detected are listed.

Organ	Earliest Onset (Days PI)	Mean Latency (Days PI)	Penetrance
Adrenals	14	24	50%
Brain	14	67	40%
Ovary	38	38	20%

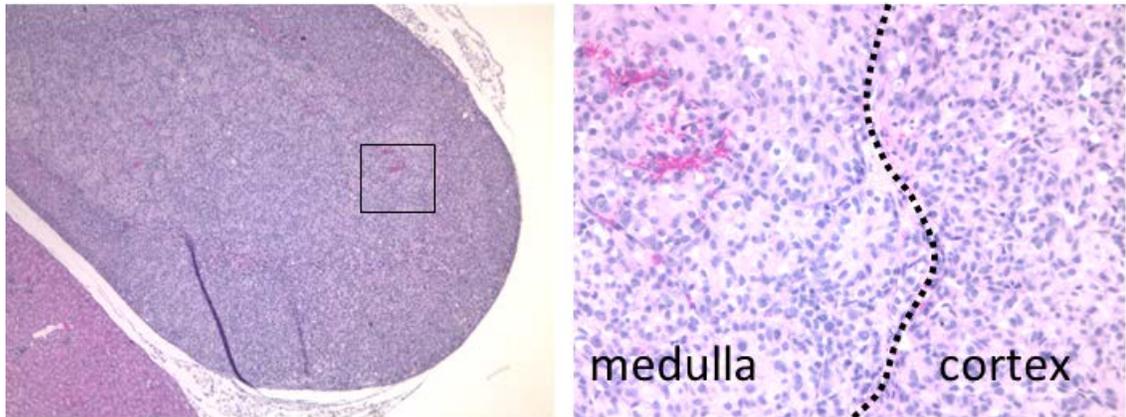
### III.3.1 Histopathology of 361-luc adrenal tumors

The adrenal glands were swollen in size up to one centimeter with bumpy white tumors but the membranes containing them appeared to still be intact. H&E staining of early adrenal tumors shows the most pronounced growth in the medulla followed shortly thereafter by displacement of the cortex (Figure III.10). The integrity of the adrenal capsule

was not found to be compromised in any tumors. Some growths consisted only of one primary mass while others had distinct lobes, possibly due to separate initiating events. The higher resolution H&E section in Figure III.11 show that the boundaries between the medulla and cortex are still discernable. Erythrocytes can be seen throughout the tumors indicating that they are well vascularized. As adrenalectomy is fatal the adrenal glands were still able to provide their vital functions even in those animals with bilateral colonization.



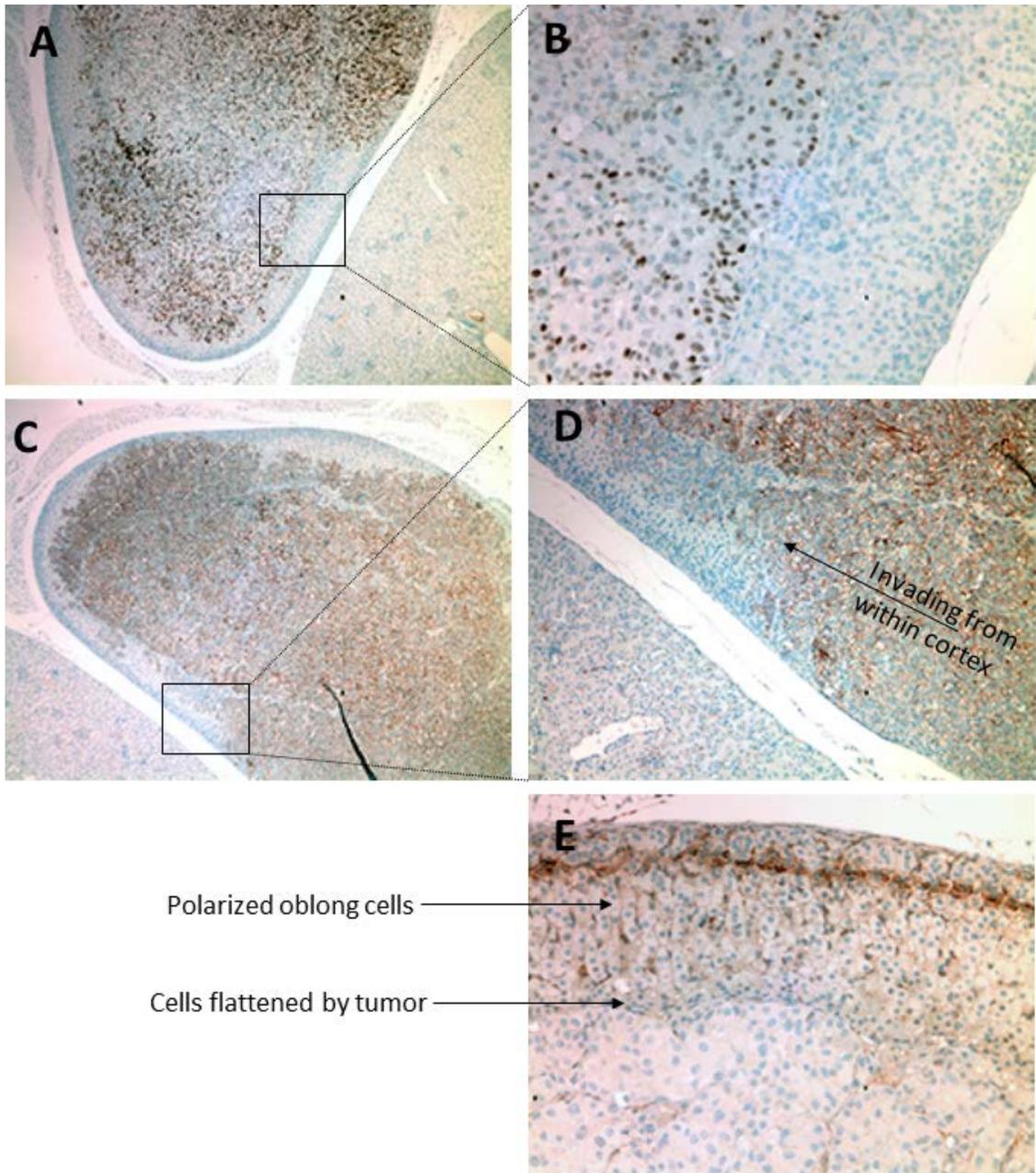
**Figure III.10 Adrenal metastases from intracardiac injection of 361-luc.** The top row are normal adrenal glands from a control animal without injection photographed at the same magnification. \* demark early growths originating in the adrenal medulla. Several tumors have distinct lobes possibly due to multiple initiation events within one adrenal gland. All images at the same magnification.



**Figure III.11 Higher magnification of adrenal boundaries.** While the tumor has invaded the cortex, the two layers are still distinguishable.

### III.3.2 Immunohistochemical analysis of adrenal 361-luc tumors

ER staining in Figure III.12A and B can better distinguish the tumor cells from the adrenal cells revealing the full extent of the tumor invasion within one adrenal gland. The adrenal medulla is clearly fully occupied by tumor cells, however in the inset portion of the growth the cortex has not been penetrated. At the top of panel A the tumor has fully invaded the cortex and no boundary is discernable. Additionally Her2 staining in panels C and D at the surface of the tumor cells more completely reveals the pushing border of the medullar tumor cells displacing the cortex. This border looks very different than the 361-luc orthotopic tumors invading the adipose stroma in the breast. A phalanx of tumor cells can be seen invading from the right flank having penetrated the cortex elsewhere. A PR antibody used for IHC stained mostly non-specific mouse protein as evidenced by the weak positivity of the cortex and a stronger signal at the border between the cortex and the capsule in age-matched animals that were not injected. This host staining, while useless for measuring PR, resolved the fate of the cortical cells in direct proximity to the expanding tumor revealing in panel E how they were crowded and flattened as the tumor pressed on.



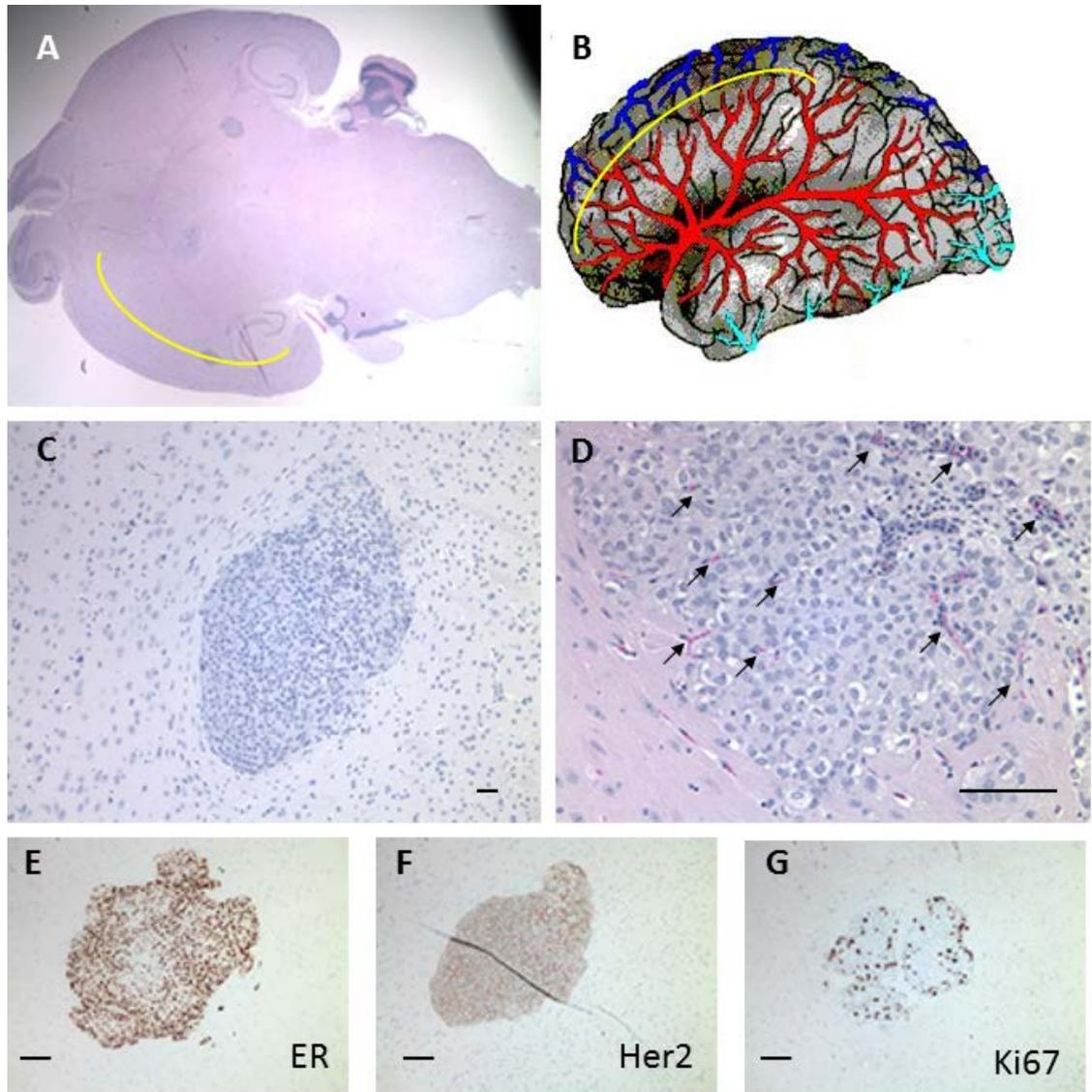
**Figure III.12 IHC of 361-luc adrenal metastases.** A & B) ER staining shows the distinction between the adrenal medullary tumor and the cortex. C & D) Her2 staining depicting an invading phalanx of tumor cells. E) Non-specific stain shows the deformation of the cells in proximity to the tumor.

### **III.3.3 Histopathological characteristics of brain metastases**

The brain tumors were all too small to be discernable by the naked eye. Tumors were found by decreasing tissue size via rounds of dissection and imaging until the bioluminescent focus had been identified. Tumors were located in the cortical parenchyma as well as the pituitary.

There were multiple foci present in most brains. Interestingly, as represented in Figure III.13, tumors in the cortex arose along a boundary between the outer and inner cortex which correlates with the most distal blood supply to this region. Blood flow from two of the three arteries that supply the brain, the anterior and middle cerebral arteries, ends here. The border between the tumor and brain parenchyma was distinct with host cells along the perimeter again visibly compressed and flattened by a pushing border as in the adrenal cortex. Capillary or arteriole blood access is evident with many intratumoral erythrocytes depicted within micro vessels. The now familiar lobular tumor growth pattern described in the mammary fat pad and the adrenals was evident.

The brain tumors were all much smaller than the mammary, adrenal and ovarian tumors. The proportion of Ki67 positive cells was lower in the brain tumors when compared to the adrenal tumors and Ki67 staining was excluded from the center of the tumors. The cells continued to display strong ER and Her2 staining.



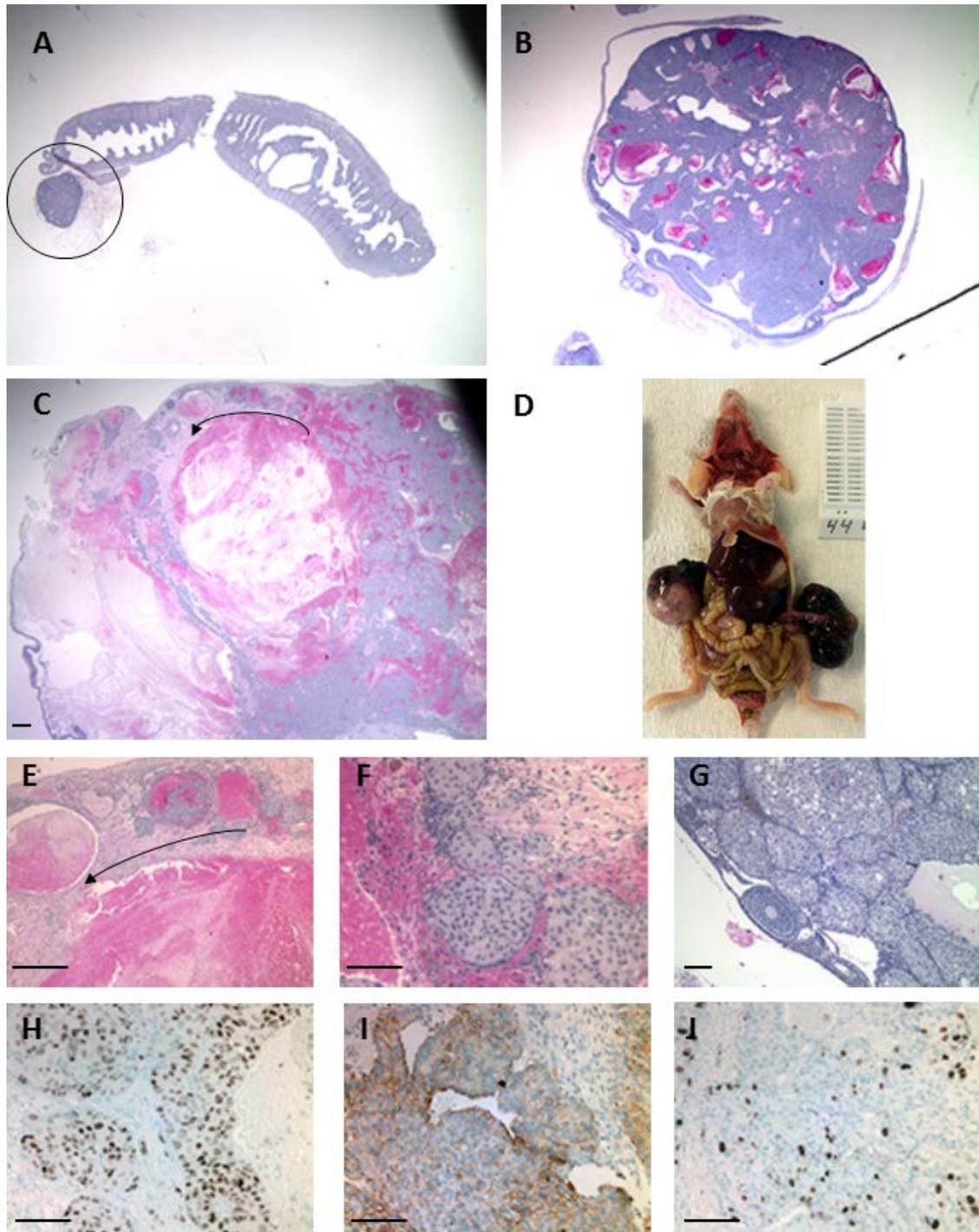
**Figure III.13 Histopathological staining of 361-luc brain metastases.** A) Three foci are evident in his horizontal section of the brain. One is growing on the yellow trace in A & B which represents the boundary between two capillary beds in the cerebrum. B) A depiction of the blood supply to the brain adapted from (Nonaka et al., 2003). C) A pushing border is evident flattening the host neurons. D) Intratumoral erythrocytes (black arrows) indicate a well vascularized tumor. IHC for E) ER F) Her2 and G) Ki67. The Ki67 signal is prevalent at the tumor surface but lacking in the tumor core.

### **III.3.4 Histopathological characterization of ovarian metastases**

Anatomically the ovary can be divided up into a cortex containing a loose arrangement of follicles in various stages of maturity and a medulla, all held together by a thin capsule. Under the influence of pituitary hormones a follicle matures into a secondary follicle made up of a large antrum and oocyte. The follicle produces a great amount of estrogen until it ruptures releasing the oocyte into the oviduct. A hormonal switch then occurs where the remaining follicle stops producing estrogen, changes to a yellowish color and condenses to form a corpus luteum and begins synthesizing progesterone in order to maintain habitability of the uterine lining by a fertilized oocyte. Progesterone synthesis continues as the corpus luteum regresses in size. Eventually it is completely degraded and the cycle begins again with the maturation of a new follicle (2014b).

The ovarian tumors had a pronounced angiogenic property with obvious tumor established vasculature that was not present at the other sites (Figure III.14). Tumors varied from approximately 3 to 15 mm in diameter and were spherical. The ovaries themselves were intact but enlarged and swollen. Externally the tumor coloring was patchy containing pockets of fluid alternating with bumps which were the color of clotted blood. The fimbriae of the oviduct still encased the ovary proximal to the uterus and there were no other signals in the viscera.

These tumors were the largest of any organ growing ten or more times larger than the ordinary size of the ovary housing them and were very cystic. Figure III.14 panels B C E & F shows a great amount of fluid filling in all of the interstitia in the organ implying that the hydrostatic pressure was high.



**Figure III.14 Histopathologic characterization of 361-luc ovarian metastases.** A) The circled region is a normal ovary. B & C) Equal magnification to A demonstrating high blood and fluid content and disruption of ovary structure. C) A follicle progression filled with blood can be seen at the crest of C. D) Mouse with large bilateral ovary tumors. E-G) Growth of tumor cells within and around follicles. H) IHC staining of ER I) Her2 J) Ki67. Arrows show the progression of follicles. Bar = 100um

Fluid leaked into the usually distinct follicles although some adjacent to the smaller tumors could still be seen developing. In many of the tumors the volume of fluid was half or more of the total tumor. The ovarian capsule was compromised so it is difficult to estimate the local origin of these tumors as the ultrastructure of the organ was no longer apparent. Tumors were found growing within follicles as well as throughout the entire ovary with the typical 361-luc tumor lobes evident (Fig. III.14G).

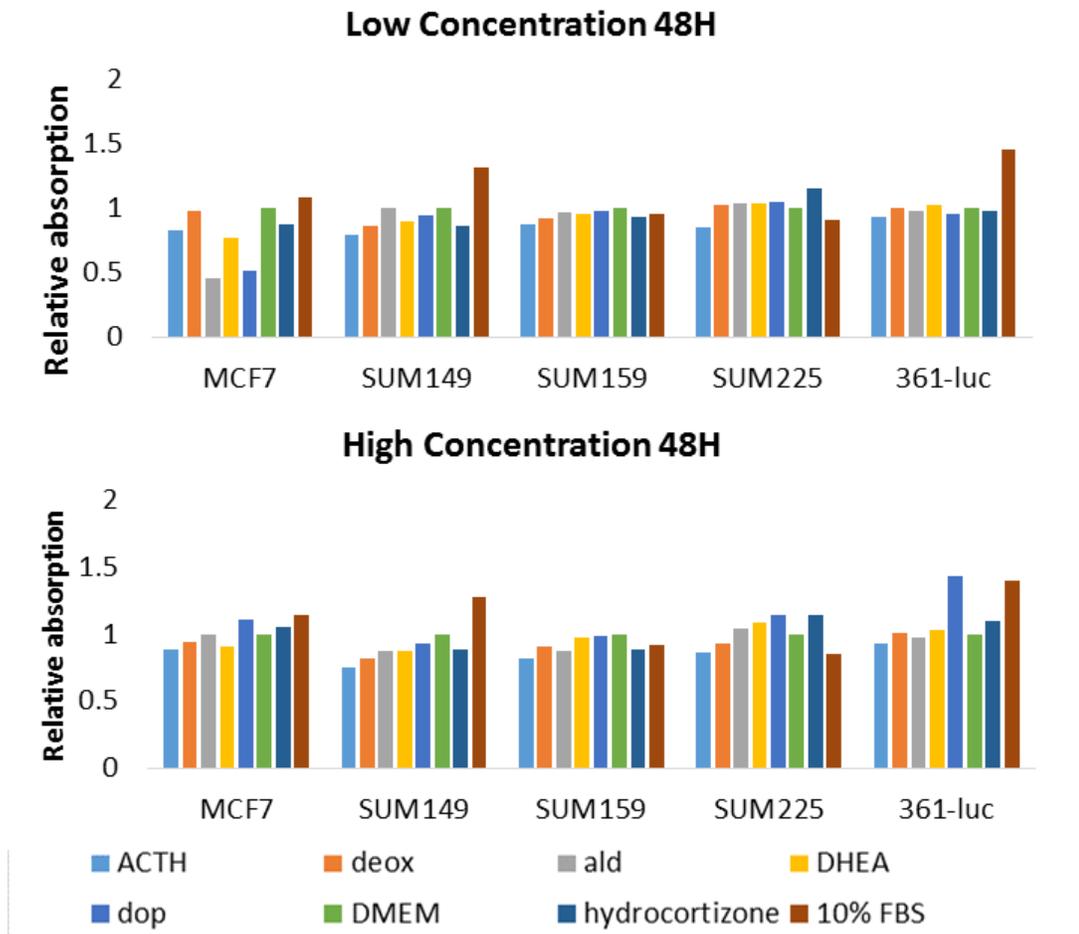
The 361-luc cells in ovarian tumors exhibited strong nuclear ER staining uniformly throughout the tumor (Fig. III.14H). Her2 staining (Fig. III.14I) at the cell surface was present but was not as uniform as ER staining with the strongest signal along the tumor periphery. The proportion of cells positive for Ki67 was higher than the brain tumors but lower than adrenal tumors (Fig. III.14J).

#### **III.4 Inherent adrenal tropism of the 361-luc line**

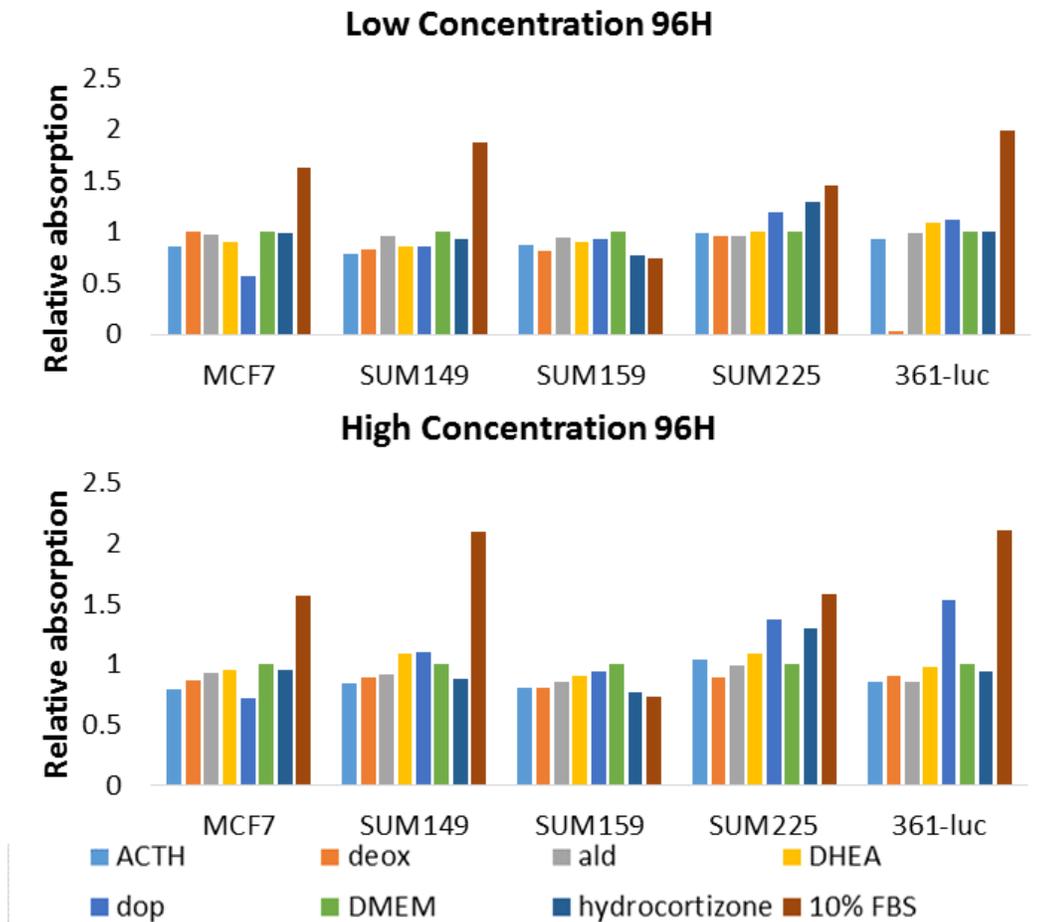
The primary organ colonized in the 361-luc model using either the parental line or any of the transduced cells in the coming chapter was the adrenal gland. To determine what adrenal specific traits may be involved in colonization here an *in vitro* screen was performed testing the effects of proliferation of the 361-luc line when treated singly with hormones of adrenal origin. After overnight serum starvation, the hormones in Table III.II were added in the listed physiologically relevant high and low doses and proliferation was measured by a viability assay. Tumor lines representing all of the breast cancer subtypes were screened. Figure III.15 and III.16 show at both 48 and 96 hours the 1uM concentration of dopamine exerted a statistically significant increase in viability in the 361-luc line. Only the positive control of 10% serum showed a parallel effect.

**Table III.II Reagents and cell lines used for *in vitro* screen of adrenal gland hormone viability.**

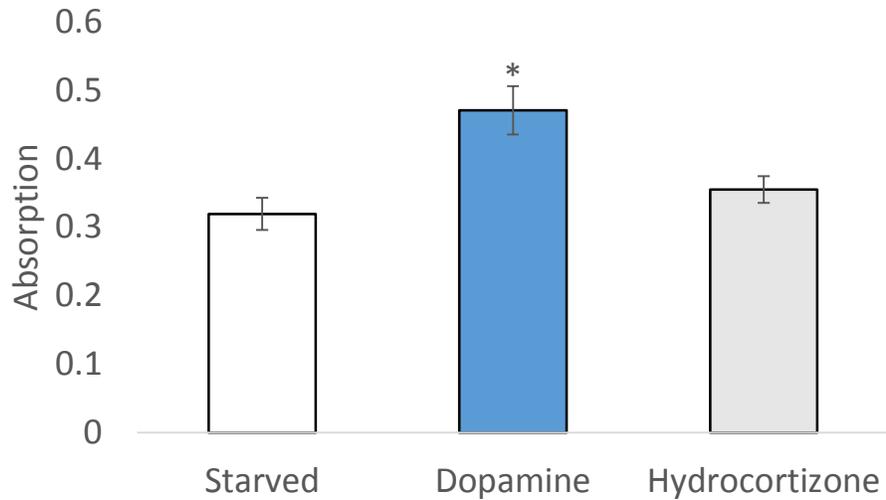
Hormone	Low dose (nM)	High dose (uM)	Cell line	Subtype
Adrenotropic hormone	10	1	MCF7	Luminal A
Aldosterone	10	1	SUM149	Basal
Deoxycorticosterone	10	1	SUM159	Mesenchymal
Didehydroepiandrosterone	10	1	SUM225	Her2
Dopamine	10	1	361-luc	Luminal B
Epinephrine	10	1		
Hydrocortisone	28	2.8		
Norepinephrine	10	1		



**Figure III.15** *In vitro* viability of 361-luc cells at 48 hours treated with hormones synthesized by the adrenal glands. Cell lines representing the molecular subtypes of breast cancer were treated at two concentrations with adrenal hormones and cell viability was measured 48 hours later. Fold change in absorption of light at 490 nm, directly proportional to the number of living cells, is displayed relative to cells growing unsupplemented.



**Figure III.16 *In vitro* viability of 361-luc cells at 96 hours treated with hormones synthesized by the adrenal glands.** Cell lines representing the molecular subtypes of breast cancer were treated at two concentrations with adrenal hormones and cell viability was measured 96 hours later. Fold change in absorption of light at 490 nm, directly proportional to the number of living cells, is displayed relative to cells growing unsupplemented.



**Figure III.17 Dopamine increases viability of 361-luc cells.** 361-luc cells were treated with 1uM dopamine and viability was significantly increased relative to starved cells as measured by light absorption in the MTS assay. The adrenal hormone hydrocortizone was included for comparison. \*p = 8.6E-05

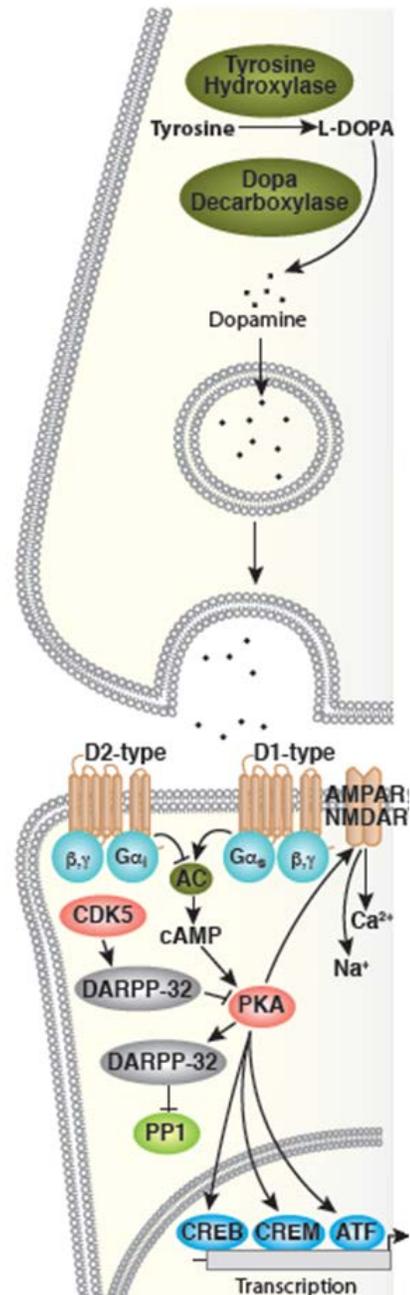
Repeated testing of dopamine against hydrocortizone as a negative control adrenal hormone revealed a statistically significant increase in relative viability compared to starved cells (Fig. III.17). These data suggest that dopamine provides a metastatic advantage to 361-luc cells.

#### III.4.1 Investigation of dopamine as a contributor to 361-luc viability

Dopamine is recognized by two families of G-protein coupled receptors. The type 1 dopamine receptors DRD1 and DRD5 activate adenylyl cyclase increasing cyclic adenosine monophosphate (cAMP) which is a second messenger that activates the protein kinase A (PKA) complex by dissociating the regulatory subunits freeing the catalytic subunits to carry out a phosphorylation cascade (Figure III.18). The type 2 receptors DRD2, DRD3 and DRD4 inhibit adenylyl cyclase and thereby attenuate PKA signaling. Quantitative polymerase chain reaction (qPCR) was used to measure the expression of the dopamine receptors on 361-luc cells and all five were found to be expressed at levels

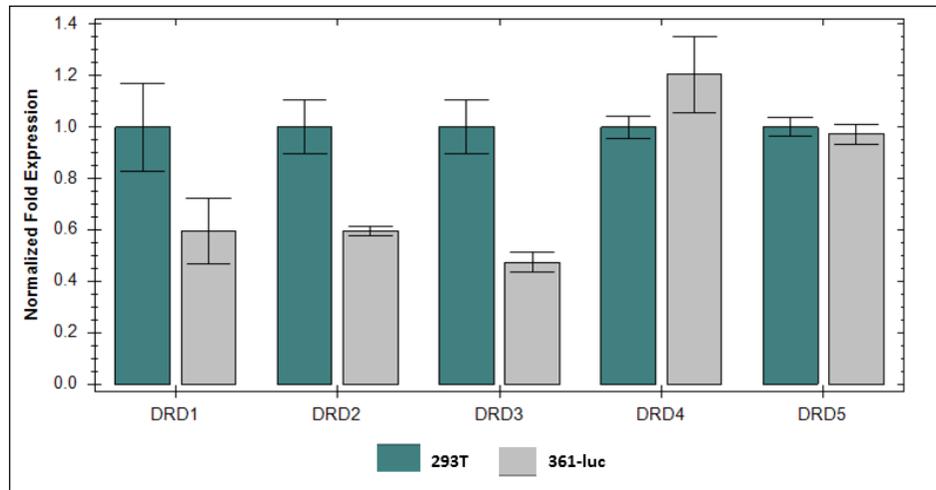
similar to the positive control embryonic kidney derived neural 293T cell line (Figure III.19). Additionally, dopamine can be bound by reuptake receptors SLC6A3 and SLC29A4 ordinarily found on presynaptic membranes. These proteins are not known to activate intracellular signaling but instead shuttle dopamine to packaging proteins that store it in vesicles. These two receptors were also found by qPCR to be expressed in 361-luc cells. 24 hours after stimulation of 361-luc with 1uM dopamine DRD4 expression was found to be significantly decreased and returned to baseline by 48 hours. Similarly, the level of reuptake receptor SLC6A3 message was also found to drop after 24 hours and recover by 48 (Fig. III.20A).

The biosynthetic enzymes involved in dopamine biosynthesis were also investigated by qPCR. There are three dopamine specific enzymes in this pathway displayed in Figure III.20B. Expression of tyrosine hydroxylase, the rate limiting enzyme in dopamine biosynthesis, changed significantly after 24 hours when 361-luc cells were stimulated with dopamine.

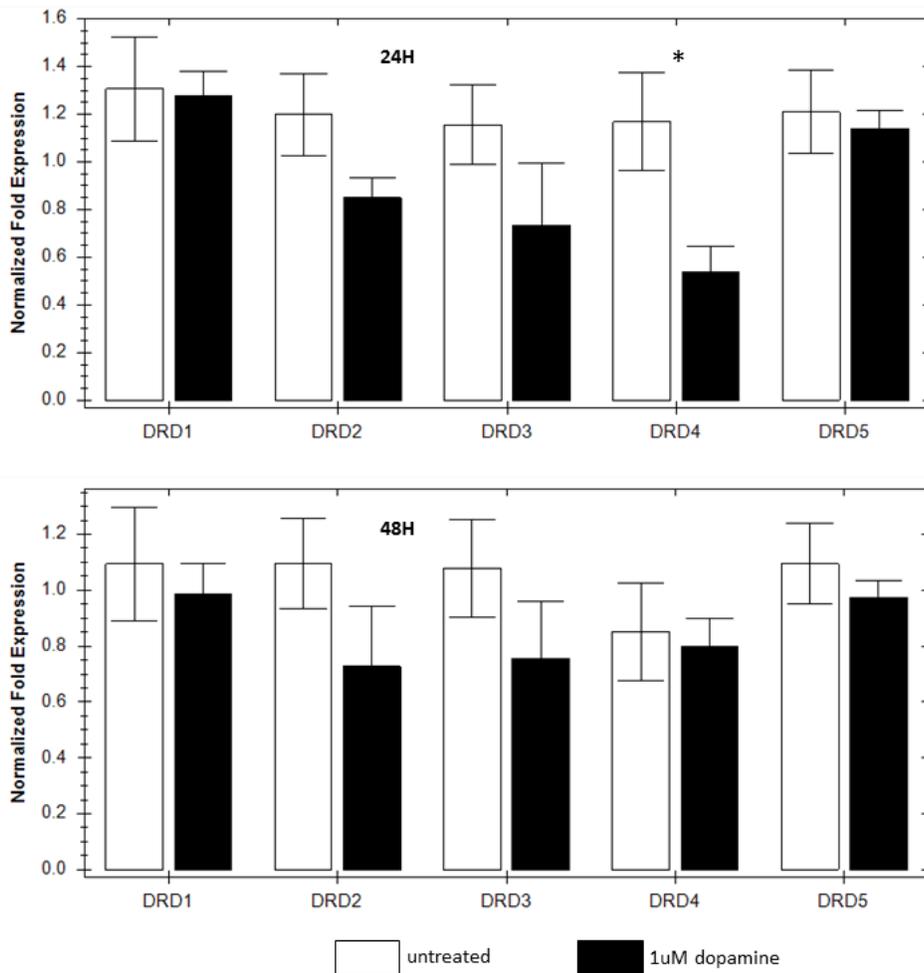


**Figure III.18 Dopamine signal transduction.** Illustration reproduced courtesy of Cell Signaling Technology, Inc. ([www.cellsignal.com](http://www.cellsignal.com)).

A

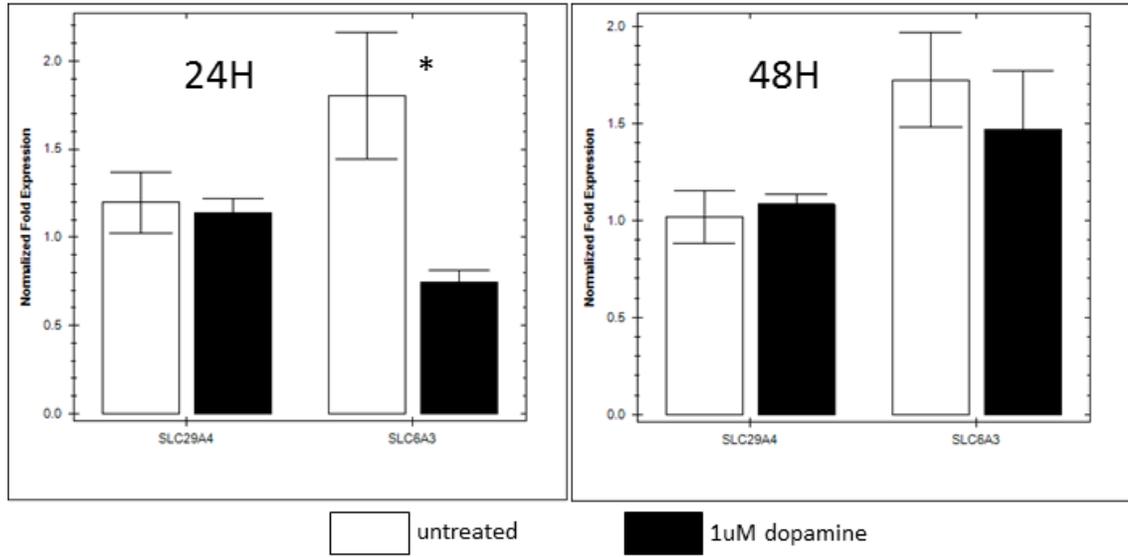


B

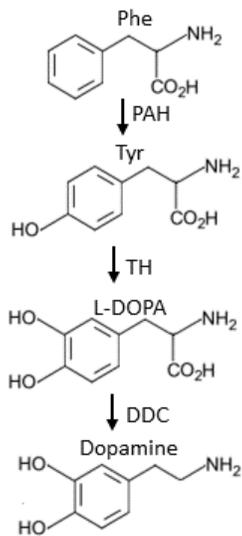


**Figure III.19 Dopamine receptor expression in 361-luc cells.** A) All five dopamine GPCRs are expressed by 361-luc similar to 293T levels. B) 24H after treatment with 1uM dopamine DRD4 levels are significantly different compared to untreated cells. By 48H levels have normalized. \*  $p < 0.05$

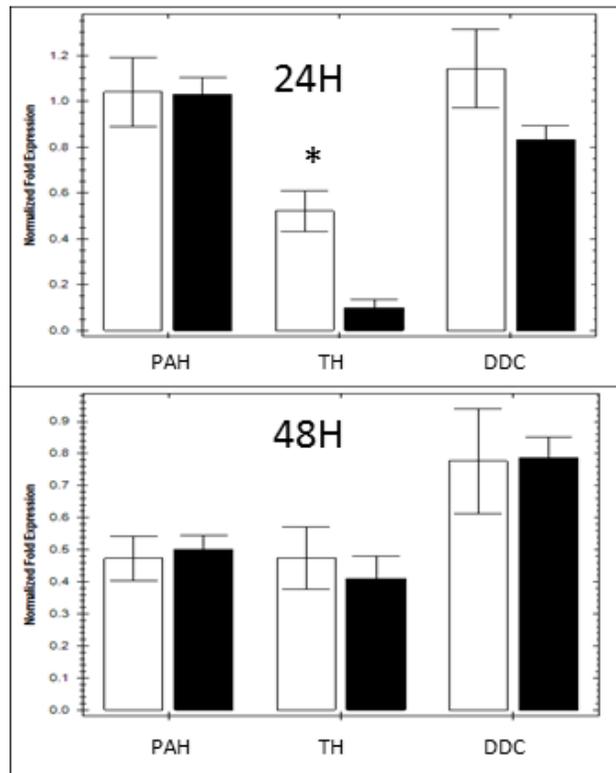
A



B

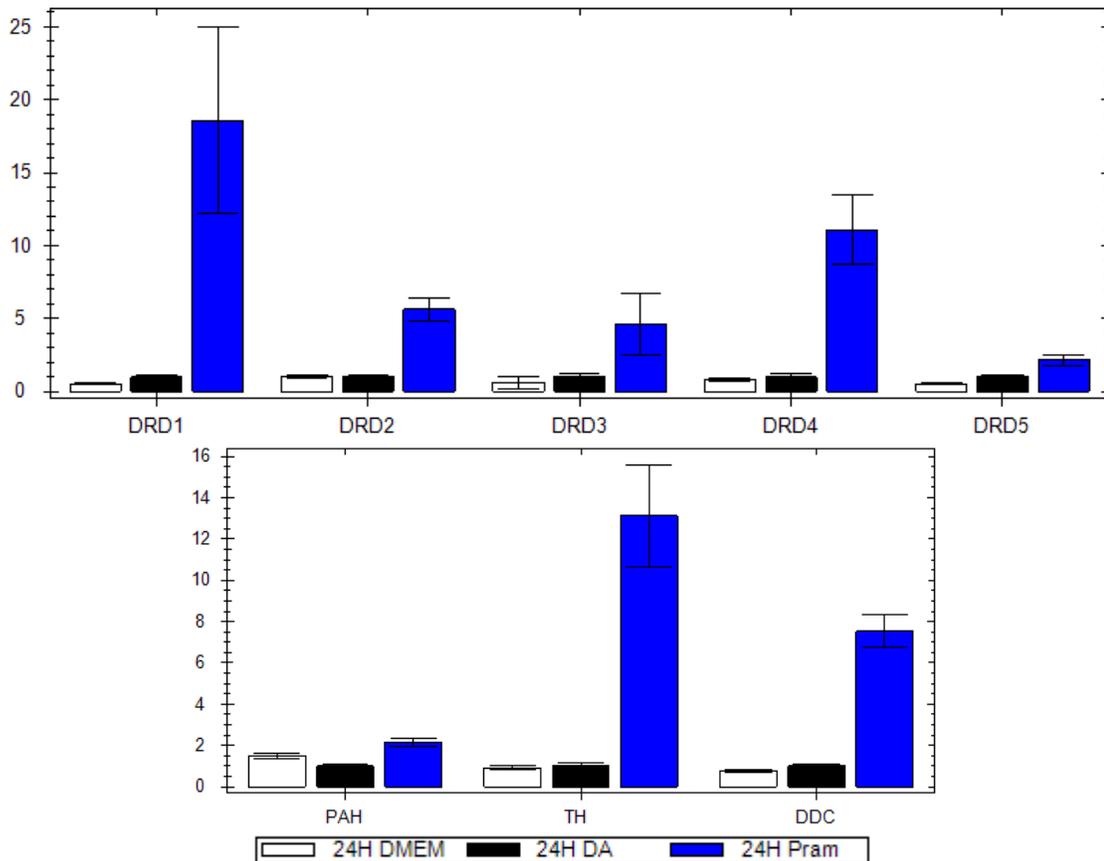


C



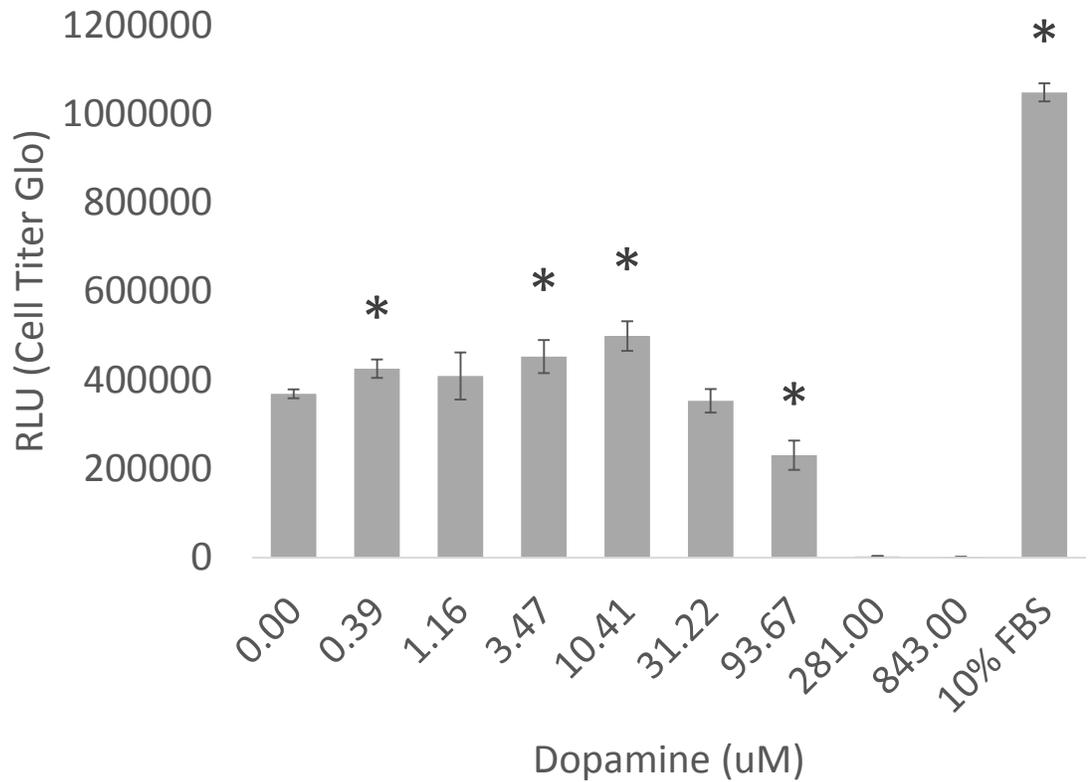
**Figure III.20 Dopamine reuptake receptors and biosynthetic enzymes in 361-luc cells.** A) Levels of the reuptake receptor SLC6A3 drop at 24H and recover by 48H when stimulated with 1uM dopamine. B) The dopamine biosynthetic pathway. C) The rate limiting enzyme TH changes significantly 24H after dopamine treatment.

Again levels returned back to baseline by 48 hours. Pramipexole (Px) is a potent type-2 dopamine receptor agonist. When 361 cells were treated with Px the expression of all five dopamine receptors increased from three to eighteen fold (Figure III.21). All three enzymes in the dopamine biosynthetic pathway were higher with TH increasing 13 fold in 24 hours.



**Figure III.21 Pramipexole stimulation of 361-luc.** 24 hours after treatment with Px all five dopamine GPCRs increased as did the biosynthetic enzymes.

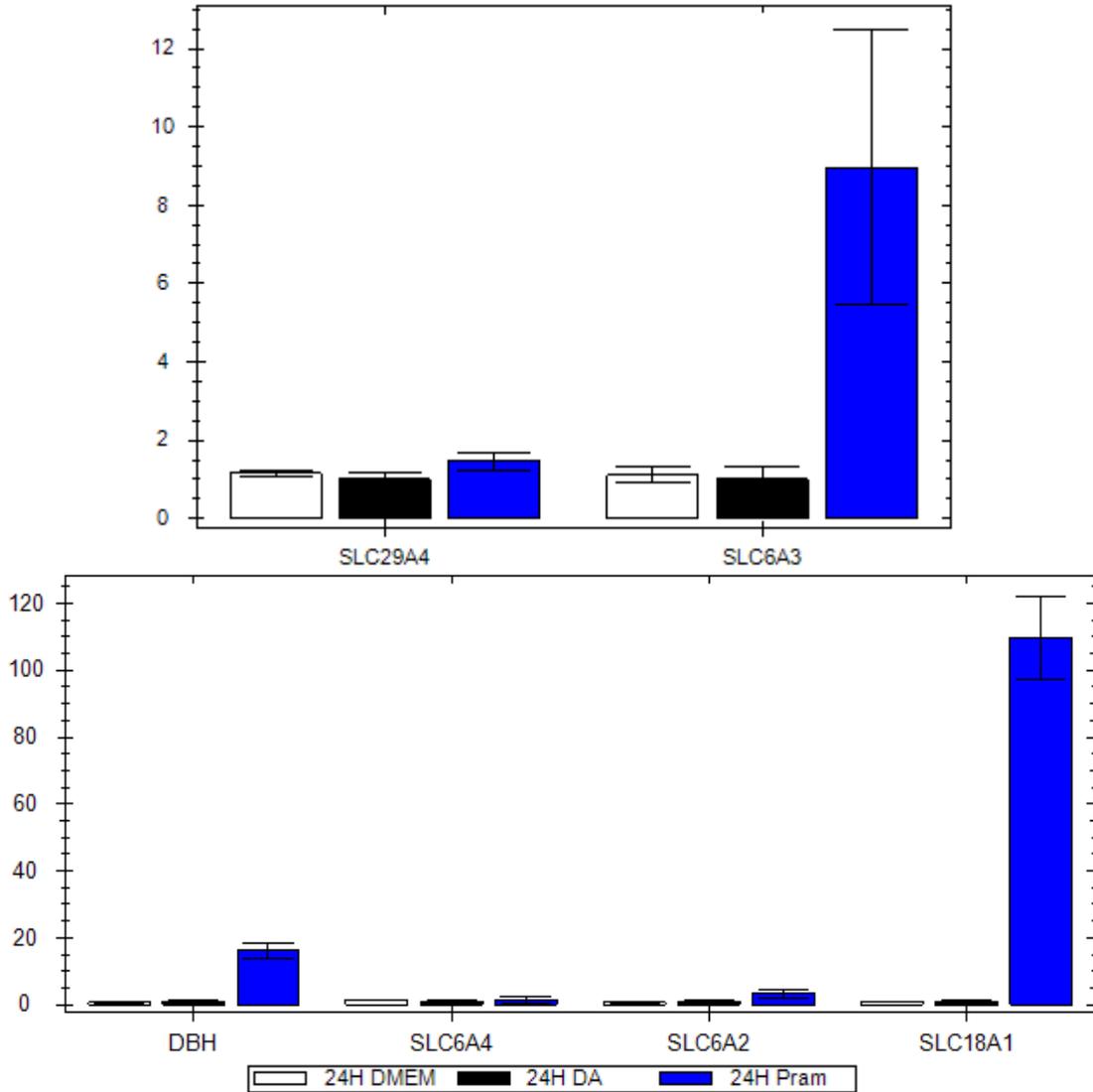
In neurons, dopamine is packaged in vesicles and free dopamine is a cytotoxic oxidizer. The dopamine response curve in the viability assay of Figure III.22 shows this effect in the dose dependence of the 361-luc response. Notably, cell viability decreased severely at the higher concentrations.



**Figure III.22 Dose response curve of 361-luc cells. \* p<0.05**

Px stimulation was used to investigate levels of the packaging proteins responsible for processing intracellular dopamine. In Figure III.23 levels of the vesicular monoamine transporter SLC18A1 also known as VMAT1, an integral membrane protein which directs cytosolic dopamine and other monoamines into vesicles increased over 100 fold. Both the dopamine reuptake pump SLC6A3 and the norepinephrine transporter SLC6A2 which can also reuptake dopamine from extracellular excess increased nine and four fold respectively. The related serotonin recycler SLC6A4 which also acts on monoamines, but not dopamine, remained unchanged. Dopamine in the cytosol can also be converted by dopamine beta-hydroxylase into norepinephrine which is further oxidized to inactive metabolites. This

enzyme was found to increase 18 fold by Px stimulation providing further evidence that dopamine can be processed by 361-luc cells.

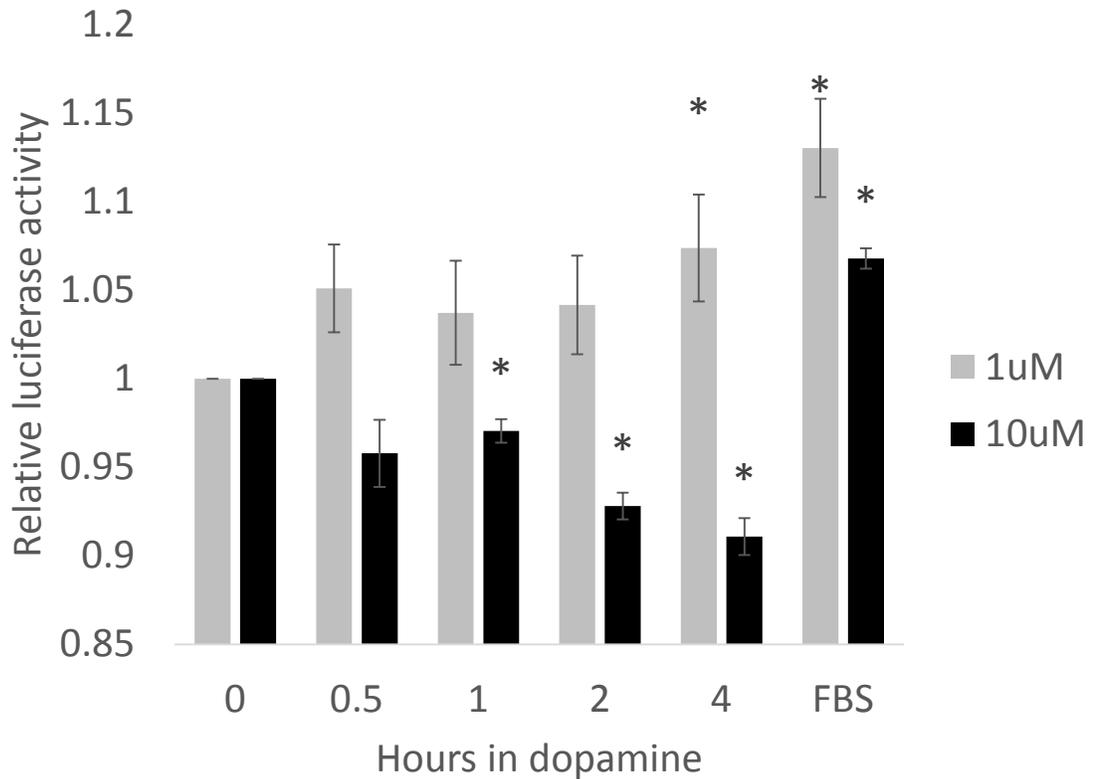


**Figure III.23 Px stimulation and vesicular machinery.** Reuptake receptor SLC6A3 levels increase 24 hours after Px stimulation. The vesicular transporter SLC18A1 (VMAT1) is potently induced as is the dopamine processing enzyme DBH.

### III.4.2 Dopamine signaling in 361-luc cells

After finding changes in the receptor, biosynthesis and packaging dynamics of dopamine related proteins in 361-luc cells I looked downstream to see if cAMP levels were

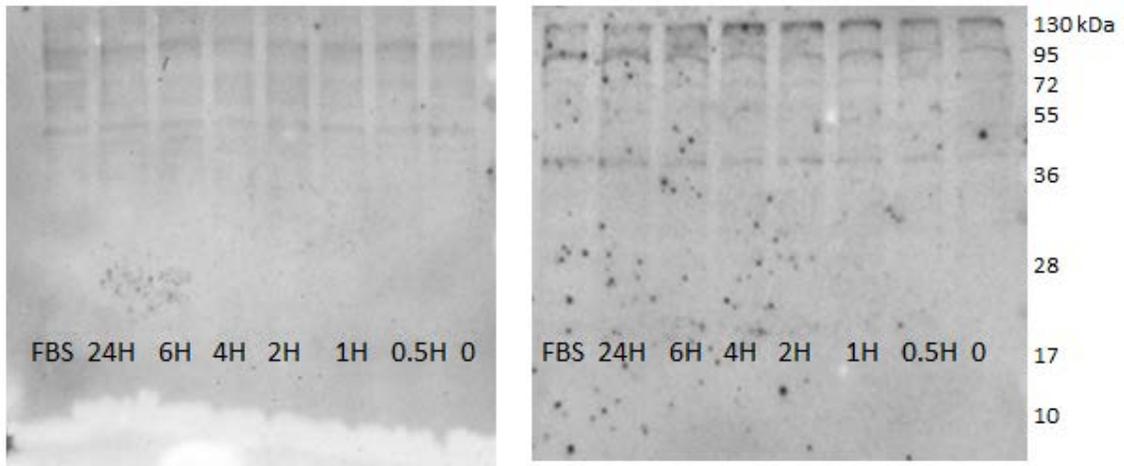
changing in response to dopamine using the cAMP-glo bioluminescent assay. This assay measures PKA activity in response to cAMP concentration changes. The amount of luminescence is inversely proportional to cAMP concentration. There were conflicting results in Figure III.24 with this assay showing a slight decrease in cAMP level after stimulation with 1uM dopamine, significant only at the four hour time point. But when the cells were treated with 10uM dopamine there was a significant increase in cAMP beginning at one hour.



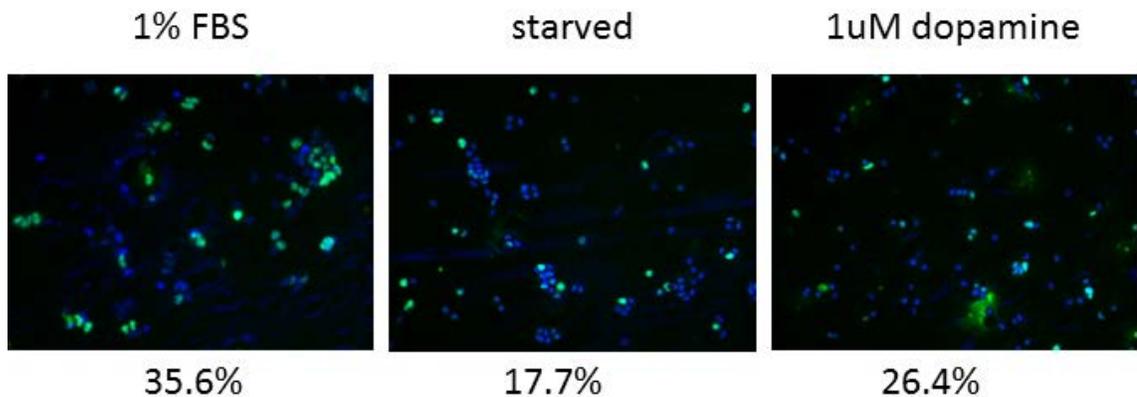
**Figure III.24 Bioluminescent signal due to PKA activity upon dopamine stimulation.** Relative luciferase activity is inversely proportional to cAMP concentration. With 1 uM dopamine cAMP levels decrease at 4 hours. But with 10uM dopamine cAMP levels increase from 1 to 4 hours.

A downstream substrate of PKA which is important in dopamine signaling in Parkinson's disease is DARPP-32 (Figure III.18). It serves as both a target and negative

regulator of PKA and propagates dopamine signaling to cell cycle regulation via type 1 protein phosphatase and Rb. This axis represents a mechanism whereby dopamine signaling controls cell cycle progression. However, the western blots in figure III.25 show that DARPP-32 levels were undetectable in 361-luc cells.



**Figure III.25 361-luc cells do not express the dopamine signal transducer DARPP-32.** After various time points of stimulation with 1uM dopamine 361-luc lysates were stained for DARPP-32 on the left and phospho-DARPP-32 on the right.

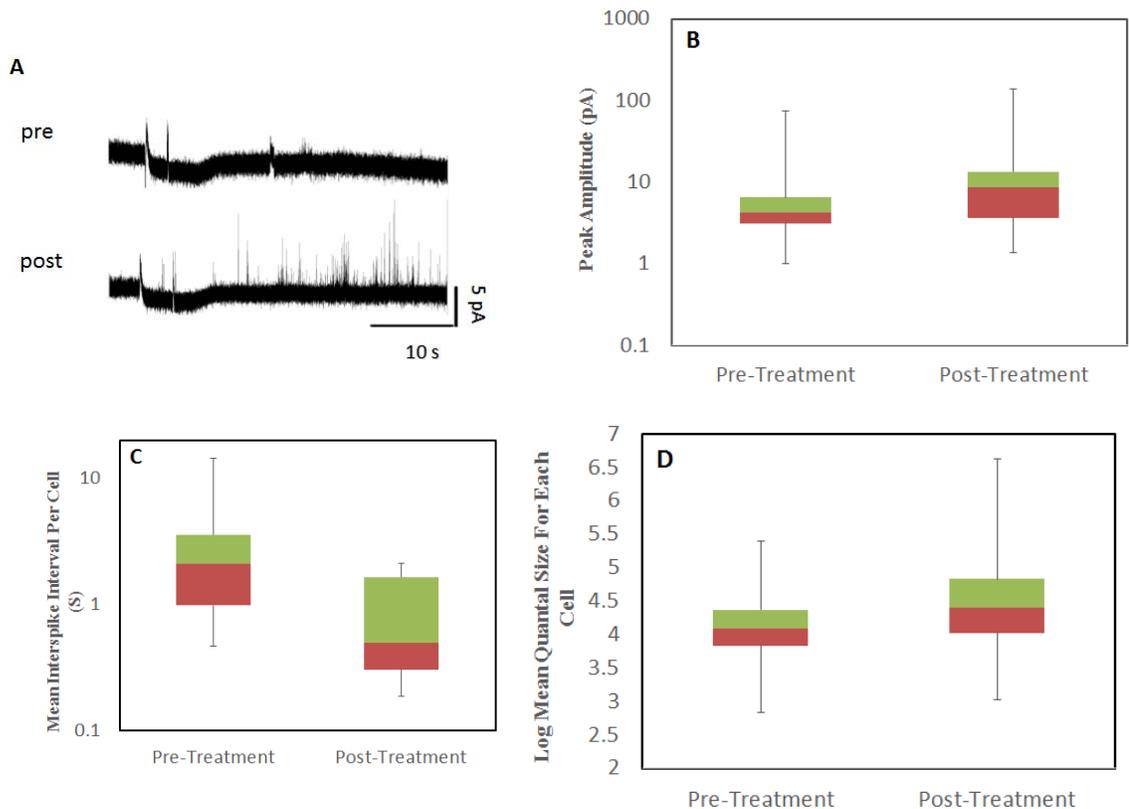


**Figure III.26 Immunofluorescent staining of Ki67 in 361-luc cells treated with 1uM dopamine.** The percentage of green nuclei are listed for each treatment. Dopamine treated cells had significantly higher Ki67 positive nuclei than serum starved cells.

I next looked at the proliferation marker Ki67 and Figure III.26 shows preliminary results that Ki67 staining is increased in cells treated with 1uM dopamine.

### **III.4.3 Functional dopamine electrophysiology**

In a collaboration with Dr. Emmanuel Pothos' lab, 361-luc dopamine related electrophysiology was examined. When individual cells were isolated under depolarizing conditions and treated with the dopamine precursor L-DOPA, quantal release was induced. The pre-treatment amplitude plot in Figure III.27A shows the baseline current peaks before addition of L-DOPA. The frequency of spikes caused by catecholamine release can be seen to increase significantly upon L-DOPA treatment. Each spike represents the exocytotic fusion of a single vesicle. The box plots demonstrate statistically significant increases in peak amplitude (B), and mean quantal size per cell (D) as well as a decrease in the mean interspike interval (C). Thus 361-luc cells are capable of storing and releasing catecholamines. As L-DOPA is known to upregulate dopamine synthesis and quantal release (Pothos et al., 1998) and is itself the immediate precursor to dopamine, the catecholamine being released is presumed to be dopamine. It is important to note that these experiments were done without addition of exogenous dopamine. This functionally suggests that 361-luc cells are capable of synthesizing and storing dopamine.



**Figure III.27 Depolarization induced quantal release in MDA-MB-361-luc cells.** A) Representative traces are shown from a cultured cell pre-treatment with L-DOPA and the response by the same cell post-treatment with L-DOPA. B-D) Box-and-whiskers plots from groups of cells pre-treatment and post-treatment with L-DOPA (unpaired samples) are demonstrated for the mean peak amplitude B) mean interspike interval C) and mean quantal size D) All graphs, post-treatment, are significantly different from the pre-treatment graphs ( $P < 0.05$ ) as indicated with the ANOVA test. Unpublished experiments conducted by Alireza Shirazian and Armen Ghazarian in the laboratory of Dr. Emmanuel Pothos.

To confirm the observation that dopamine imparts a viability advantage I examined expression of the receptors, biosynthetic machinery and processing enzymes. Changes consistent with active dopamine signaling were found for all three but I was unable to link the signaling biochemically to increased proliferation. Functional assays showed increased proliferation and the ability to store and release dopamine. Combined, these data suggest that dopamine can signal to increase proliferation of 361-luc cells and that the cells may be able to produce dopamine endogenously.

# Chapter IV:

## *An in vivo* screen of chromatin modifiers in metastatic colonization

## **Abstract**

Metastatic colonization involves survival and adaptation to a foreign, possibly hostile environment. As tumors grow out years after surgical removal of the primary growth, cells seeding distant organs experience a state of dormancy. How cells awaken from this dormant state is poorly understood. Chromatin state can control expression programs and chromatin dysregulation is known to contribute to malignant progression, so I hypothesized that chromatin modifiers may be involved in mediating this dormancy switch. To test this hypothesis I used two overlapping lentiviral RNAi hairpin libraries of human chromatin modifiers to singly knock down known chromatin modifiers and then injected the transduced cells using the system described in the previous chapter. I found that the organ tropism changed little but there was a difference in the incidence of adrenal vs nonadrenal tumors. I discovered that four genes EHMT1, JMJD5, PRDM14 and SIRT3 were targeted more frequently in the nonadrenal group. Loss of any of these genes, particularly SIRT3 showed a poor prognosis in clinical outcomes of metastasis. However, *in vitro* assays of invasion and stemness as well as *in vivo* colonization failed to reproduce a difference in organ tropism upon SIRT3 knockdown.

#### **IV.1 A screen for chromatin modifier involvement in metastatic colonization**

Genetic instability is a hallmark of the cancer cell state. It contributes to the stepwise nature of malignant progression. However, once a cell or micro metastasis becomes separated from the primary tumor, it evolves on its own trajectory. In heterogeneous tumors like breast cancer, this nascent tumor will have a limited founding genetic repertoire compared to the primary tumor's many clones. So it is logical to assume that the new tumor would need to acquire additional mutations that lead to a metastatic phenotype enabling this cell to survive and grow out in spite of the foreign milieu. Indeed there are reports of new mutations in genes occurring in metastases and some of these mutations have been confirmed to function in this context. However, increasing experimental evidence has proven that the primary tumor has everything it needs to spawn clones which can successfully colonize a distant site without any new driver events. Proof that metastases without new drivers are the norm rather than the exception can be found from the most recent reports of single cell sequencing of clinical primary and secondary tumors. However, given that we know that a small proportion of cancers can arise without any known significant driver mutations at all means that we may never be able to rule out the synergistic effect of multiple mutations of smaller driver magnitude. It may well be that the additional genetic changes can push a cell beyond dormancy, but it appears more frequently the existing oncogenic machinery is retooled in a way that does not require mutation.

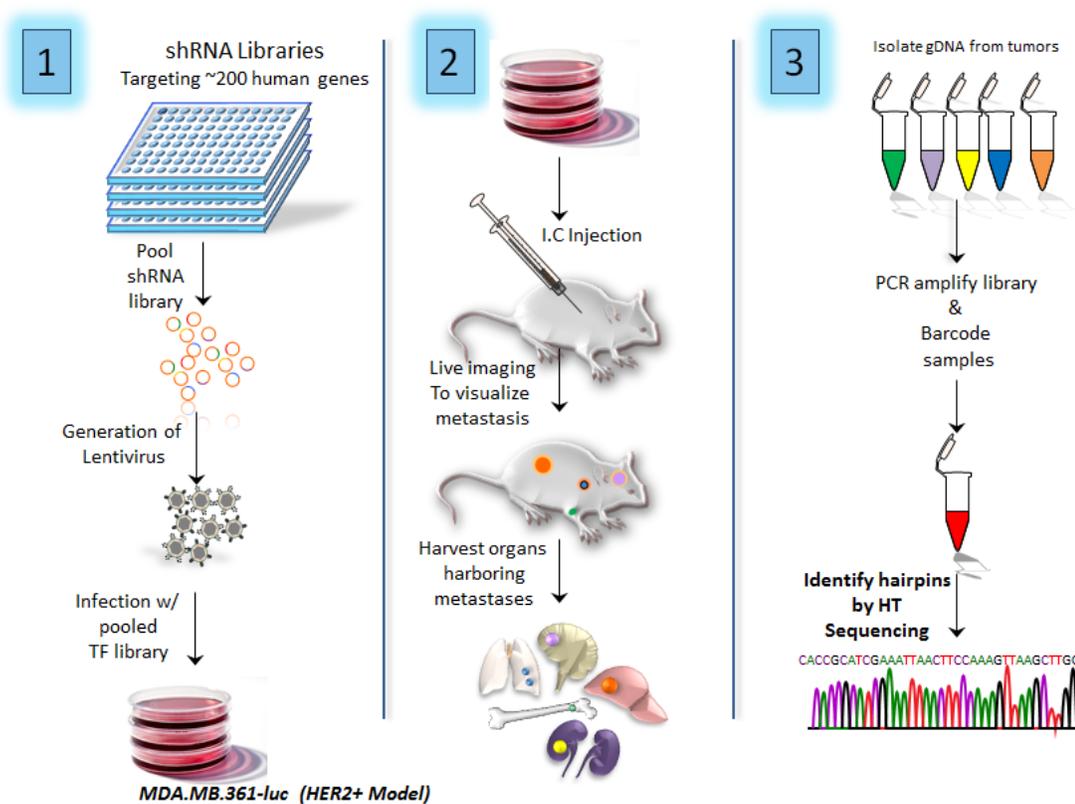
Learning that metastases can grow out without a significantly different genome than their founders caused me to speculate as to how micro metastatic dormancy is overcome. We also know that dormant single cells do not proliferate which means their capacity to

continuously acquire genetic damage is limited. Recognizing that there are changes in chromatin state in cancer cells and the existence of common mutations found in chromatin modifiers in specific cancers which are believed to drive progression suggested to me that chromatin modification itself was very likely to be a non-genetic change that facilitates awakening from metastatic dormancy. Even if it was not the initiating event, it appeared in my mind likely to be a necessary one. The formal hypothesis I wanted to test was that chromatin modifiers could regulate the dormancy of solitary cells after successful extravasation and limited colonization at secondary sites.

We turned to RNA interference and used the *in vivo* 361-luc metastasis system I characterized in the previous chapter to test this hypothesis. Collaborators in the Elledge lab supplied us with lentivirus containing a pooled library of virions encoding hairpins to knock down all known human chromatin modifiers. None of these hairpins had been experimentally validated so we decided to also include an overlapping library using the standard RNAi consortium vector pLKO for knock down provided by the Gupta lab.

It is noteworthy that when I began this project the 361-luc line had only been well characterized *in vitro* and its potential for metastasis was not established. The preliminary data available then, based on an experiment with three mice, showed that it did not metastasize when delivered systemically. But there were clearly signs of local invasion with tumor cell presence in vessels, inside ductal lumen and along nerves from the orthotopic site (Figure III.4). We reasoned that there was some deficiency, or perhaps an as yet unactivated latent program in the 361-luc cells that prevented them from successful colonization.

The experimental outline can be visualized in Figure IV.1 which shows that after ultrasound guided intracardiac injection (UGIC) and tumor growth the genomic DNA of the tumors was harvested. Each tumor was bar coded by including an index in the PCR primers used to recover the hairpins. This step also added flanking nucleotide adapters to allow flow cell binding for next generation sequencing. Thus each hairpin was identified and linked to an individual tumor.

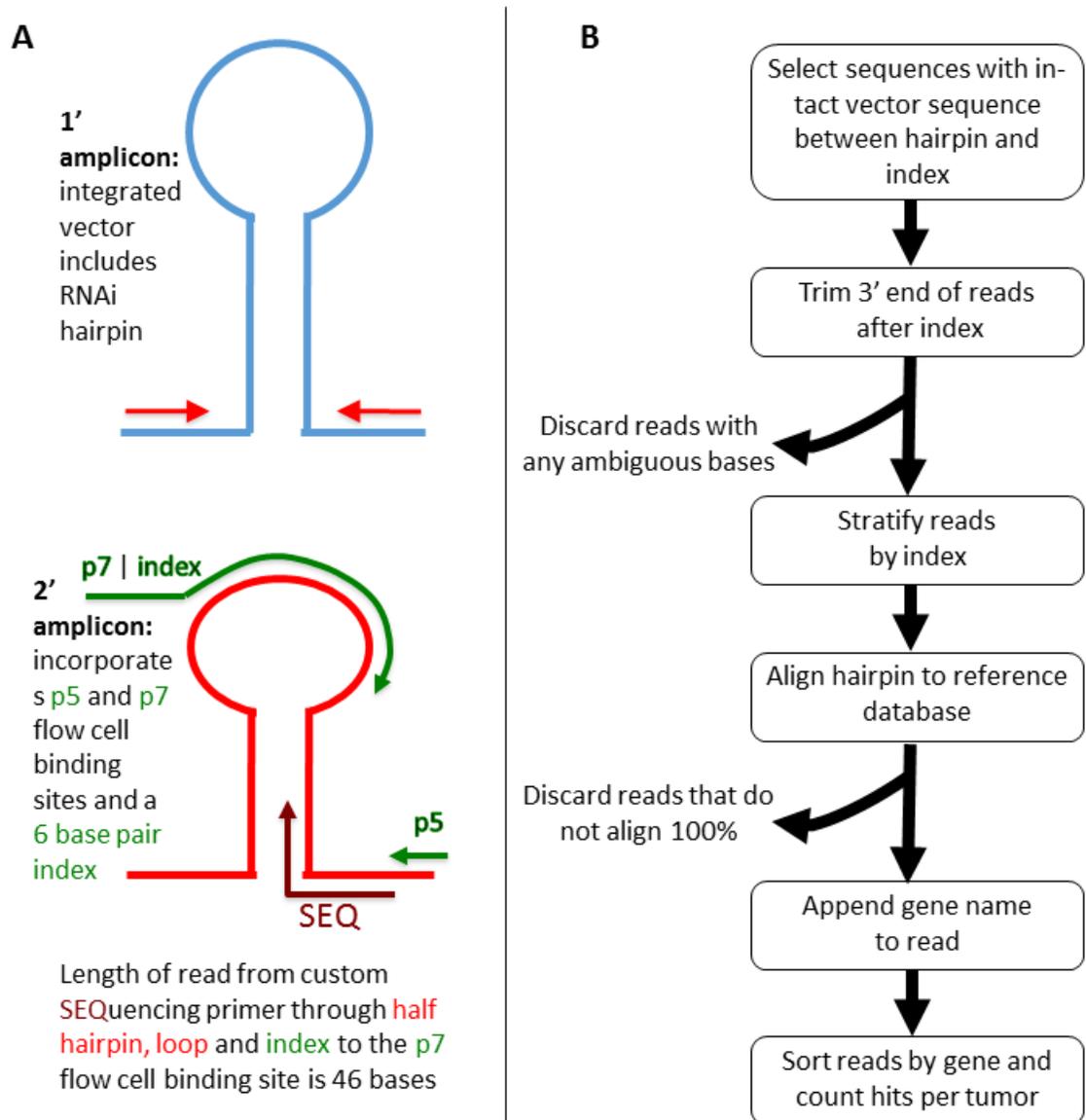


**Figure IV.1 Experimental approach for chromatin modifier knock down screen of 361-luc cells and metastatic colonization.** Adapted from Charlotte Kuperwasser

The bioinformatics flow used to recover and group the hairpins is shown in Figure IV.2. The number of reads per tumor was expected to be on the order of at least tens of thousands so the strictest possible criteria were used at all of the quality checkpoints to

determine hits. Hits were then matched to the reference hairpin and bar code databases for each library to determine the hairpin targeted and the tumor of origin.

Unfortunately, there was no positive control used in this experiment because one has not been proven to exist. It would have required the knock down of a gene that results in growth of a dormant metastatic colony.



**Figure IV.2 Bioinformatics scheme for recovery and analysis of hairpins from tumors.** A) 2-step PCR was used to add on flow cell binding adapters p7 and p5 and provide an index or bar code. B) Process for quality control and grouping of hairpins per tumor.

## **IV.2 Expected outcome**

A measure of success for this screen would have been a shift in metastatic behavior between the control cells and the library transduced cells, preferably one targeted by more than one hairpin for the same gene in more than one animal. Originally, I anticipated this would have been manifested by growth of secondary tumors anywhere in the body as the line was not known to be capable of colonization. However as the weeks went on it became apparent that there were metastases growing out in several of the control animals. This did not disprove the hypothesis but the focus was changed from an all-or-none colonization phenotype to a more nuanced shift in organ tropism. We now expected to define a baseline metastatic tropism in the control cells and some difference in the colonization of organs or growth patterns in the knock down libraries.

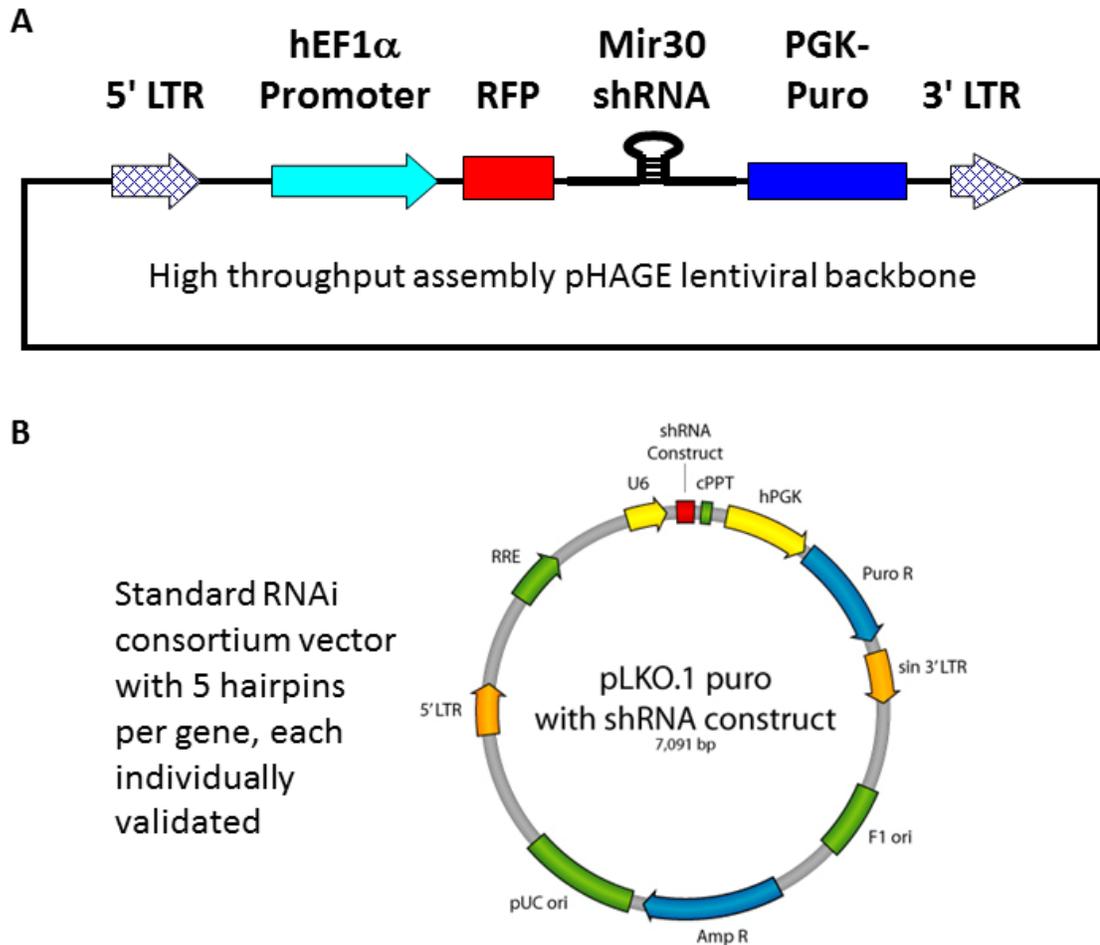
## **IV.3 Execution of the screen**

The two libraries used were provided by Dr. Steven Elledge and Dr. Piyush B. Gupta in the pHAGE and pLKO expression contexts respectively. The 217 target genes listed in red in Table IV.I were chosen by Dr. Elledge via curation of several comprehensive lists of chromatin modifiers based on proteins known to modify chromatin as well as those with similar structural domains. This pHAGE system uses Mir30 for hairpin processing (Figure IV.3) utilizing the human EF1a pol II promoter to transcribe a single message encoding an RFP protein, the Mir30 RNA and puromycin. Approximately fifty hairpins per gene were cloned to ensure a broad spectrum of knockdown efficiency with the potential for greater confidence in hits if multiple hairpins were present for a specific gene. This seemed especially well suited to chromatin modifiers as the amount of

knockdown could be important because chromatin modifier tumor mutations are usually heterozygous without total loss of the wild type protein (Garraway and Lander, 2013).

**Table IV.I Genes targeted by RNAi hairpin libraries.** Red indicates presence only in the pHAGE library, blue only in pLKO and purple in both libraries.

ACTL6B	CBX2	DNMT3L	HELLS	KDM4D	NAP1L1	PRDM2	SETD1A	SMYD1
ALKBH2	CBX3	DOT1L	HIRA	KDM5A	NAP1L2	PRDM4	SETD1B	SMYD2
ALKBH3	CBX4	EDF1	HLTF	KDM5B	NAP1L3	PRDM5	SETD2	SMYD3
ASH1L	CBX5	EED	HPSE2	KDM5C	NCOA1	PRDM6	SETD3	SMYD4
ASH2L	CBX6	EHMT1	HR	KDM5D	NCOA2	PRDM7	SETD4	SMYD5
ASXL1	CBX7	EHMT2	HSPBAP1	KDM6A	NCOA3	PRDM8	SETD5	SRCAP
ASXL2	CBX8	EIF4B	ING1	KDM6B	NR0B2	PRDM9	SETD6	SSRP1
ASXL3	CDH1	ELP3	ING2	L3MBTL1	NSD1	PRKAA1	SETD7	SUPT16H
ATAD2	CDY1	EP300	ING3	MBD1	PADI4	PRKAA2	SETD8	SUV39H1
ATAD2B	CDY2B	EP400	ING4	MBD2	PARP1	PRKCD	SETDB1	SUV39H2
ATRX	CHAF1A	ERCC6	ING5	MBD3	PARP2	PRMT1	SETDB2	SUV420H1
BAZ1A	CHAF1B	EZH1	INO80	MBD4	PAX5	PRMT2	SETMAR	SUV420H2
BAZ1B	CHD1	EZH2	JARID2	MECOM	PAXIP1	PRMT3	SFMBT1	SUZ12
BAZ2A	CHD1L	FANCM	JHDM1D	MECP2	PBRM1	PRMT5	SFMBT2	TAF1
BAZ2B	CHD2	FBXL19	JMJD1C	MEN1	PCGF2	PRMT6	SIN3A	TAF3
BLM	CHD3	FKBP1A	JMJD4	METTL8	PCMT1	PRMT7	SIN3B	TET1
BMI1	CHD4	FKBP2	JMJD5	MGMT	PCNA	PRMT8	SIRT1	TET2
BOP1	CHD5	FKBP5	JMJD6	MINA	PHF1	PSIP1	SIRT2	TET3
BPTF	CHD6	GTF3C4	JMJD7	MLL	PHF2	RAD54B	SIRT3	TRIM24
BRAF	CHD7	H2AFZ	JMJD8	MLL2	PHF20	RAD54L	SIRT4	UBE2A
BRD1	CHD8	HAT1	JUN	MLL3	PHF20L1	RBBP4	SIRT5	UBE2B
BRD2	CHD9	HCFC1	KAT2A	MLL4	PHF8	RBBP5	SIRT6	UBE2E1
BRD3	CLOCK	HDAC1	KAT2B	MLL5	PKD2	RECQL	SIRT7	UBE2I
BRD4	CPA4	HDAC10	KAT5	MORF4L1	POLR2B	RECQL4	SMARCA1	UHRF1
BRD7	CREBBP	HDAC11	KDM1A	MSL3	PPARGC1A	RECQL5	SMARCA2	USP17L6P
BRD8	CTNBN1	HDAC2	KDM1B	MST1	PRDM1	RING1	SMARCA4	USP22
BRD9	CTSL1	HDAC3	KDM2A	MTA1	PRDM10	RNF2	SMARCA5	USP27X
BRDT	CXXC1	HDAC4	KDM2B	MTA2	PRDM11	RNF20	SMARCA1	USP51
BRIP1	DEFA1	HDAC5	KDM3A	MTA3	PRDM12	RNF40	SMARCA1	UTY
C14orf169	DMAP1	HDAC6	KDM3B	MYST1	PRDM13	RNF8	SMARCC2	WDR5
C2orf60	DNMT1	HDAC7	KDM4A	MYST2	PRDM14	RPA1	SMARCD1	WDR82
CARM1	DNMT3A	HDAC8	KDM4B	MYST3	PRDM15	RPA3	SMARCD3	WHSC1
CBX1	DNMT3B	HDAC9	KDM4C	MYST4	PRDM16	SATB1	SMARCE1	WHSC1L1

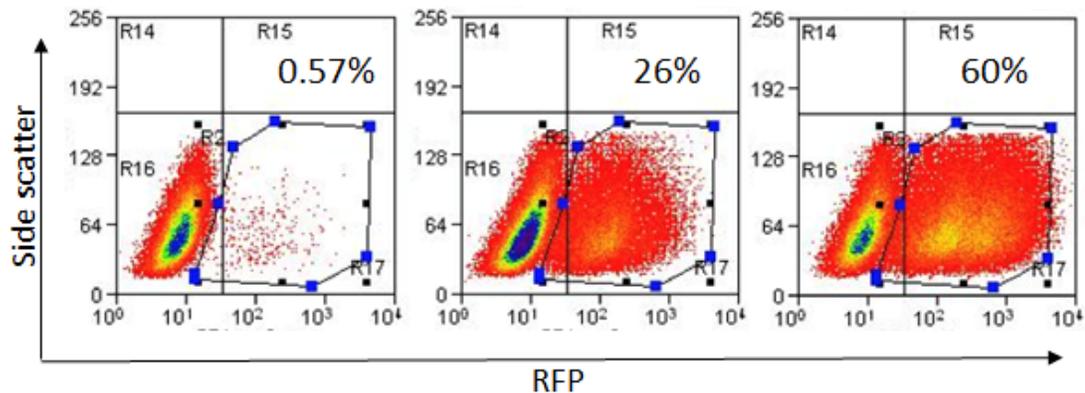


**Figure IV.3 Expression contexts for the two chromatin modifier libraries.** A) pHAGE library had approximately 50 hairpins per gene that were assembled using on chip ligation but were not validated. B) pLKO library had five hairpins per chip which were all verified to knock down their gene target.

The pLKO library was provided as 237 plasmids which I packaged into VSV-G pseudotyped lentivirus. pLKO is a polIII driven construct which also uses the Mir30 context for hairpin processing and includes puromycin selection (Fig. IV.3B). This library contained a more focused and individually validated five hairpins per gene. The two libraries targeted 157 common genes listed in purple in Table IV.I.

In order to properly represent all of the hairpins the Elledge lab suggested, based on their personal experience with *in vitro* screens, that one thousand cells per hairpin were

needed to be able to derive meaningful results beyond the stochastic nature of hairpin knock down screens. The pHAGE library had a total of 10,605 individual hairpins which would require around 11 million cells to represent them all effectively. The complexity of the pLKO library was nearly a full order of magnitude lower at 1,185 total hairpins. I estimated that an adult mouse could tolerate 1 million cells via UGIC, so I decided to inject 15 mice assuming there would be some morbidity. We used the same numbers of mice and cells for the pLKO library recognizing there would be a tenfold increase in the number of cells per hairpin because this library was less complex.



**Figure IV.4 MOI determination for pHAGE library transduction.** The cells from the center plot were used to ensure an MOI of 0.3 or less.

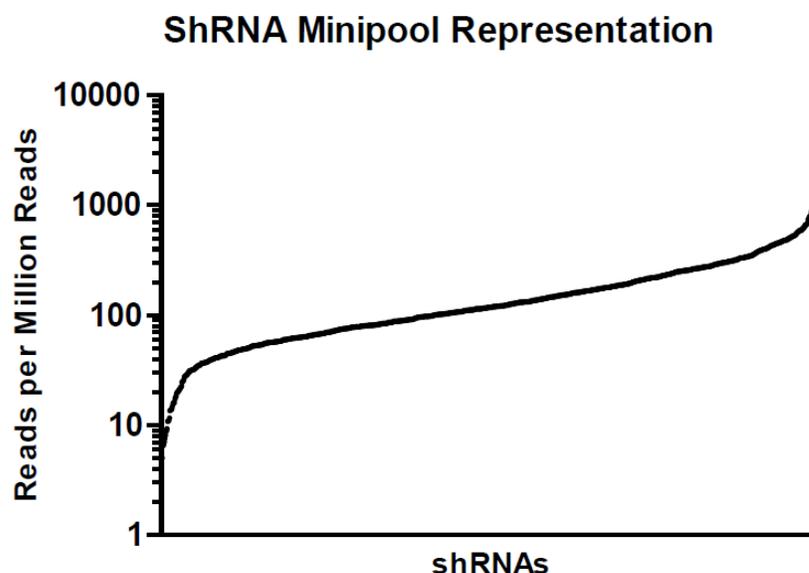
The virus was titered on 361-luc cells using dilutions to find a range where the effective multiplicity of infection (MOI) was below 0.3 (Figure IV.4). This ensured mathematically that no more than one hairpin would be transduced per cell. The cells were selected at the lowest possible puromycin concentration of 1.5 ug/ml that killed all of the non-transduced control cells in 48 hours in order to avoid selection of hairpins that might put cells at a proliferative or survival disadvantage in culture. The pHAGE library MOI was determined based on flow cytometry using the RFP readout and the pLKO readout was survival for 48 hours of puromycin selection. Control hairpins for both expression

vectors were used in an additional group of five mice each for a total of 40 mice according to the experimental groups listed in Table IV.II. One more group of twelve control mice was injected at a later date to raise the total animals harboring scramble hairpins (shScr) to that of the library groups.

**Table IV.II Experimental groups used for chromatin modifier metastasis screen.**

Group	Purpose	Animals injected	Animals surviving injection	Animals with thoracic leakage
pLKO	Library	15	15	2
pHAGE	Library	15	15	1
shScr5	Non-targeting pLKO control	5	5	0
shFF	Off-target pHAGE control	5	5	1
shScr10	Non-targeting control	12	10	1

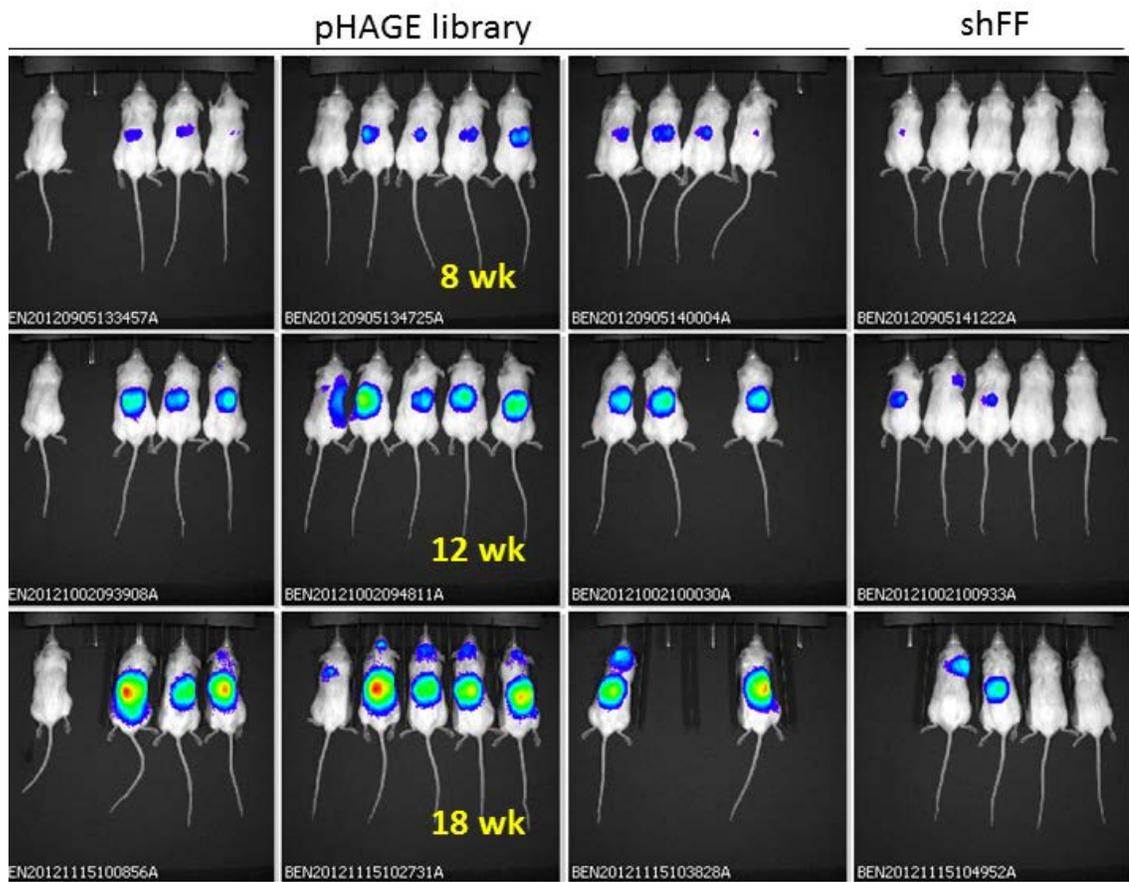
The pLKO plasmids were sequenced before I received them and every plasmid was confirmed to be present although the distribution in Figure IV.5 shows they were not all at the same copy number. Some transduced cells from each of the libraries were frozen, instead of being injected, as pre-screen pellets to be processed with the tumors, in order to determine the distribution of hairpins that actually made it into the cells before going into animals.



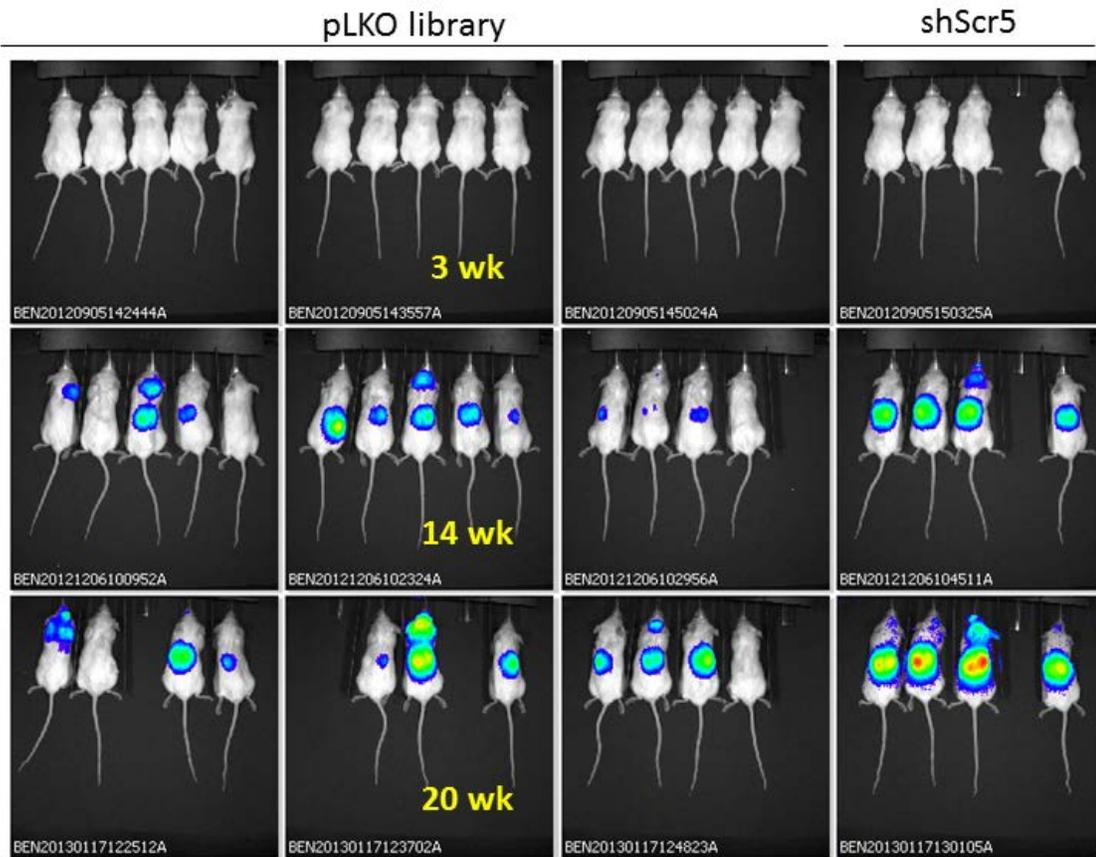
**Figure IV.5 Representation of all hairpins in the pooled pLKO plasmids.** Figure Provided by Rob Mathis, Gupta Lab, Whitehead Institute.

The mice tolerated the procedure very well with 48 out of 50 animals surviving UGIC. Thoracic leakage was detected in 12% of the animals (Table IV.II). As with the parental line described in Chapter III bioluminescent signal was detectable by the third week post-injection. The first signals to appear came from the adrenal glands and then later the brain. Table IV.III lists the latency and penetration for each of the libraries as well as the combined shScr control group. A total of 35 tumors were recovered from the pHAGE library group and 38 from pLKO.

The organ tropism observed with the parental line was recapitulated in the experimental groups for the metastasis screen as well. The site most frequently colonized was the adrenal gland followed by the brain. Figures IV.6 and IV.7 visualize the metastases growth over time in the library groups. There were other organs than these two colonized but by only one tumor for each site within an experimental group. Individual tumors were found in the library groups in the ovary, lung, pleura, thoracic wall, esophagus, rib, heart, eye, and viscera. None of these sites were colonized in the control animals.



**Figure IV.6 Metastatic outgrowth over time of the pHAGE library.**



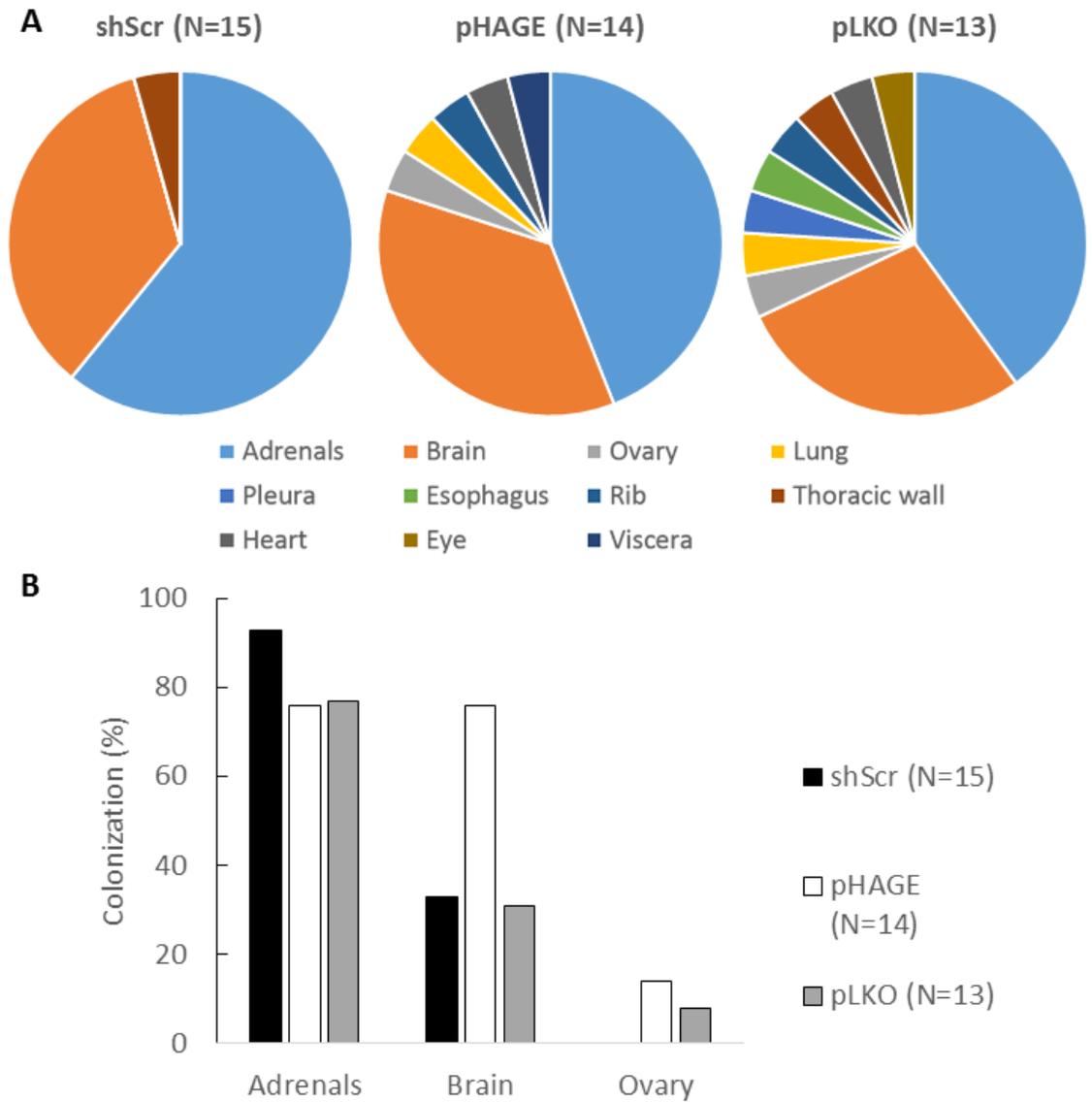
**Figure IV.7 Metastatic outgrowth over time of the pLKO library.**

Figure IV.8 shows a complete representation of the bioluminescent signal detected per organ. There is a change in the distribution of sites colonized in the library groups compared to either the parental line or the shScr transduced control cells. In figure IV.8A the libraries clearly have more slices in their pies. However, statistical analysis to determine the significance of the change in distribution is complicated because the new slices were only one hit thick. As the adrenal tropism was the most pronounced, the organs were divided into two groups of adrenal hits and nonadrenal hits and chi squared analysis rejected the null hypotheses for both the parental and the shScr expected distributions with p values of 0.0008 and 0.002 respectively.

**Table IV.III Latency and penetrance of bioluminescent tumors after injection of 1 million 361-luc cells into the left ventricle of NOD/SCID mice.** The organs colonized and the number of days post-injection (PI) when the signal was first detected are listed.

Organ	Earliest Onset (Days PI)	Mean Latency (Days PI)	Penetrance
pLKO (N=13)			
Adrenals	21	51	77%
Brain	48	58	31%
Ovary	33	33	8%
pHAGE (N=14)			
Adrenals	23	42	76%
Brain	65	86	76%
Ovary	46	98	14%
shScr (N=15)			
Adrenals	20	37	93%
Brain	42	83	33%
Ovary	N/A	N/A	0%

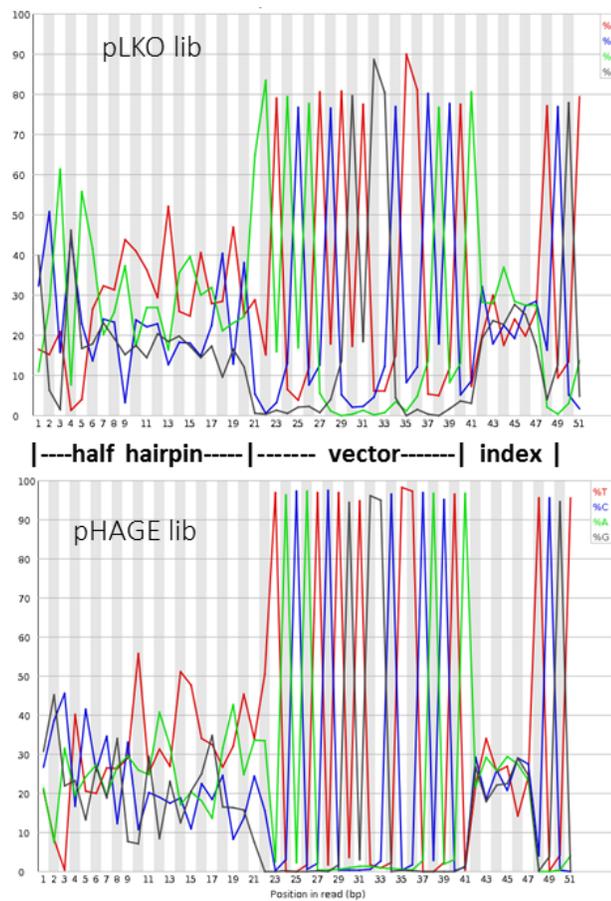
To test if chromatin modifiers could be involved in the awakening of dormant solitary or micro metastatic colonies I used RNAi hairpins to knock down all known human chromatin modifiers. Several organs were colonized, but aside from the dominant adrenal and secondary brain and ovarian tropisms, only single tumors at any one organ site were colonized more frequently in the library animals. When the tumor distributions were divided into nonadrenal and adrenal groups there was a statistically significant difference in the two samples. This suggested the hits might be meaningful so the hairpins were next recovered from the tumor genomic DNA to identify any emerging trends related to tropism.



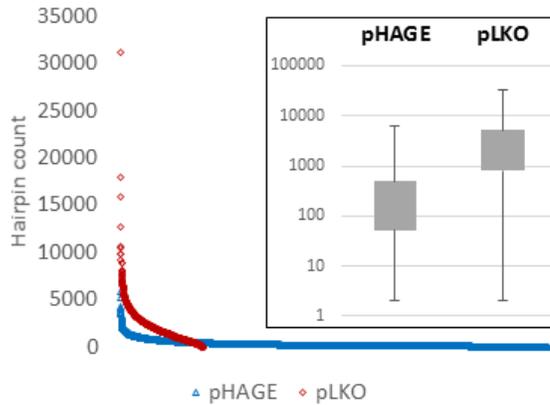
**Figure IV.8 Bioluminescent signal detected per organ.** A) Pie charts displaying the number of organs with signal. N listed is number of mice. B) Percentage of animals with growths in the adrenal gland, brain and ovary for each library group.

### IV.3.1 Bioinformatic analysis

Next generation sequencing is dependent upon being able to define distinct clusters of amplicons on the flow cell to accurately call each base added on in one round of synthesis. Cluster positions are determined in the first few base reads so it is essential that there be a good amount of nucleotide diversity immediately adjacent to the sequencing primer. The amplicon containing my hairpins was mostly identical vector sequence so a custom sequencing primer which sat down exactly at the start of the hairpin was used. This ensured maximum diversity in the first few bases and allowed read through the half hairpin and the bar code.



**Figure IV.9 Per base sequence content of library reads.** Both libraries show maximum sequence diversity in the first bases representing the hairpin followed by the identical miR30 adapter sequence and then the index.



**Figure IV.10: Distribution of pre-screen pellet hairpins.** The line graph shows a distribution of the count for each hairpin present before injection. The inset shows a tenfold difference in scale between the two libraries. The pLKO spread was skewed towards the larger count hairpins whereas the pHAGE distribution was flatter. The top twenty hairpins are listed in the table.

pHAGE		pLKO	
hairpin	count	hairpin	count
HDAC5_4949	5878	MYST2_4	31144
SUV420H1_2193	5185	PRDM4_2	17947
PRMT8_324	4304	WHSC1L1_5	15818
MST1_2283	4031	RPA3_3	12738
SETD1A_1749	3965	WDR82_2	10553
SETD3_1209	3682	SMYD2_2	10459
KDM5B_6116	3566	MSL3_4	9884
MYST2_2861	3556	JMJD5_5	9786
SMARCAD1_4050	3409	BMI1_5	9258
KDM5B_1533	3335	PRDM10_1	8944
SETD3_1179	3087	METTL8_4	8797
DEFA1_446	3038	FKBP2_3	8106
SETD1A_1900	2978	MYST4_4	8104
PRDM8_59	2916	MGMT_2	8025
SETD1A_2544	2896	SMYD1_1	7876
TET3_6342	2871	KDM1A_3	7759
BMI1_737	2848	KAT2B_1	7643
SIRT2_1017	2646	SMARCC2_1	7603
WHSC1L1_738	2595	KDM5D_3	7465
EZH1_1047	2485	DNMT1_5	7289

For the pHAGE pellets there were nearly 93 million reads and over 196 million for pLKO. The quality of the reads was judged by several criteria including read length uniformity, and completeness of the hairpin, the barcode, and a 20 base pair adapter region between the hairpin and the barcode. Figure IV.9 shows that per base sequence content varied exactly as expected with variability only at the hairpin and barcode regions. This indicates the custom sequencing scheme worked well. Reads with a single base difference in length, a single mismatch in the adapter or hairpin or a single ambiguous base were all discarded. 19% and 5% of the pHAGE and pLKO reads were eliminated in this manner.

The pre-screen pellets were each assigned a barcode and their reads were similarly aligned to the reference database. There were 1.8 million reads for pHAGE and 4.2 million

for pLKO. Figure IV.10 shows the distribution of hits per hairpin for each library. All of the genes were represented and the curve of the graph for both showed a negative exponential distribution with a small portion of the hairpins overrepresented. The pHAGE range crossed four orders of magnitude from one to 5,878 hits with a flatter distribution than pLKO which varied across five logs from 31,144 hits per hairpin to one. pLKO was skewed more towards the higher count hairpins. The twenty hairpins with the highest representation are listed in Figure IV.10 and the box plot shows the spread of the hairpins before being injected into mice.

In the pHAGE bioinformatics analysis, a non-tumor negative control amplicon from NOD/SCID genomic DNA was included and a hairpin sequence targeting firefly luciferase was placed in the reference library to judge noise. Hits were determined at the tumor level and only reads greater than 1% of the most frequent read within a tumor were considered. The firefly hairpin count, if present in a tumor, was also used as a low threshold. The only hit which aligned to the NOD/SCID negative control amplicon was the firefly targeting hairpin indicating negligible noise when aligning to the reference database.

**Table IV.IV Tumor and hit measurements for pLKO and pHAGE grouped by nonadrenal vs adrenal.**

	pLKO		pHAGE	
	Adrenal	Nonadrenal	Adrenal	Nonadrenal
Total Hits	242	139	235	54
Tumors	20	18	16	16
Hit mean	12.1	7.7	14.7	3.4
Hit minimum	1	1	4	1
Hit maximum	67	46	37	7
Monoclonal tumors	5%	33%	0%	25%

The number of hits per tumor varied from one to thirty-seven implying that some of the tumors were founded by more than one cell as the likelihood of thirty-seven separate transduction events occurring within an experiment with a functionally measured MOI of 0.3 seems improbable. As table IV.IV shows, when the hits were broken out into adrenal and nonadrenal groups the adrenal tumors showed a mean clonality of 14.7 with zero monoclonal tumors. The nonadrenal tumors ranged from one to seven hits per tumor averaging 3.4 hits. Of the sixteen pHAGE nonadrenal tumors with valid hits, 25% of them were monoclonal. Without knowing the hairpin targets it could already be seen that there were differences in the colonized sites based on clonality. The adrenal tumors had many more hairpins and therefore most likely more founding cells than the other sites.

#### **IV.4 Nonadrenal tumor analysis reveals targeting of EHMT1, JMJD5, PRDM14 and SIRT3**

Using the above thresholds, there were four gene targets enriched in the nonadrenal tumors: EHMT1, JMJD5, PRDM14 and SIRT3. Table IV.V shows how many hairpins were present in how many tumors and animals for each hit. Three of these genes also had adrenal hits. The proportion of the hits that were nonadrenal varied from fifty to one hundred percent.

**Table IV.V: pHAGE hairpin enrichment analysis of non-adrenal vs adrenal tumors.**

Gene	Non-adrenal tumors				Adrenal tumors				%Non-adrenal
	Hits	Tumors	Animals	Hairpins	Hits	Tumors	Animals	Hairpins	
EHMT1	4	3	3	3	2	2	2	2	67
JMJD5	7	5	5	2	3	1	1	3	70
PRDM14	4	2	2	2	0	0	0	0	100
SIRT3	4	4	1	3	4	4	4	4	50

In a parallel analysis of the pLKO tumors, the non-tumor NOD/SCID negative control amplicon unexpectedly aligned with 98 reference hairpins with counts ranging from one to 936. Concernedly, the top three hairpins in the NOD/SCID control were also the top three in the pre-screen pellet. A scramble hairpin tumor that was included in the sequencing run also had six hits that matched the reference database when it was expected to have had zero. Four of these six hits were hairpins listed at the high end of the pre-screen pellet distribution in Figure IV.10. The control samples and hairpins indicated the signal to noise was lower in the pLKO library and suggested that interpretation of these results would not be straight forward.

As shown in Table IV.IV the twenty adrenal tumors in the pLKO library also contained a large number of clones with one monoclonal tumor, and an average of 12.1 hits per tumor ranging up to 67. The hairpins present in the monoclonal tumors for both library are listed in Table IV.VII. The nonadrenal tumors had a mean of 7.7 hits per tumor with a range of one to 46. There were six monoclonal out of 18 total tumors. Table Segregation into adrenal and nonadrenal groups found 24 genes to be targeted in nonadrenal tumors. All 24 of these genes were also targeted in the adrenal tumors to varying degrees. The thirteen of these genes which were found as nonadrenal hits more frequently than adrenal hits are listed in table IV.VI. Normalization across the five log distribution present in the pLKO pre-screen pellet was deemed not to be useful for these hits. Instead a subjective measure of bias is included in table IV.VI where the rightmost column is the pre-screen rank of the gene. None of these genes were found in the pHAGE results. The four genes enriched in the pHAGE analysis were used for follow up analyses as the data from this screen were cleaner and interpretation was direct.

To examine the clinical relevance of loss of expression of these genes in breast cancer a Kaplan Meier analysis was performed using the publicly available kmplot online tool to interrogate microarray data for 4,142 breast cancer patients. Figure IV.11 shows relapse free survival probability was significantly lower for patients with low expression of any of these genes. Focusing in on the 1,166 tumors with the same luminal B molecular subtype as the 361-luc line also revealed a poor prognosis for all four genes. But when distant metastasis free survival was considered, only SIRT3 loss was significant. SIRT3 expression showed a significant correlation with survival for distant metastasis free survival and post-progression survival.

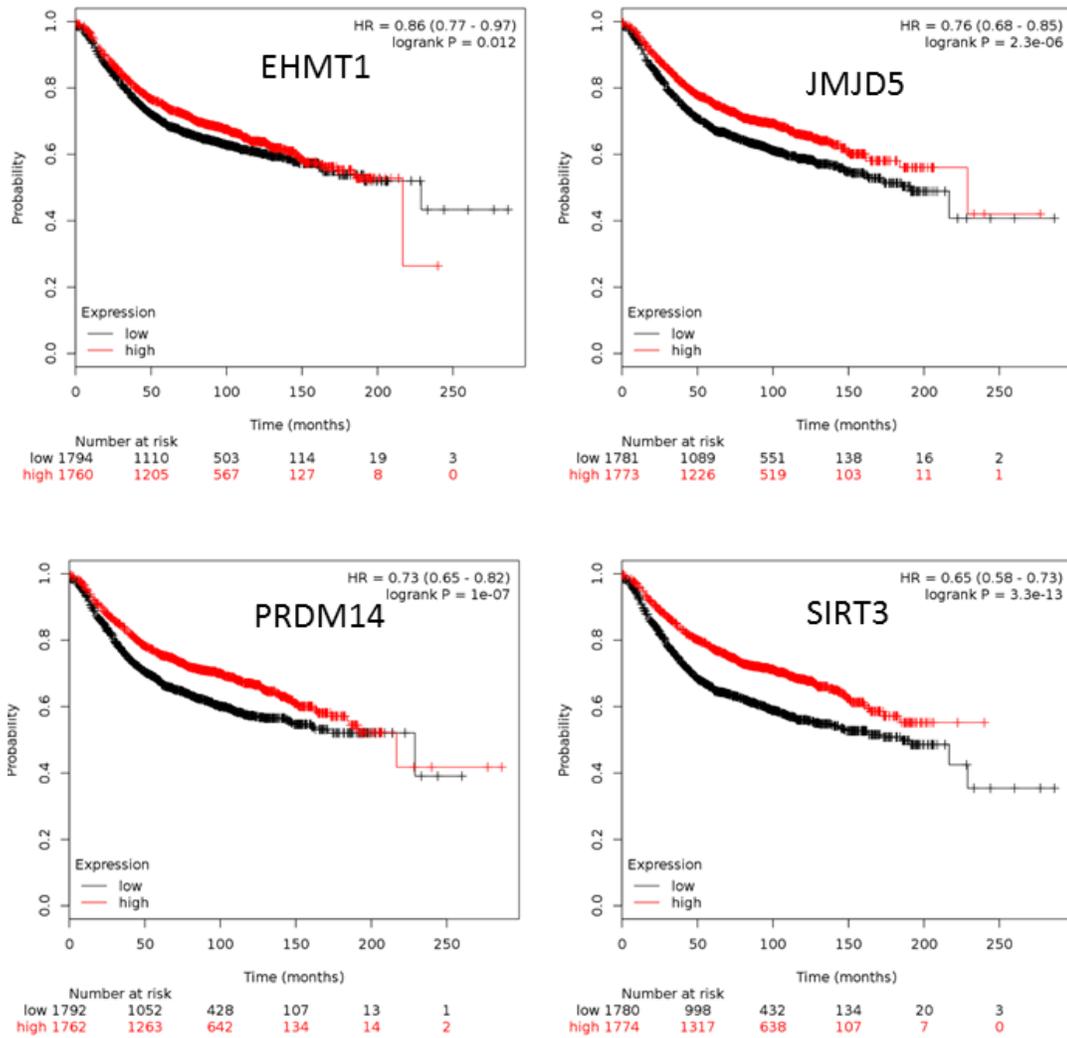
**Table IV.VI pLKO hairpin enrichment analysis of non-adrenal vs adrenal tumors.**

Gene	Nonadrenal hits	Adrenal hits	Nonadrenal %	Pre-screen rank
KDM5C	5	1	83%	163
PCNA	4	1	80%	62
CHAF1B	3	1	75%	73
HDAC11	3	1	75%	16
USP22	3	1	75%	177
SMYD2	5	2	71%	29
JMJD6	6	3	67%	53
GTF3C4	4	2	67%	81
MLL3	4	2	67%	182
BRAF	3	2	60%	98
CHD6	3	2	60%	130
EDF1	3	2	60%	104
KDM1A	3	2	60%	24

**Table IV.VII Hairpins present in monoclonal tumors.**

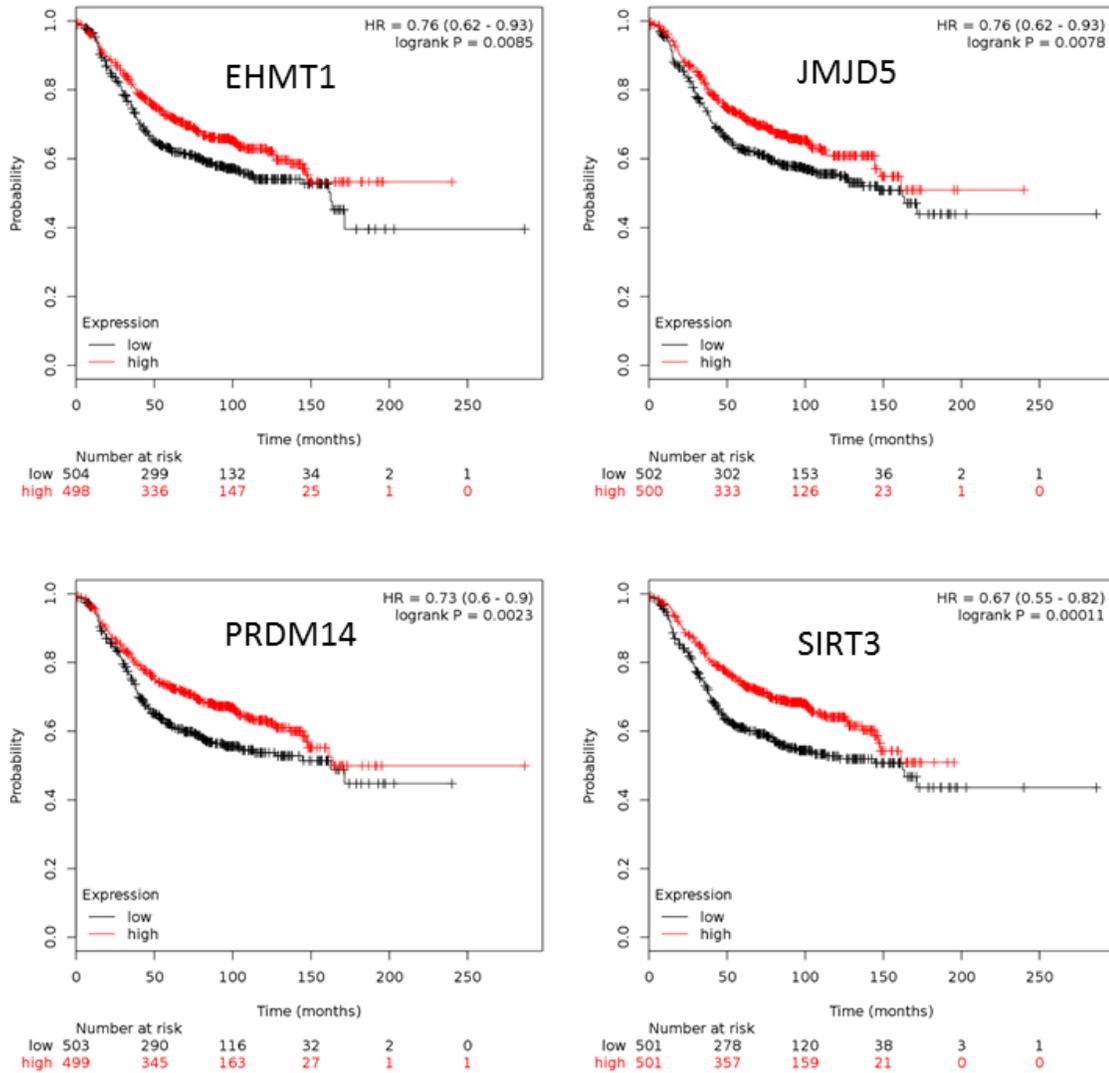
Library	Hairpin	Hits	Mouse	Tissue
pHAGE	SIRT3_2294	1779930	46	brain
pHAGE	SIRT3_2294	559332	46	pituitary
pHAGE	JMJD5_1945	1953622	47	pituitary
pHAGE	KAT2A_1866	1567543	8	pituitary
pLKO	PCGF2_3	600371	46	left adrenal
pLKO	SMARCC2_1	4309082	46	brain
pLKO	MYST2_4	5153407	47	brain1
pLKO	METTL8_3	3969243	47	brain2
pLKO	MEN1_4	949509	5	brain
pLKO	UBE2B_5	4838313	6	brain
pLKO	WHSC1L1_5	4717195	8	ovary

### Relapse free survival (N=4,142)

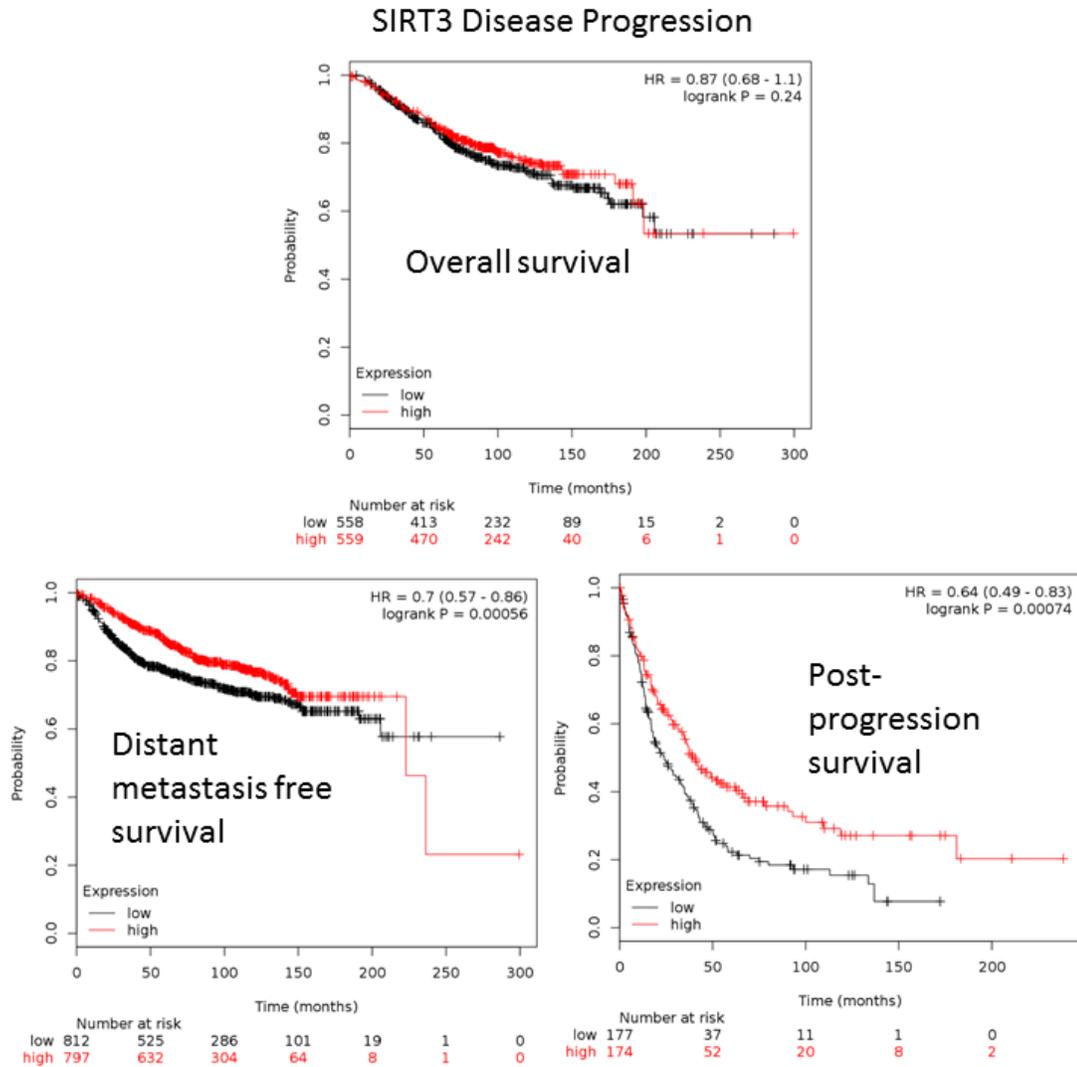


**Figure IV.11 Probability of relapse free survival of breast cancer patients.** Patients with all subtypes were grouped as high or low expressing for each gene and probability of relapse free survival over time was plotted. Low expressing patients for all four genes were significantly less likely to survive without relapse. [kmplot.com](http://kmplot.com)

### Relapse free survival luminal B type (N=1,166)



**Figure IV.12** Probability of relapse free survival of luminal B breast cancer patients. Patients were grouped as high or low expressing for each gene and probability of relapse free survival over time was plotted. Low expressing luminal B patients for all four genes were significantly less likely to survive without relapse. [kmpplot.com](http://kmpplot.com)



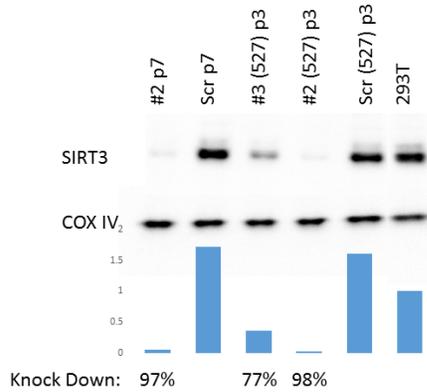
**Figure IV.13 SIRT3 expression and probability of clinical disease progression in breast cancer patients.** Patients were grouped as high or low expressing for SIRT3 and probability of overall survival, distant metastasis free survival and post-progression survival over time was plotted. Low SIRT3 expressing patients were significantly less likely to survive metastasis free and post-progression. [kmplot.com](http://kmplot.com)

Put together these data suggest that all four of these hits could play a role in human disease, especially in luminal B tumors, and SIRT3 loss in particular showed an effect in three out of four clinical measures of metastatic progression. The next step was to experimentally validate the SIRT3 hit.

#### IV.4.1 Validation of the SIRT3 hit

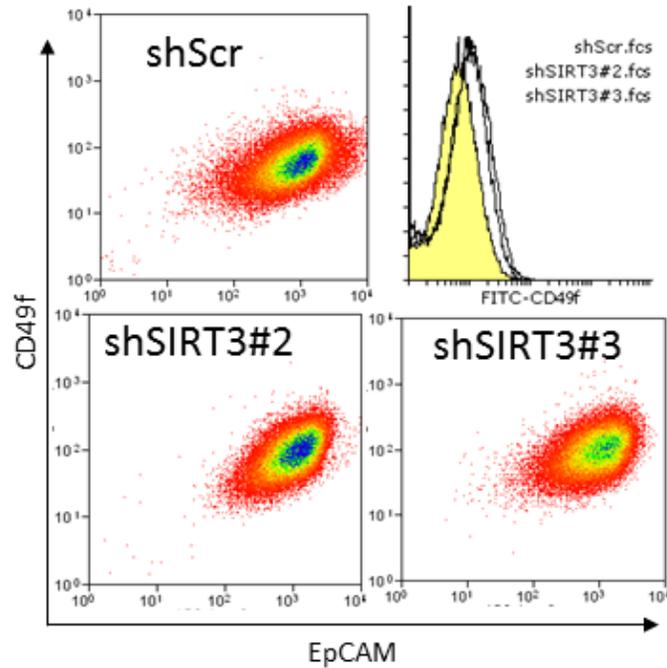
SIRT3 is a mitochondrial localized deacetylase that is required for maintenance of mitochondrial integrity (Kim et al., 2010). There are reports of it having a tumor suppressive function and experimental knock down leads to high levels of reactive oxygen species (ROS), genomic instability and increased HIF-1 $\alpha$  signaling (Alhazzazi et al., 2013; Finley et al., 2011). SIRT3 loss contributes to metabolic reprogramming which can drive a cancer cell into a more stem-like state. Acquisition of stem cell characteristics is a known contributor to metastatic progression so I investigated if SIRT3 knockdown would result in changes of mammary epithelial progenitor markers in 361-luc cells.

SIRT3 knock down was measured at the message and protein level and was shown to be stable across multiple passages *in vitro*. Five hairpins were tested and the two with the greatest knockdown, #2 at 98% and #3 at 77% (Figure IV.14), were stained with antibodies for the mammary epithelial marker EpCAM and the progenitor markers CD49f and CD24 in Figure IV.15. EpCAM staining was nearly absolute regardless of SIRT3 expression confirming the luminal nature of the 361-line. CD49f expression did increase slightly in the SIRT3 knockdown cells, suggesting there could be a shift from a mature luminal state to a luminal progenitor state. However, the shift was modest and no clear change in the overall FACS defined populations was evident.

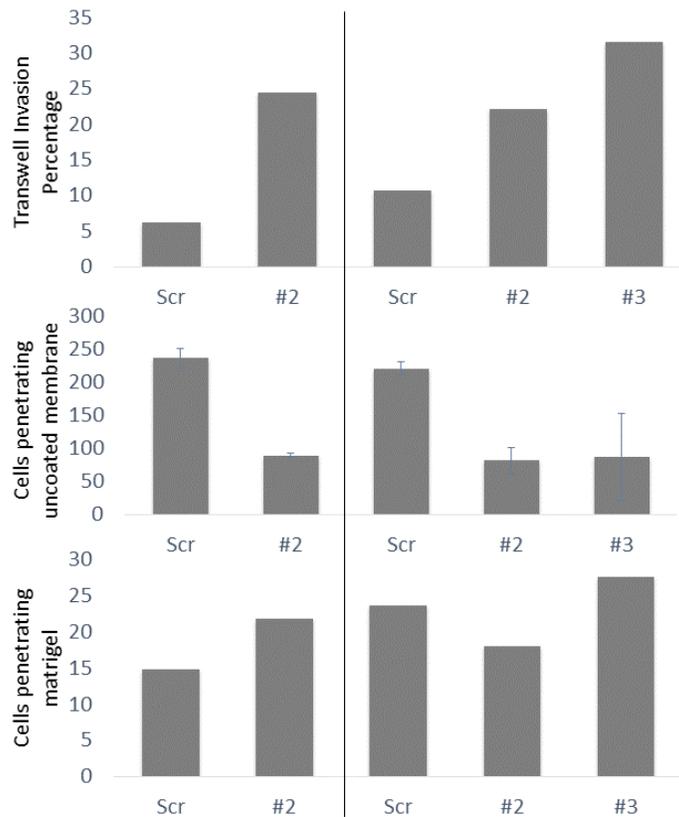


**Figure IV.14 Knock down of SIRT3 in 361-luc cells.** The two hairpins with the biggest decrease in SIRT3 mRNA were validated by western blot. Hairpin #2 and #3 exhibited greater than 95% and 75% knockdown respectively. Cells at passage 3 and passage 7 were tested to ensure durable knock down.

The tumor sphere assay is another way to measure stem cell activity. Cells are plated sparsely in low adhesion conditions and those cells capable of self-renewal proliferate to form unattached monoclonal spheres. In Figure IV.17B it can be seen that knockdown of SIRT3 did not change the tumor sphere forming ability of 361-luc cells regardless of the size cut-off used to define a sphere. Testing the hypothesis that SIRT3 loss led to changes in stemness did not yield a clear answer with these assays. In addition to stemness other aspects of metastasis were considered.

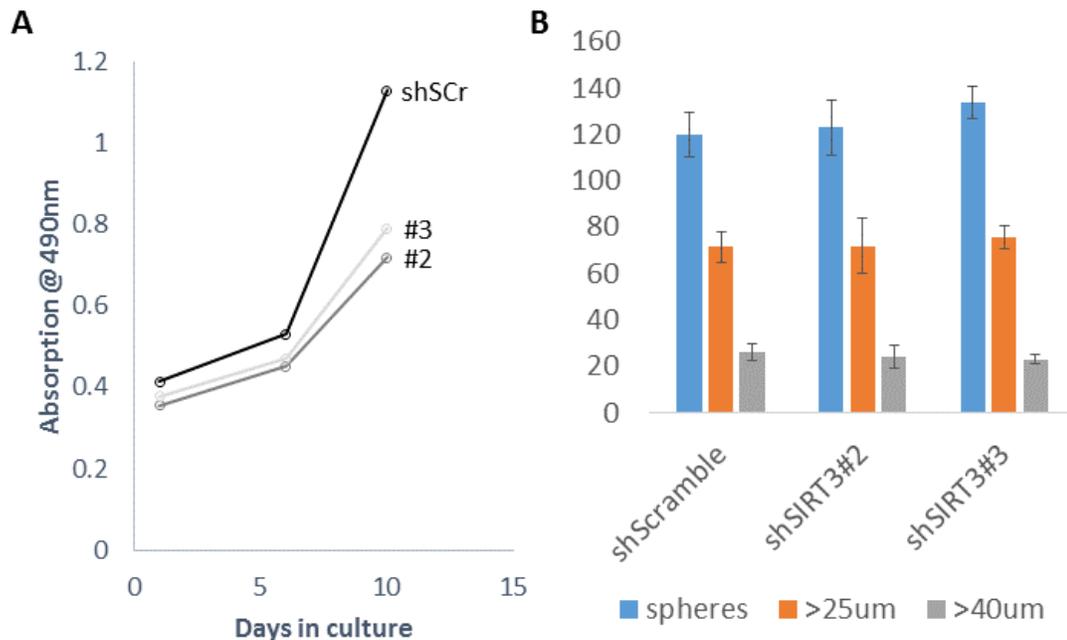


**Figure IV.15 FACS analysis of CD49f and EpCAM mammary progenitor markers on shSIRT3 361-luc cells.** CD49f levels are higher by FACS staining for two SIRT3 hairpins.



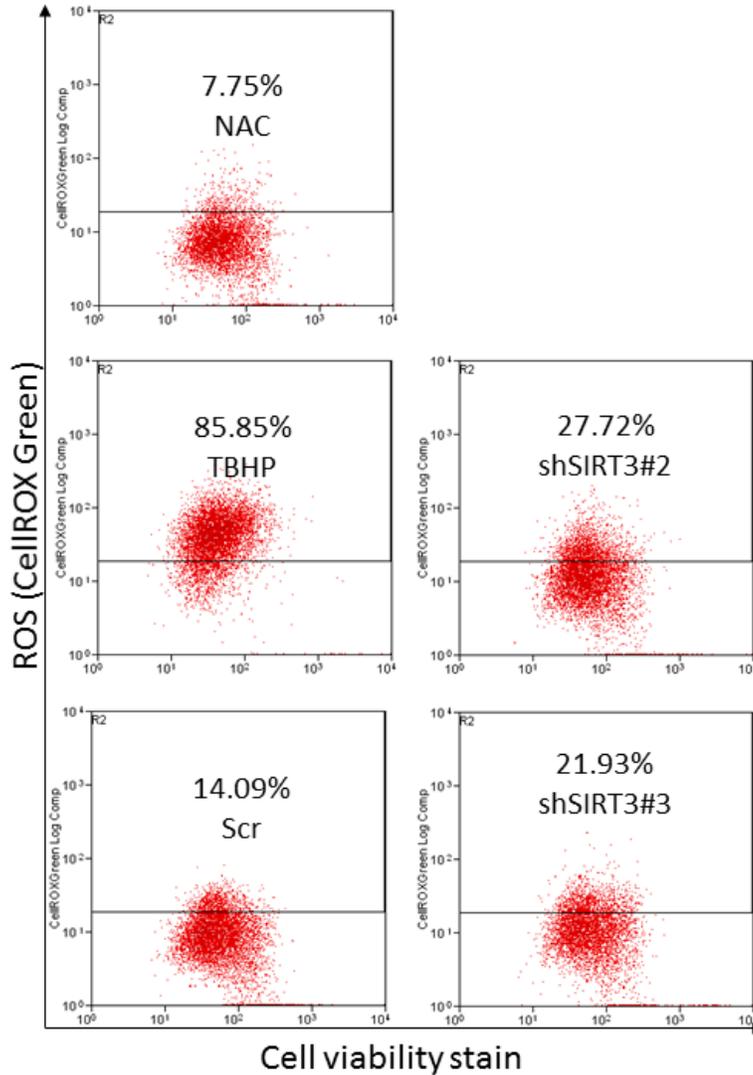
**Figure IV.16 Transwell invasion by shSIRT3 361-luc cells.** The top graph shows an increase in invasion by shSIRT3 cells. This is a compound measurement taken by dividing the number of Matrigel penetrants (bottom graph in absolute numbers) by the uncoated membrane penetrants (middle graph). The bar delineates two separate experiments.

To measure changes in invasion and motility due to SIRT3 loss the cells were subjected to a transwell assay using a layer of Matrigel as a surrogate for an intact basement membrane. Optimal conditions for the 361-luc line were established by varying the density of cells plated and the length of time allowed to traverse the transwell. 100,000 cells were plated on top of an 8µm porous transwell with and without Matrigel. Cells that migrated to the other side after 24 hours were then counted. Invasion was measured by dividing the absolute number of cells per well able to penetrate the Matrigel by those penetrating a separate uncoated well. Interestingly, the number of cells able to cross the uncoated transwell was lower in the SIRT3 knockdown cells compared to the shScr cells. But there was no measurable difference in the number of cells crossing the Matrigel barrier. Thus the SIRT3 knock down cells were less motile, but more invasive. A cell viability assay to measure proliferation of the shSIRT3 cells showed that both hairpins grew more slowly than the scramble control cells (Figure IV.17A).



**Figure IV.17 Viability and stemness of shSIRT3 361-luc.** A) Viability assay (MTS) for 361-luc cells shows decreased proliferation. B) Tumor sphere assay shows no differences in sphere forming between shScr and shSIRT3

As SIRT3 knockdown is reported to lead to increased ROS the CellROX Green reagent which is non-fluorescent when reduced but stains strongly green when oxidized was used in Figure IV.18 to test ROS levels in the knockdowns.

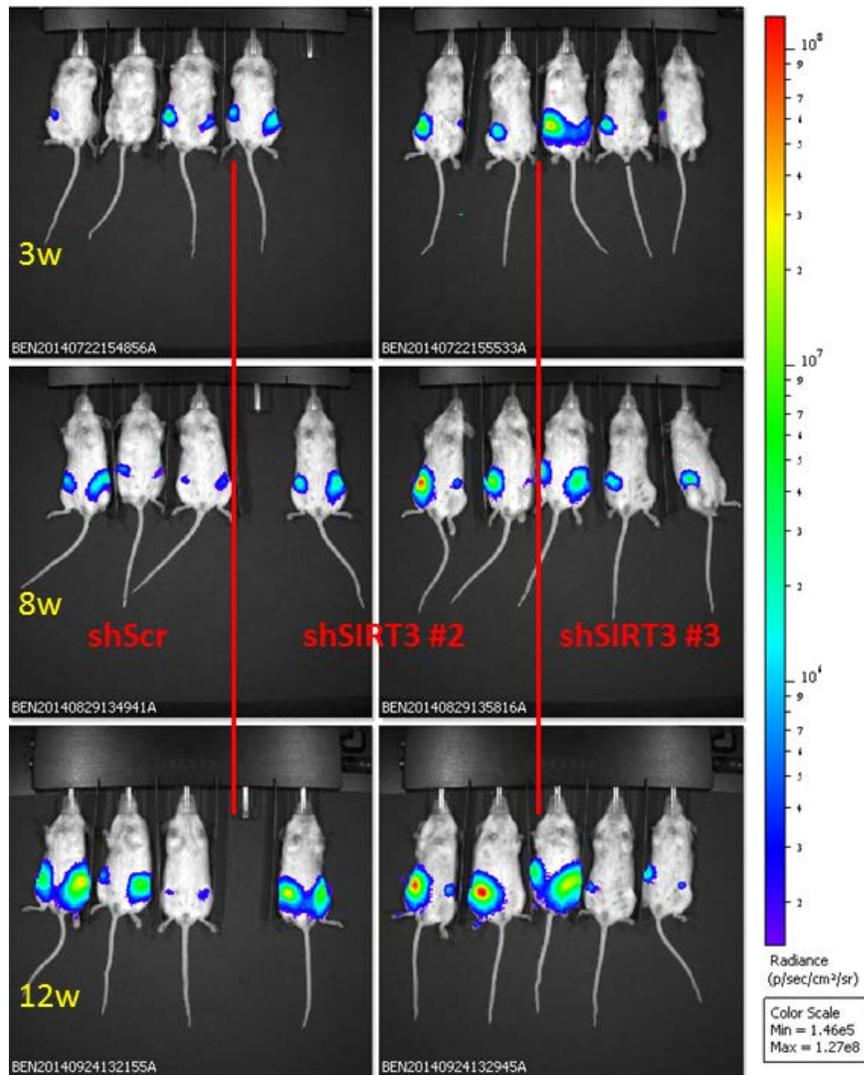


**Figure IV.18 ROS measurement of shSIRT3 knock down 361-luc cells.** The left plots are controls for reducing (NAC) and oxidizing (TBHP) conditions and the non-targeting shScr hairpin. The right plots are the SIRT3 targeting hairpins. SIRT3 knock down with either hairpin increased ROS levels.

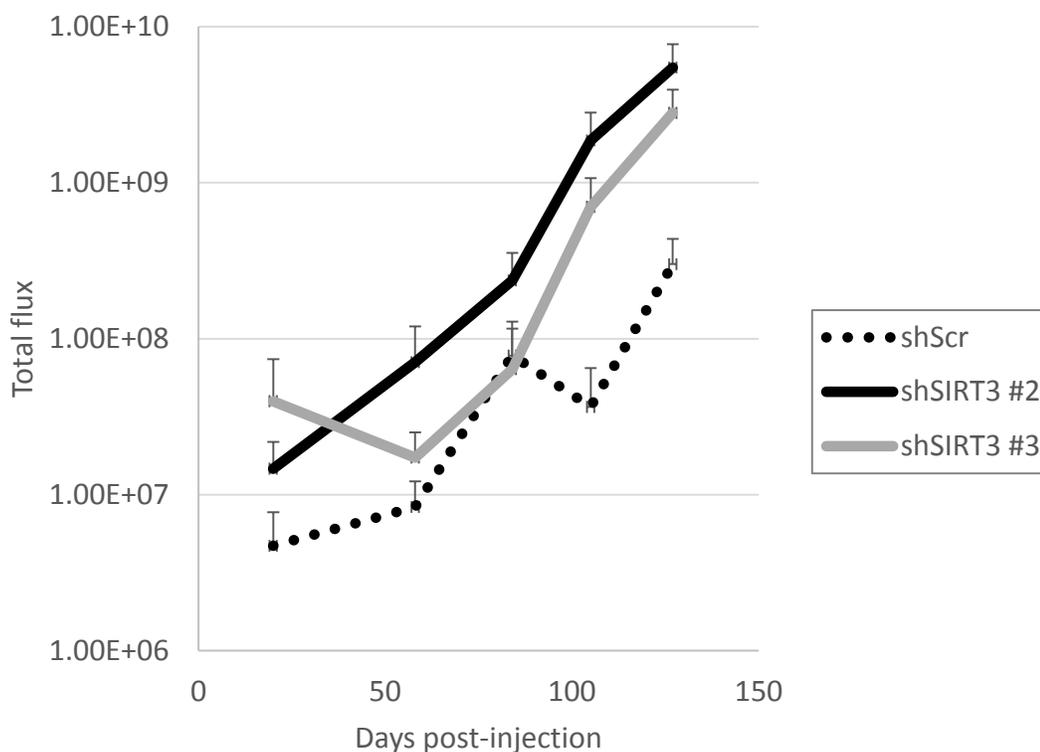
Under standard culture conditions, a positive control treated with tert-butyl hydro peroxide (TBHP) demonstrated increased fluorescence and pre-treatment with the reducing agent

N-acetyl-cysteine attenuated the ROS signal. SIRT3 knock down increased cells staining for ROS by 2 and 1.5 fold in hairpins 2 and 3 compared the shScr control.IV.4.2 SIRT3 knockdown *in vivo*

The *in vitro* assays were inconclusive and it was possible the SIRT3 loss phenotype may only be detectable *in vivo*. To determine the effect of SIRT3 on local invasion *in vivo* shSIRT3 cells and shScr cells were injected orthoptically into the mammary fat pads of NOD/SCID mice. Three mice were used with injections into the 4<sup>th</sup> pair of mammary glands on both sides (Fig. IV.19). Tumors grew out in seventeen of the eighteen injection sites. There was no spontaneous metastasis to any secondary site detected. The tumor growth difference between the shScr and shSIRT3 as judged by mean bioluminescent total flux per tumor was not significantly different at any time point over a period of 18 weeks (Figure IV.20).



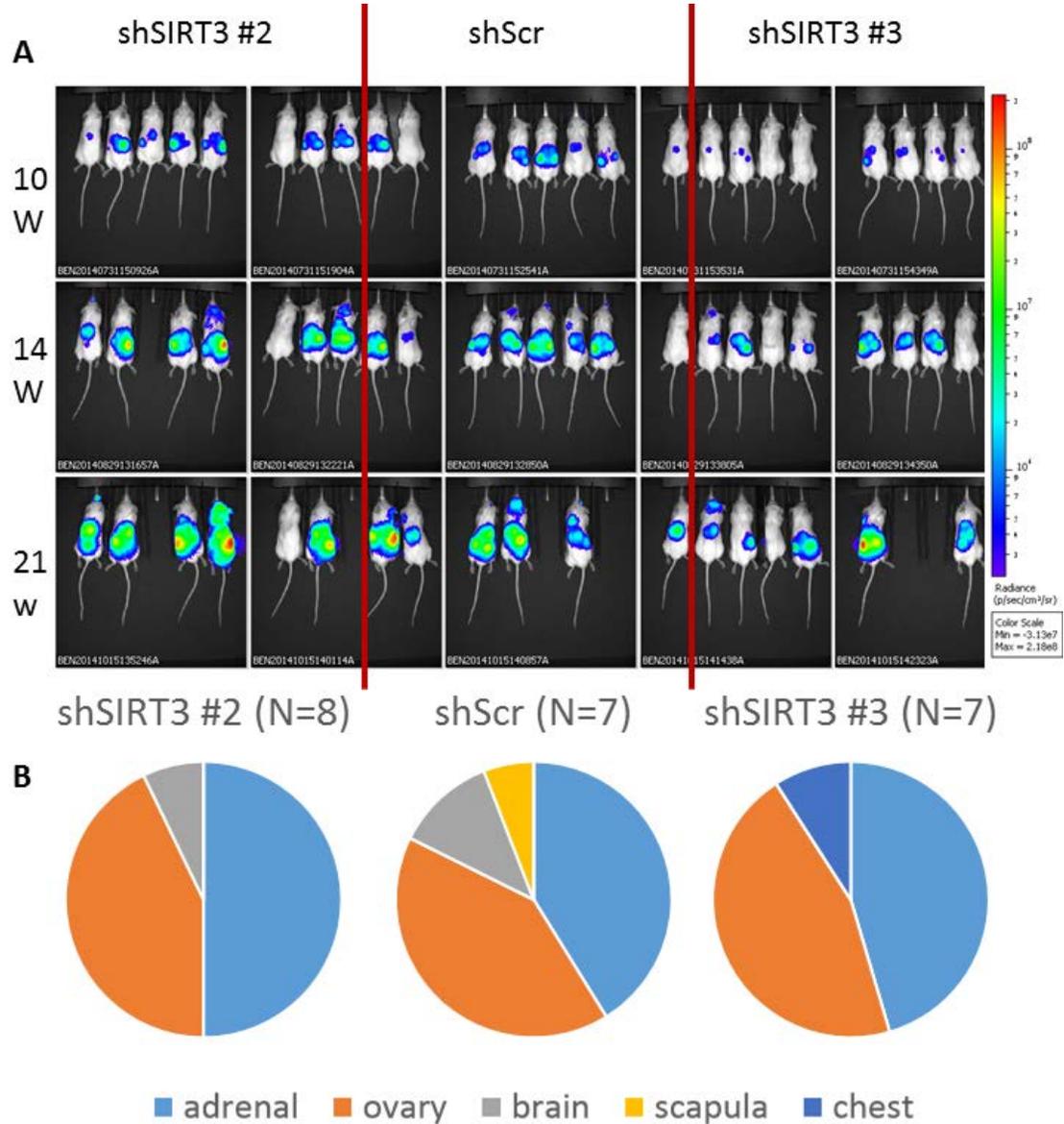
**Figure IV.19 Orthotopic injection of shSIRT3 361-luc cells.** 1 million cells in 50% Matrigel were injected into the mammary fat pads of 3 animals per hairpin.



**Figure IV.20 Mean total flux per hairpin group.** The mean total flux (p/s) for six tumors per hairpin over 20 weeks is presented. There was no statistical difference at any time point.

To investigate a role in colonization eight animals per group underwent UGIC of 1 million cells. Figure IV.21 shows signal was detected in multiple organs for all three groups. Two mice with early thoracic signal were excluded from the analysis. The original screen showed that the organs colonized were more diverse in the library groups than in the control. However, the dominant tropism was the adrenal gland followed by the brain. Adrenal tropism was evident again as expected in the shSIRT3 follow up experiments, but surprisingly the ovaries were the second most colonized organ in these experiments and brain growths were seen in only two animals. There was no statistical difference between the colonized organ distributions between any of the three groups. Variation in the location of the tumors, specifically the amount of tissue that light would need to penetrate, made a luciferase based analysis of tumor growth rate invalid as photon flux is not proportionate

to tumor mass or volume in this case. *In vivo* experimentation was unable to confirm a change in outgrowth or organ tropism with SIRT3 loss.



**Figure IV.21 Colonization by shSIRT3 361-luc cells following UGIC.** A) Tumors grew out largely in the adrenal glands and ovary regardless of hairpin. B) Pie charts showing all of the organs colonized in each group.

# Chapter V: Discussion

## **V.1 Conclusions and significance of findings**

### **V.1.1 361-luc system to model metastasis**

Metastatic colonization can be broken up into three phases: 1) Escape, which includes survival in the circulation, 2) Deposition or arrival at a foreign site and 3) Adaptation. In all three of these phases, the tumor cell is not a lone player and its role can vary from passive to active. It seems most probable that there are active and passive parts played in each phase. A thorough analysis of these phases as they relate to the system described herein is useful to discuss my findings and their merits.

Escape in the 361-luc system is an entirely passive process insofar as the tumor line is dissociated by hand by trypsinization from a plate and the cells are injected directly into the animal. However the durability of the tumor cells and potential survival signaling in transit are more active properties also at play. In this phase the host environment is the circulation which includes shear forces and deformation of the cell as it encounters artery walls and decreasing diameter passages. There may be chemical messengers in the blood to which the tumor cell may be able to respond. One could even speculate that active communication and exchange of genetic information is possible via blood borne nucleic acids or exosomes.

The most passive outcome in the deposition phase is a cell simply getting stuck somewhere like a capillary bed which is too narrow for the cell to navigate. Cancer cells are known to change their integrin expression and cellular adhesion molecules demonstrating a more active role in this phase. Depending on the receptor profile of a circulating tumor cell it may be able to respond to cytokines in a homing process that

contributes to the ultimate destination and may as well provide survival, proliferation and angiogenic functions as is the case with the stromal derived factor 1 chemokine and its CXCR4 receptor (Teicher and Fricker, 2010). Studies have not yet been published able to measure the extremes of the wanderlust of an individual metastatic cell but the example of melanoma which may engage a latent neural crest program to travel from the skin of a distal appendage all the way to the brain is striking. Upon autopsy, melanoma metastases to the brain are found in 75% of patients with metastatic progression (Sampson et al., 1998).

Personally adaptation is the most interesting phase to me because it seems the cell is forced to play an active role or perish, the figurative Hunger Games of tumor progression, or perhaps the Thunderdome for those of us non-millennials (Collins, 2008; Miller and Ogilvie, 1985). But this active role assumedly must be reciprocated in some regard by the cells of the host environment in many cancers that lay dormant for a long period of time and then suddenly switch to a proliferative state. One of the most exciting fields of research in metastasis, and malignancy in general, is the role of the microenvironment and the immune system. Having described the phases we can now use them as a framework for the organ specific colonization of 361-luc cells cataloged here.

An important aspect of deposition which could easily be overlooked when modeling the disease in mice is the natural circulatory flow downstream of the primary tumor. In breast cancer there are no arteries leaving the breast to supply other organs and tumor cells nearly always habituate local sentinel lymph nodes. These nodes and breast capillaries eventually drain into veins where the fluid rejoins the systemic circulation passing through the right ventricle and pulmonary vessels. Thus a cell that escapes the primary tumor could passively be lodged in the sentinel lymph nodes or the lungs. In the

361-luc system direct access to all organs is provided, bypassing lymphatics entirely. This is effectively placing the tumor cells into capillaries throughout the body. This feature of the system is useful for examining mechanisms for specific organ colonization which might not be possible using a spontaneous metastasis model.

### **V.1.2 361-luc cells exhibit organ specific colonization**

As 361-luc cells grow out mainly in the adrenal glands, the brain and the ovary one must consider the circulatory hurdles necessary to relocate to each of these organs. An invariable blood supply of nutrients is essential particularly to the brain because of extremely limited fuel stores (Berg et al., 2002). But many non-glucose elements in the blood have high potential for neurotoxicity. To prevent this, capillaries in the brain have astrocyte processes that cooperate with pericytes and endothelial cells to establish unusually tight junctions (Figure I.7). *Ex vivo* brain imaging in my chromatin screen, and histologic examination of the 361-luc brain tumors demonstrated foci that were located at the distal end of blood micro vessels in the cerebrum. Notably the second most common site of brain colonization was the pituitary gland which is remarkable here not only because of its well-known endocrine function but also because it exhibits faster blood flow than the rest of the CNS and contains fenestrated capillaries where the blood brain barrier is not intact (Mullier et al., 2010). While there are not lymph nodes present directly in the brain, there are recently discovered lymphatics that drain into the jugular vein (Louveau et al., 2015). Of the three organs most readily colonized by 361-luc, the brain has the highest barrier to entry and most likely demands an active role for the cell. 361-luc cells are capable of overcoming this highest burden exemplified by tumor presence in regions of the brain where the blood brain barrier is fully functional.

The ovary shares a common blood supply with the uterus and has cycles of angiogenesis and regression corresponding to follicular demands and the reproductive cycling of the organ. It has established permanent capillaries of the traditional type but the capillaries supporting developing follicles are dynamic, undergoing budding, sprouting, elongation and sinusoid formation (Macchiarelli et al., 2006). The nulliparous ovary thus could provide easy access to circulating tumor cells. It is not clear exactly where in the ovary 361-luc metastases originate as the tumors were very large when examined in my work and had obliterated the ovarian tissue. They were found juxtaposed to developing follicles and intact follicles were observed surrounded by tumor cells (Figure II.15E & G).

For the adrenal glands with open sinusoidal micro vessels, the cells may very well be pushed directly into the cortex and medulla without any active cell intrinsic process. This gland has the lowest entry burden via arterial blood. However, the liver also has a high rate of blood flow and sinusoidal capillaries so assumedly just as many cells, possibly more due to its larger size, could get stuck here. But liver metastases were not detected in any animals. So this appears to be a dead end for metastatic progression for these cells while the adrenal gland is not.

In fact, the most striking feature of the 361-luc system described in this work is its propensity to colonize the adrenal gland. After deposition, the data presented here suggest that these tumors originate in the adrenal medulla made up largely of chromaffin cells which are modified postganglionic neurons.

### **V.1.3 361-luc cells have an inherent neurotropism**

There are several other pieces of evidence which tempt speculation that the adaptation phase of 361-luc adrenal colonization involves an inherent neurotropism. First, the 361-luc line originated from a breast cancer that metastasized to the brain. So before it became a tumor cell line it spent some time in the brain and most likely adapted in some ways to the CNS environment as all metastases must do in order to grow in foreign soil. It is unclear if this was through new genetic gains in function or the utilization of the primary tumor's existing repertoire. Second, in the original orthotopic injections of this line done by Ina Klebba in the lab which were photographed by Charlotte Kuperwasser, tumor cells were found growing along nerve fibers (Figure III.4). Third, this line also colonizes the brain and regularly was found in multiple foci there. Fourth, there is a recent PNAS report of brain localized Her2 positive breast carcinomas adapting to the central nervous system environment by metabolizing the neurotransmitter  $\gamma$ -aminobutyric acid (GABA) as a NADH source (Neman et al., 2014). Critically, the paper is preliminary and compared patient cells to only one token reference tumor cell, nevertheless this is evidence that CNS adaptation does occur in brain metastases of breast cancer. As a whole these observations support the hypothesis that 361-luc cells may be somehow attracted to neural tissues.

However, 361-luc cells are also ER positive cells and the adrenal cortex can make weak androgens like DHEA and androstenedione. Figures III.8 and III.12 demonstrate that ER staining is maintained by these cells in tumors. What is not known is if the small amount of androgens made by the adrenal glands could influence the growth of nearby 361-luc cells which exited the sinusoidal capillaries and became embedded in the gland.

Some breast cancers are known to express aromatases and could possibly convert androgens into estradiol to fuel growth. There is one report looking at aromatase expression in 361 cells which shows they do not express the enzymes to convert DHEA and androstenedione into estradiol but the question is not settled as this same report shows expression of PR which we did not see by ICC or IHC (Hevir et al., 2011).

As the ease of distribution to the adrenal gland by the circulation alone is not sufficient to explain why 361-luc cells grow out preferentially in the adrenal gland and ER ligands are not likely involved, I performed an *in vitro* viability screen treating the cells with adrenal hormones. I reasoned that whatever mechanism was at play in the observed adrenal tropism was related to proliferation as this must ultimately be accomplished by cells that have adapted and exited dormancy. Figure III.15 confirms that these cells were not responsive to DHEA. In fact DHEA provided no viability increase in any of the experimental conditions. The 1 $\mu$ M dose of dopamine, a neurotransmitter also produced in the brain, showed increased viability relative to untreated cells at both 48 and 96 hours. In repetitions this increase was found to be significant and reproducible (Figure III.17).

#### **V.1.4 361-luc cells respond to and are capable of synthesis, storage and release of dopamine**

I next looked at dopamine signaling beginning with the receptors. By qPCR all five of the putative dopamine receptors are expressed in 361-luc cells. Receptor levels of DRD4 were found to change upon stimulation with dopamine when measured at 24 hours returning to baseline by 48. One of the two dopamine reuptake receptors, SLC6A3 showed a similar pattern. Use of the type-2 agonist Pramipexole (Px) to hyper stimulate dopamine signaling showed upregulation of all five GPCRs as well as SLC6A3. From these data we

can conclude that the machinery for responding to dopamine as a ligand does exist in 361-luc cells and it is responsive to dopamine.

Looking downstream of the receptor, canonical dopamine signaling (Figure III.18) involves the second messenger cAMP with DRD1 and DRD5 activating adenylyl cyclase to increase cAMP and DRD2 DRD3 and DRD4 decreasing cAMP concentration. cAMP works directly on PKA so I measured PKA activity via bioluminescence when 361-luc cells were stimulated with 1uM and 10 uM dopamine (Fig. III.24). At 1uM PKA activity was unchanged but with 10uM dopamine there was a significant increase in cAMP concentration. This implies that dopamine is signaling through the type 1 receptors. There is also substantial evidence that dopamine GPCRs can act via phospholipase C, calcium ion channels, intracellular calcium store mobilization, beta-arrestins and G-protein coupled receptor kinases and that Type 1 and Type 2 heterodimers are functional (Beaulieu and Gainetdinov, 2011). As with many GPCR related signal transduction networks, signal propagation is complex and highly context specific. For instance dopamine receptor 2 signaling is known to be anti-proliferative in adrenal cortical cells through angiotensin II stimulated phosphorylation of cyclin D1 (Chang et al., 2014). But in the sub ventricular zone of the brain the same receptor signals to selectively promote self-renewal of a multipotent stem cell population via protein kinase C mediated release of EGF (O’Keeffe et al., 2009). There is also an evolving understanding of differential functions of DRD2 isoforms resulting in a long version and a short version. The long form has been termed the canonical form and has post-synaptic function while the short form appears to be auto stimulatory found on the pre-synaptic membrane (Neve et al., 2013).

I shared my findings with dopamine researcher Dr. Emmanuel Pothos and he was intrigued as to how breast epithelial cancer cells may have adapted to the CNS environment by processing dopamine. As the cells expressed the reuptake receptor, we considered the cells may be able to store dopamine. Otherwise they would be importing a highly cytotoxic substance. The dose dependent nature of the viability response (Figure III.22) also shows they are able to tolerate and actually benefit from dopamine free in the media. By qPCR I was able to demonstrate Px induction of the SLC18A1 gene also known as the vesicular monoamine transporter 1 (VMAT1) dopamine processing protein which is responsible in cells of neuroendocrine origin for transport of dopamine into vesicles (Weihe et al., 1994). Dopamine beta hydroxylase which breaks down free dopamine was also induced providing additional evidence that these cells can package and also process dopamine.

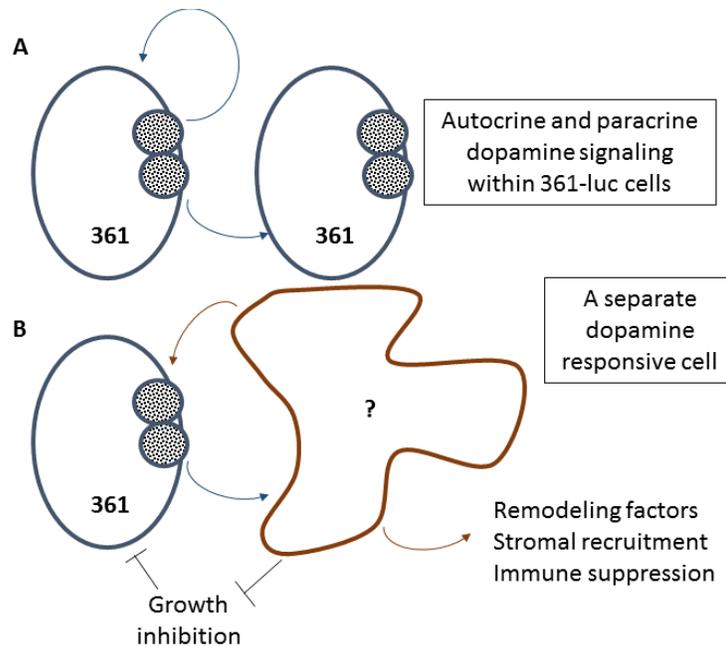
Dr. Pothos' lab has experimental expertise in measuring the quantal release of monoamine containing vesicles. We tested if when held in depolarizing conditions, 361-luc cells were capable of releasing catecholamines. Figure III.27 shows that they are and when the cells were pre-treated with the immediate precursor to dopamine, L-DOPA, sustained release of monoamine containing vesicles was observed. This implies that 361-luc cells are capable of synthesizing dopamine.

The dopamine biosynthetic pathway can begin with the conversion of phenylalanine to tyrosine by phenylalanine hydroxylase or start directly with tyrosine (Fig. III.20B). The rate limiting step is the conversion of tyrosine to L-DOPA by tyrosine hydroxylase. The DOPA decarboxylase produces dopamine which can be further metabolized by dopamine beta hydroxylase. The rate limiting TH was found to increase

after dopamine treatment and transcription of all of these enzymes was induced by Px stimulation of 361-luc cells (Figs. III.21 & III.23).

In summary, 361-luc cells colonize the brain and adrenal glands *in vivo*, both organs that synthesize dopamine. Dopamine provides an *in vitro* viability advantage relative to serum starved cells and all of the putative dopamine receptors are expressed. The expression of DRD4 and a dopamine reuptake receptor both fluctuate after dopamine stimulation to return back to baseline by 48 hours. The D2-type agonist Px upregulates dopamine receptors and biosynthetic enzymes, particularly the rate limiting biosynthetic TH and the dopamine vesicle transporter VMAT1. Additionally 361-luc cells respond to treatment with L-DOPA by releasing catecholamines in quantal bursts. Combined, these data are compelling evidence that 361-luc cells, of breast epithelial origin, are able to synthesize, store, release and respond to dopamine utilizing proteins that are ordinarily exclusive to neuroendocrine cells.

The evidence presented here introduces a novel concept for breast to brain metastasis of dopamine response being a mechanism that confers metastatic advantage. While the complete biochemical underpinnings are not evident, this is true of many processes believed to govern metastasis. Additional experiments are necessary to prove these findings, which will be discussed shortly, but if we assume the major results that 1) 361-luc cells increase proliferation when stimulated with dopamine and 2) they are capable of making, storing and releasing their own dopamine are true it is still not obvious why these cells would grow out in areas rich in dopamine production like the brain and the adrenal gland. If they can make their own dopamine, why do they need an external source? As illustrated in Figure V.1 there are several possible explanations.



**Figure V.1 Potential mechanisms for dopamine in mediating metastasis.**

It may very well be that autocrine and paracrine dopamine signaling can stimulate proliferation within 361-luc populations but there could be involvement of another dopamine responsive cell. In cartoon B a separate neuron could be dopamine responsive and stimulate 361-luc cells with additional dopamine stimulation or through an entirely different mechanism. The cell may also provide an indirect benefit by acting to relieve growth suppression, recruit stromal effectors that provide support to the 361-luc cells, promote local microenvironmental changes or suppress immune cells. If this second cell is common to the adrenal gland and the brain, as acetylcholine releasing neurons are known to be, a unifying mechanism could be at play here.

### **V.1.5 Limitations and future directions**

qPCR analysis is limited in that it only measures transcripts. While dopamine receptor transcription is strictly regulated (Bence et al., 2009; Seo et al., 2012), GPCR

internalization upon ligand binding is typical but proceeds on the order of minutes. Consequently, surface staining of the specific receptors at earlier time points could confirm the significance of the changes in message level reported here for the dopamine receptors.

As the receptor driving the signal is critical in dopamine, overexpression of the individual receptors, including the long and short form of DRD2 may be useful in determining which one is mediating the viability effect in 361-luc cells. There are some characterized activating mutations, but they do not exist for all five receptors and their mechanism of action is not well defined (Zhang et al., 2015). The use of receptor specific agonists could also confirm this line of experimentation, and knock down or genetic removal of the receptor(s) via CRISPR could be used to validate this.

While the finding that there is a significant difference in cell viability due to dopamine is intriguing the physiologic significance of this result is unclear. As the baseline for relative comparison is serum starved cells this assay is measuring the effects of dopamine on cells which are or will soon be dying. This may very well be relevant to what is happening in metastasis, and looking at the proliferative marker Ki67 did show dopamine increased proliferation relative to starved cells. However there have not yet been any biologic replicates for this experiment so statistical significance is not established. There also seemed to be a high amount of Ki67 staining with almost 18% of the serum starved cells staining green in Figure III.26. In addition to repetition, pharmacologic agents could be used to show amplification or muting of the proliferative effect and a complimentary apoptotic marker like cleaved caspase 3 could explain if dopamine imparts a survival advantage *in vitro* as well.

To prove this hypothesis beyond a reasonable scientific doubt would necessitate a few additional experiments. In addition to the experiments listed earlier to nail down the receptor involvement, electromicroscopic evidence of dopamine containing vesicles in 361-luc cells has long been the gold standard for intracellular packaging. HPLC experiments on intracellular vesicles could also confirm the presence of dopamine. Tracing dopamine signal transduction beyond the receptor can lead to a very long rabbit hole of individual hypotheses to be tested. It might be more productive to stimulate 361-luc cells with specific agonists or antagonists based on the receptor validation and then perform a timescale of metabolic and phosphoprotein profiling. This has the advantage of being able to verify changes in the packaging and biosynthetic machinery while also providing for discovery of the signal transduction network activity.

If it is true that 361-luc cells have adapted to use dopamine for a proliferative or cell death avoidance advantage *in vitro*, experiments must be done to judge the breadth of this effect *in vivo*. One way to do this would be to supply an ectopic source of dopamine before 361-luc UGIC and show colonization at this site. Dopamine is highly cytotoxic so continuous exposure should be avoided. There are several cell lines known to secrete dopamine, these could potentially be injected into a highly vascular site like the kidney capsule or the liver. But it might be more feasible to instead alter the dopamine processing ability of the 361-luc cells by knocking out and overexpressing the receptor involved and comparing UGIC colonization. There are also pharmacologic agents which degrade dopamine receptors that could be used before injection for comparison to undamaged cells. Animals could be treated with receptor agonists and antagonists to demonstrate changes in colonization.

These experiments could also address the question of dopamine utilization being a common mechanism for colonizing both the brain and the adrenal gland. Fortunately the growth rate of these cells *in vivo* is slow so brain metastases could be detected months after the adrenal tumors begin. The skull provides a significant barrier to luminescent signal detection and it would be necessary to image the brain *ex vivo* in order to accurately quantitate the brain foci. Staining of tumor sections for potential dopamine responsive partner cells, remodeling proteins and CNS growth inhibitors would address the questions illustrated in Figure V.1.

## **V.2 Chromatin modification and metastasis of 361-luc cells**

### **V.2.1 Chromatin modifier knock down screen of systemic colonization**

The question driving the chromatin modifier hairpin screen was if loss of a specific chromatin modifier could provide a metastatic advantage in the context of a luminal B tumor line. I knew that colonization was the rate limiting step and we wanted to test if we could wake up dormant tumor cells. The original expected outcome was that specific hairpins would result in the colonization of specific organs. At the onset of the project the 361-luc line was not known to be capable of spontaneously colonizing any organs from UGIC. Similarly to the original metastatic characterization of the line in the literature, I tested orthotopic injection and did not see any metastases (Fig. III.6). However, once the screen began, UGIC delivery had very different results (Fig IV.6 & IV.7). As I continued to collect the *in vivo* images over time it became apparent that the control cells which were not transduced with the library had multiple tumors growing out. So the question then became if there would be any difference in the distribution of tumors at specific sites in the library cells. In the end the adrenal and brain tropism was obvious regardless of treatment.

What was different is that there were some organs colonized in the library groups that were not in the controls. But each organ was colonized no more than once in either library. This made a statistical analysis of the tumor distribution difficult because of the very low sample sizes of zero in the control groups and one in the experimental groups. Grouping of the tumors into adrenal and non-adrenal and comparison by chi square did reveal an overall statistical difference in the tumor distributions for both libraries.

I designed a scheme to amplify a region of the miR-30 context hairpin vector in one low cycle PCR round followed by a second round of PCR to focus in on the hairpin and add the adapters including the flow cell binding sites and the bar codes (Figure IV.2A). This was accomplished by using a custom sequencing primer which sat down immediately adjacent to the half hairpin to provide maximum diversity in the first few bases read, necessary to facilitate accurate clustering on the chip. The sequence then read through the half hairpin, and into the hairpin loop before the bar code or index and then the flow cell adapter at the other end of the amplicon were added. The tumors were barcoded and sequenced and the scheme worked well for both libraries with millions of reads per tumor recovered. When mapped back to the reference hairpins there were ample reads to do very strict filtering such that every base read was unambiguous and 100% of the vector, hairpin and index sequences aligned.

However, despite the excellent quality of the sequences, the pLKO library had some problems. The negative control amplicon amplified from mouse genomic DNA which had never been virally transduced unexpectedly aligned with 98 hairpins with the top hairpin having 936 reads. On its own, this could be insignificant and a noise threshold could be set at 1000 reads. But the pre-screen pellet also varied from 31,144 reads for the most prevalent

hairpin down to one. When the pre-screen pellet results were compared to the negative control results it was apparent the two shared a similar bias because the top three hairpins were the same in both samples. I still do not understand how this happened as the same tools, procedures, purification methods and even the same plate of oligos and PCR machine were used for both libraries. Somehow a great deal of skew had been introduced to the pLKO library. It may very well have been present in the virus before transduction. But it never should have been possible to amplify 98 hairpins from a sample with assumedly no genetic material in the tube. I attempted to normalize results by defining a multiplier for each hairpin that leveled the pre-screen distribution. But the results applying the multiplier were nonsensical as hairpins that were present only once in the pre-screen pellet were now weighted over 30,000 times more than the most prevalent. This ended up with an enrichment for extremely low count hairpins.

Fortunately at the same time I was realizing that interpretation of the pLKO hits was going to be difficult, I also received the pHAGE reads. The highest count in the pre-screen pellet was tenfold lower than pLKO and only four hairpins were present in the negative control sample. The top negative control hit aligned exclusively to a firefly targeting hairpin included in the reference database as a decoy to determine noise in each sample.

There were only four genes which were targeted by more than one hairpin in the nonadrenal tumors: EHMT1, JMJD5, PRDM14 and SIRT3. The clinical survival data in Figures IV.11 and IV.12 implied that all four could be relevant to human disease, but SIRT3 loss appeared to have the best correlation with disease progression. Based on these data and reports in the literature of SIRT3 being a tumor suppressor we chose this hit for

the initial follow up. In hindsight, if I had done the analysis in table IV.V I would have probably concluded that PRDM14 and JMJD5 might have been better hits to go after because they had a greater number of overall hits, in more tumors and animals with better enrichment in the nonadrenal group.

Ironically, SIRT3 has nothing to do with chromatin. It is a deacetylase but it is exclusively mitochondrial and is important for maintaining mitochondrial integrity and balancing ROS. As SIRT3 loss has been demonstrated to affect ROS and ROS sensors can control metabolic reprogramming and regulate expression of stem like characteristics (Bigarella et al., 2014) I hypothesized that a metabolic shift, possibly one that led to dedifferentiation of the 361-luc cells could be responsible for the metastatic changes observed in the screen.

To investigate this *in vivo* I used five hairpins from the RNAi consortium which were individually validated to knock down SIRT3 and chose the two with the best mRNA reduction in 361-luc cells for injection. The knock down was verified to work at the protein level in Figure IV.14 as well. I injected 8 mice each with scramble control hairpins and hairpins #2 and #3 via UGIC to see if any changes in the organs colonized would be apparent.

While the tumors grew out I performed assays related to viability, self-renewal, ROS, motility and invasion and looked at progenitor markers by FACS. The shSIRT3 cells grew more slowly in culture and this was verified in the viability assay in Figure IV.17A. There were no obvious changes in self-renewal when the cells were cultured in low adhesion plates as tumor spheres. FACS staining with EpCAM and CD49f to visualize changes in mammary progenitor populations showed no change in the expression of

EpCAM with a slight increase in CD49f (Fig. IV.15). This pattern could be interpreted as a muted shift from mature luminal to luminal progenitor, but would require additional functional evidence to be meaningful which was lacking in the tumor sphere experiment. I measured ROS levels in the knock down cells using a fluorescent reporter via FACS. Figure IV.18 shows that control treatments ensured the assay could measure ROS changes in 361-luc cells and that loss of SIRT3 increased ROS.

Lastly, transwell migration through a simulated basement membrane showed no changes between shSIRT3 and shScr (Figure III.16). However in the same experiments when baseline motility was established without coating the porous transwell membrane fewer shSIRT3 cells were counted. As invasion in this assay is measured as the proportion of basement membrane penetrants (unchanged) divided by the uncoated membrane penetrants (fewer), there was an increase in invasion because of a smaller denominator. The validity of this finding is not clear to me. If the cells are less motile that alone seems enough to say they do not have increased invasive ability. If an equal number of cells are found on the other side of the membrane when coated with Matrigel, what does that mean? Perhaps there is a population of cells that will penetrate the Matrigel regardless of SIRT3 knock down. I used growth factor reduced Matrigel, but as Matrigel is a sarcoma secretion of undefined content this assay could very well be measuring chemo attraction to something in the Matrigel. This could be repeated with a different membrane coating, to help settle the question. The invasion assay results were enough to justify looking at local invasion *in vivo*, so six glands per hairpin were injected with one million cells (Fig. IV.19). Considered together, the *in vitro* results did not support a role for SIRT3 in metastasis.

*In vivo* the growth rate of the mammary fat pad tumors as judged by bioluminescent signal was not significantly different (Fig. IV.20). There were no spontaneous metastases detected. Unexpectedly in the UGIC mice there were many tumors found in the ovary (Fig. IV.21). In all three experimental groups there were ovarian growths when previously there had only been one ovarian tumor reported in each library and none in control cells. Statistically there was no difference in the distribution of the observed organs colonized in the shScr UGIC animals compared to either of the shSIRT3 hairpins.

### **V.2.2 Significance of findings and future directions**

It seems appropriate to say that SIRT3 knock down failed to recapitulate the original phenotype of the screen but a problem arises in defining exactly what that phenotype was. The result of the library screen was that more organs were colonized in the library cells. But each organ was only colonized at most once per library making it difficult to exclude chance as the main factor for growth in that organ. I found certain hairpins enriched in nonadrenal tumors. But three of those four hits were also found in the adrenal tumors. Of course the hairpin being in an adrenal tumor does not exclude it from having a role in metastasis. Thus, while I may not have had confidence in the overall strength of these hits, I was confident they were the strongest hits from this screen.

There are three other hits which could be examined. But the emergence of a more pronounced ovarian tropism in the shSIRT3/shScr UGIC animals suggests there is metastatic variability in the 361-luc line that was not recognized even in my screen with a total of 47 animals. The emergence of larger numbers of ovarian tumors further marginalizes the individual tumor hits that were in the library animals. The most obvious answer to why 361-luc cells grow in the ovary is that they are deposited there in proximity

to a follicle producing estrogen and these ER expressing cells respond with cell growth and proliferation.

Animal variability by age or strain, genetic drift *in vitro* and viral transduction parameters were all potential sources of variation which I attempted to minimize. The mice were all the same age plus or minus a week when the injections were performed, and they were all the same NOD/SCID background but they were ordered from two different vendors. The cells I used in this study were all within ten passages of each other. The shSIRT3 transduction was slightly different in that I did not measure the MOI. Transduction was performed in a 6 well plate seeded with 300,000 cells which seems like a large enough founder population to maintain all the heterogeneity of the parental line. After transduction and expansion, approximately 50% of the cells died during selection, so the MOI could not have been so high that insertional mutagenesis was likely a factor.

The chromatin modifier screen project evolved as the results were being realized and faced some technical challenges. While the bioinformatics process I designed and implemented successfully yielded high quality reads amenable to bar code stratification, the interpretation of the biological significance of the results was difficult due to small sample sizes and unexpected metastatic potential inherent in the line.

If I were to start over and maintain the focus on chromatin modifiers I would still use hairpins, even with the advent of CRISPR technology, as chromatin modifiers are essential cellular machinery and complete loss of wild type function does not occur in those cancers with consistently mutated chromatin modifiers. The screen could be done *in vitro* using an assay with defined positive and negative controls like EMT induction or invasion through Matrigel. This would allow a larger sample size and the screen could be made

more robust by including other tumor lines. However the results would be limited initially to one step in the metastatic cascade.

To truly model metastatic colonization requires *in vivo* experimentation. The 361-luc UGIC system stands out as a tractable luminal B metastasis model, but I was unable to leverage its strengths with the chromatin screen. To apply my findings to successfully do this I would explore a strategy known to be productive with other lines by serial passage through the brain and adrenal glands of 361-luc cells in multiple mice to elucidate changes in gene expression by RNA sequencing. This would provide data to determine how changes in expression of specific genes or gene signatures contribute to specific organ tropism and would provide fertile ground to elucidate the associated mechanisms.

# Chapter VI:

# Appendix

**Table VI.I: The histone code: chromatin marks, modifying enzymes and their functions.** Table modified from <http://www.cellsignal.com/common/content/content.jsp?id=science-tables-histone>

<b>Acetylation</b>			
Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys4	Esa1	transcriptional activation
	Lys5	Tip60, p300/CBP	transcriptional activation
	Lys7	Hat1	unknown
		Esa1	transcriptional activation
H2B	Lys5	p300, ATF2	transcriptional activation
	Lys11	Gcn5	transcriptional activation
	Lys12	p300/CBP, ATF2	transcriptional activation
	Lys15	p300/CBP, ATF2	transcriptional activation
	Lys16	Gcn5, Esa1	transcriptional activation
	Lys20	p300	transcriptional activation
H3	Lys4	Esa1	transcriptional activation
		Hpa2	unknown
	Lys9	unknown	histone deposition
		Gcn5, SRC-1	transcriptional activation
	Lys14	unknown	histone deposition
		Gcn5, PCAF	transcriptional activation
		Esa1, Tip60	transcriptional activation
		SRC-1	transcriptional activation
		Elp3	transcriptional activation (elongation)
		Hpa2	unknown
		hTFIIIC90	RNA polymerase III transcription
		TAF1	RNA polymerase II transcription
		Sas2	euchromatin
		Sas3	transcriptional activation (elongation)
		p300	transcriptional activation
	Lys18	Gcn5	transcriptional activation, DNA repair
			DNA replication, transcriptional activation
		p300/CBP	
	Lys23	unknown	histone deposition
Gcn5		transcriptional activation, DNA repair	
Sas3		transcriptional activation (elongation)	
	p300/CBP	transcriptional activation	
Lys27	Gcn5	transcriptional activation	

	Lys36	Gcn5	transcriptional activation
	Lys56	Spt10	transcriptional activation DNA repair
H4	Lys5	Hat1	histone deposition
		Esa1, Tip60	transcriptional activation DNA repair
		ATF2	transcriptional activation
	Lys8	Hpa2	unknown
		p300	transcriptional activation
		Gcn5, PCAF	transcriptional activation
	Lys12	Esa1, Tip60	transcriptional activation DNA repair
		ATF2	transcriptional activation
		Elp3	transcriptional activation (elongation)
	Lys16	p300	transcriptional activation
		Hat1	histone deposition telomeric silencing
		Esa1, Tip60	transcriptional activation DNA repair
Lys91	Hpa2	unknown	
	p300	transcriptional activation	
	Gcn5	transcriptional activation	
		MOF	transcriptional activation
		Esa1, Tip60	transcriptional activation DNA repair
		ATF2	transcriptional activation
		Sas2	euchromatin
		Hat1/Hat2	chromatin assembly

### Methylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H1	Lys26	Ezh2	transcriptional silencing
H2A	Arg3	PRMT1/6, PRMT5/7	transcriptional activation, transcriptional repression
H3	Arg2	PRMT5, PRMT6	transcriptional repression
	Arg8	PRMT5, PRMT2/6	transcriptional activation, transcriptional repression
	Arg17	CARM1	transcriptional activation
	Arg26	CARM1	transcriptional activation
	Arg42	CARM1	transcriptional activation
	Lys4	Set1	permissive euchromatin (di-Me)
		Set7/9	transcriptional activation (tri-Me)
		MLL, ALL-1	transcriptional activation
		Ash1	transcriptional activation

	Lys9	Suv39h,Clr4	transcriptional silencing (tri-Me) transcriptional repression genomic imprinting
		G9a	transcriptional repression (tri-Me)
		SETDB1	DNA methylation (tri-Me)
		Dim-5 , Kryptonite	transcriptional activation
	Lys27	Ezh2	transcriptional silencing X inactivation (tri-Me)
		G9a	transcriptional silencing
	Lys36	Set2	transcriptional activation (elongation)
	Lys79	Dot1	euchromatin transcriptional activation (elongation) checkpoint response

H4	Arg3	PRMT1/6 PRMT5/7	transcriptional activation transcriptional repression
	Lys20	PR-Set7 Suv4-20h Ash1 Set9	transcriptional silencing (mono-Me) heterochromatin (tri-Me) transcriptional activation checkpoint response
	Lys59	unknown	transcriptional silencing

### Phosphorylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H1	Ser27	unknown	transcriptional activation, chromatin decondensation
H2A	Ser1	unknown MSK1	mitosis, chromatin assembly transcriptional repression
	Ser122	unknown	DNA repair
	Ser129	Mec1, Tel1	DNA repair
	Thr119	NHK1	mitosis
	Thr120	Bub1, VprBP	mitosis, transcriptional repression
H2A.X	Thr142	WSTF	apoptosis, DNA repair
	Ser139	ATR, ATM, DNA-PK	DNA repair
H2B	Ser10	Ste20	apoptosis
	Ser14	Mst1 unknown	apoptosis DNA repair
	Ser33	TAF1	transcriptional activation
	Ser36	AMPK	transcriptional activation
H3	Ser10	Aurora-B kinase MSK1, MSK2 IKK-a Snf1	mitosis, meiosis immediate-early gene activation transcriptional activation transcriptional activation
	Ser28	Aurora-B kinase MSK1, MSK2	mitosis immediate-early activation

	Thr3	Haspin/Gsg2	mitosis
	Thr6	PKCβI	
	Thr11	DIK/Zip	mitosis
	Tyr41	JAK2	transcriptional activation
	Tyr45	PKCd	apoptosis
H4	Ser1	unknown	mitosis, chromatin assembly
		CK2	DNA repair

### Ubiquitylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys119	Ring2	spermatogenesis
H2B	Lys120	UbcH6	meiosis
	Lys123	Rad6	transcriptional activation euchromatin

### Sumoylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys126	Ubc9	transcriptional repression
	Lys6 or		
H2B	Lys7	Ubc9	transcriptional repression
H4	NTD	Ubc9	transcriptional repression

### Biotinylation

Histone	Site	Histone-modifying Enzymes	Proposed Function
H2A	Lys9	biotinidase	unknown
	Lys13	biotinidase	unknown
H3	Lys4	biotinidase	gene expression
	Lys9	biotinidase	gene expression
	Lys18	biotinidase	gene expression
H4	Lys12	biotinidase	DNA damage response

### Newly discovered histone modifications with emerging significance

Citrullination  
 Formylation  
 Propionylation  
 Butyrylation  
 Crotonylation  
 Glutarylation  
 Succinylation  
 Malonylation  
 Hydroxylation

## **VI.1 The RasGAP gene, RASAL2, is a tumor and metastasis suppressor**

### **Abstract**

This paper published in Cancer Cell was a collaboration with Karen Cichowski's lab looking at the Ras negative regulator RASAL2 which they had identified as a potential metastasis suppressor. Ras is not frequently mutated in breast cancer but Ras/ERK signaling is frequently amplified. This paper identified loss of RASAL2 as a mechanism contributing to increased Ras output without mutation. I performed the orthotopic injection of MDA-MB-361 cells containing retrovirally expressed RASAL2 mutants transduced by Sara Koenig McLaughlin. I measured the tumor size of the resultant tumors over time. Together we analyzed H&E sections of the resultant tumors and photographed them

# The RasGAP Gene, *RASAL2*, Is a Tumor and Metastasis Suppressor

Sara Koenig McLaughlin,<sup>1,2,3</sup> Sarah Naomi Olsen,<sup>1,2,3</sup> Benjamin Dake,<sup>4,5</sup> Thomas De Raedt,<sup>1,2,3,9</sup> Elgene Lim,<sup>2,3,6</sup> Roderick Terry Bronson,<sup>3</sup> Rameen Beroukhi,<sup>2,3,7,8</sup> Kornelia Polyak,<sup>2,3,6</sup> Myles Brown,<sup>2,3,6</sup> Charlotte Kuperwasser,<sup>4,5</sup> and Karen Cichowski<sup>1,2,3,9,\*</sup>

<sup>1</sup>Genetics Division

<sup>2</sup>Department of Medicine

Brigham and Women's Hospital, Boston, MA 02115, USA

<sup>3</sup>Harvard Medical School, Boston, MA 02115, USA

<sup>4</sup>Department of Developmental, Molecular, and Biochemical Biology, Sackler School of Graduate Biomedical Sciences, Tufts University School of Medicine, 136 Harrison Avenue, Boston, MA 02111, USA

<sup>5</sup>Molecular Oncology Research Institute, Tufts Medical Center, 800 Washington Street, Boston, MA 02111, USA

<sup>6</sup>Division of Molecular and Cellular Oncology, Department of Medical Oncology

<sup>7</sup>Department of Cancer Biology

Dana-Farber Cancer Institute, Boston, MA 02215, USA

<sup>8</sup>Broad Institute of Massachusetts Institute of Technology and Harvard University, Cambridge, MA 02142, USA

<sup>9</sup>Ludwig Center at Dana-Farber/Harvard Cancer Center, Boston, MA 02115, USA

\*Correspondence: [kcichowski@rics.bwh.harvard.edu](mailto:kcichowski@rics.bwh.harvard.edu)

<http://dx.doi.org/10.1016/j.ccr.2013.08.004>

## SUMMARY

*RAS* genes are commonly mutated in cancer; however, *RAS* mutations are rare in breast cancer, despite frequent hyperactivation of Ras and ERK. Here, we report that the RasGAP gene, *RASAL2*, functions as a tumor and metastasis suppressor. *RASAL2* is mutated or suppressed in human breast cancer, and *RASAL2* ablation promotes tumor growth, progression, and metastasis in mouse models. In human breast cancer, *RASAL2* loss is associated with metastatic disease; low *RASAL2* levels correlate with recurrence of luminal B tumors; and *RASAL2* ablation promotes metastasis of luminal mouse tumors. Additional data reveal a broader role for *RASAL2* inactivation in other tumor types. These studies highlight the expanding role of RasGAPs and reveal an alternative mechanism of activating Ras in cancer.

## INTRODUCTION

The Ras pathway is one of the most commonly deregulated pathways in human cancer (Downward, 2003). Mutations in *RAS* genes occur in a variety of tumor types (Karnoub and Weinberg, 2008; Pylayeva-Gupta et al., 2011); however, the Ras pathway is also frequently activated as a consequence of alterations in upstream regulators and downstream effectors, underscoring the importance of this pathway in cancer (Downward, 2003).

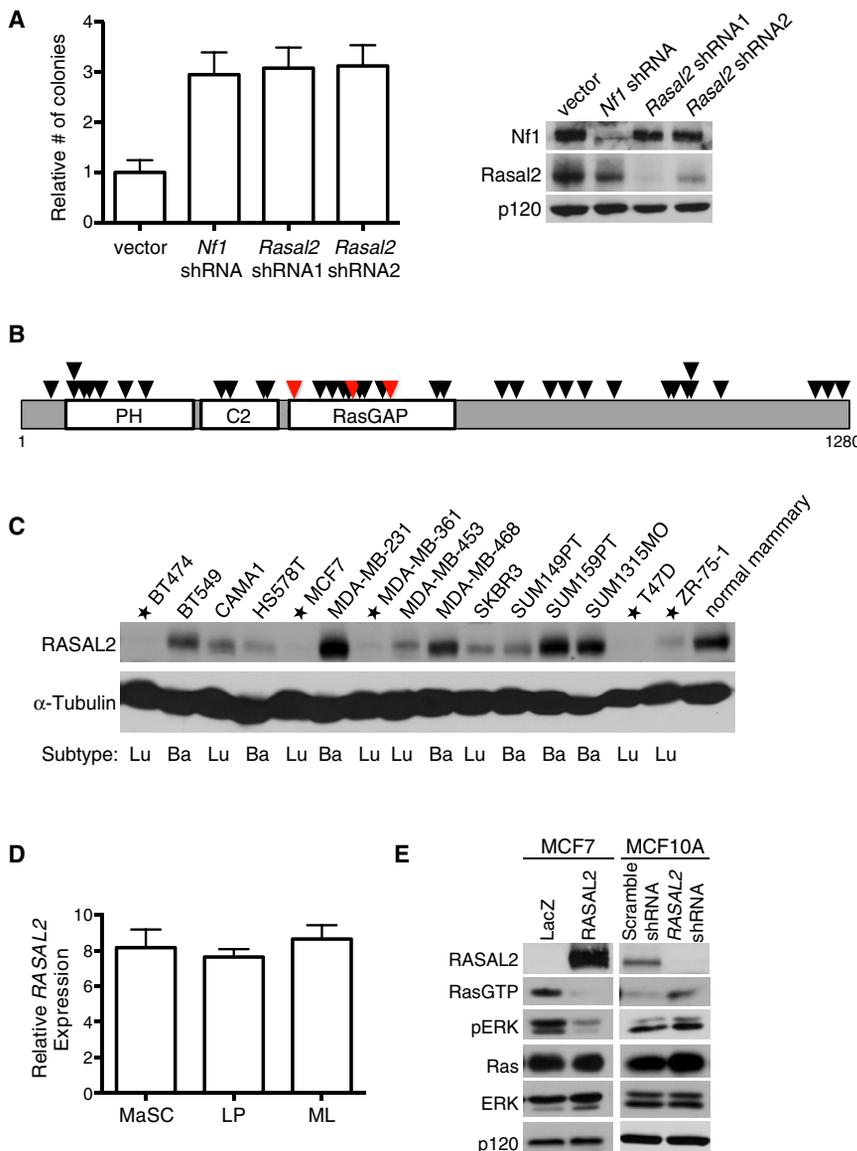
Ras is negatively regulated by Ras GTPase-activating proteins (RasGAPs), which catalyze the hydrolysis of Ras-GTP to Ras-GDP (Bernards, 2003). As such, RasGAPs are poised to function as potential tumor suppressors. Indeed, the *NF1* tumor suppressor encodes a RasGAP and is mutated in the familial cancer syndrome

neurofibromatosis type 1 (Cawthon et al., 1990). *NF1* also is lost or suppressed in sporadic cancers, including glioblastoma (Cancer Genome Atlas Research Network, 2008; Parsons et al., 2008; McGillicuddy et al., 2009), nonsmall cell lung cancer (Ding et al., 2008), neuroblastoma (Hölzel et al., 2010), and melanoma (Krauthammer et al., 2012; Maertens et al., 2012). More recently, the RasGAP gene, *DAB2IP*, has been shown to function as a potent tumor and metastasis suppressor in prostate cancer (Min et al., 2010). In total, there are 14 RasGAP genes in the human genome (Bernards, 2003). All contain a RasGAP domain but exhibit little similarity elsewhere. It is currently unknown whether any of these other genes may also function as human tumor suppressors.

Breast cancer is the most common cancer in women worldwide (Kamangar et al., 2006). *K-*, *H-*, and *N-RAS* mutations are

### Significance

The RasGAPs are direct negative regulators of Ras and are therefore poised to function as potential tumor suppressors. Here, we identify a RasGAP gene, *RASAL2*, as a tumor suppressor within this gene family. Our data suggest that *RASAL2* loss plays a causal role in the development, progression, and metastasis of breast cancer and may play a broader role in the metastasis of other solid tumors as well. Collectively, these data reveal an alternative mechanism by which Ras becomes activated in cancer and identify a role for *RASAL2* and Ras in breast cancer progression and metastasis.



**Figure 1. RASAL2 Is a Candidate Tumor Suppressor**

(A) Left: immortalized MEFs were infected with lentiviral shRNAs that target *Rasal2*, *Nf1*, or control and were plated in soft agar. Data are reported as relative number of colonies  $\pm$  SEM. Inactivation of *Nf1* or *Rasal2* induced a statistically significant increase in anchorage-independent growth ( $p \leq 0.0001$ ). Right: western blot confirming knockdown. (B) RASAL2 mutations in human tumor samples (Bamford et al., 2004). Each triangle represents a nonsynonymous mutation. Red triangles indicate breast cancer mutations. See also Tables S1 and S2.

(C) RASAL2 expression in a panel of human breast cancer cell lines in comparison to normal human mammary epithelial cells. Cell lines with very low or no RASAL2 are starred. Luminal (Lu) or basal (Ba) subtype categorization is indicated.

(D) Relative RASAL2 expression in subsets of sorted human mammary epithelial cells (Lim et al., 2009). MaSC, mammary stem cell enriched: (CD49hi EpCAM<sup>-</sup>). LP: luminal progenitor (CD49<sup>+</sup> EpCAM<sup>+</sup>). ML, mature luminal (CD49<sup>+</sup> EpCAM<sup>+</sup>). Data show relative expression  $\pm$  SD. Similar results were obtained using two additional RASAL2 probes. There were no statistically significant differences in RASAL2 expression between subsets of cells.

(E) Left: western blot of Ras-GTP and phospho-ERK (pERK) levels in MCF7 cells following expression of LacZ or RASAL2. Right: western blot of Ras-GTP and phospho-ERK (pERK) levels in MCF10A cells following shRNA-mediated inactivation of RASAL2 or control (nontargeting "Scramble" shRNA).

## RESULTS

### The RasGAP Gene, RASAL2, Is a Candidate Tumor Suppressor

We previously developed a cell-based screen to identify additional RasGAPs that might function as tumor suppressors (Min et al., 2010). Distinct small hairpin

RNAs (shRNAs) that recognize individual RasGAP genes were introduced into immortalized mouse embryonic fibroblasts (MEFs), and cells were evaluated for the ability to grow in soft agar. Three genes scored in this screen: *Nf1*, a well-documented tumor suppressor gene, *Dab2ip*, which we have since shown is a tumor suppressor in prostate cancer, and *Rasal2*, a third RasGAP gene (Min et al., 2010). Several *Rasal2*-specific shRNA sequences promoted colony growth in this assay and did so as well as *Nf1*- and *Dab2ip*-specific shRNAs (Figure 1A; Min et al., 2010). Notably, transformation was not generally promoted by the loss of any RasGAP, suggesting that only a subset of RasGAPs may function as tumor suppressors (Min et al., 2010). Upon identifying RASAL2 as a candidate tumor suppressor, we searched publicly available databases and found mutations within the catalytic RasGAP domain in human breast cancers (Figure 1B; Table S1 available online) (Sjöblom et al., 2006; Shah et al., 2012). Current genomic mutation databases indicate that RASAL2 is also mutated in several other tumor

relatively rare in this tumor type, and together they have been detected in only  $\sim 3.2\%$  of all breast lesions (Bamford et al., 2004). Nevertheless, the Ras/ERK pathway is hyperactivated in  $\geq 50\%$  of breast cancers and has been proposed to be involved in tumor progression and recurrence, suggesting that Ras may be more frequently activated by other mechanisms in these tumors (Sivaraman et al., 1997; von Lintig et al., 2000; Mueller et al., 2000). In this study, we demonstrate that the RasGAP gene, RASAL2, functions as a tumor suppressor in breast cancer. Through the analysis of human tumor samples, human xenografts, and genetically engineered mouse models, we show that RASAL2 loss plays a causal role in breast cancer development and metastasis. Additional mouse modeling studies reveal a broader potential role for RASAL2 in other tumor types. Together, these studies highlight the expanding role of RasGAP genes in cancer and reveal an important mechanism by which Ras becomes activated in breast tumors.

types, including colorectal, lung, and ovarian tumors (Figure 1B; Table S2). In total, 42 nonsynonymous mutations have been detected in *RASAL2*, 31% of which reside in the catalytic RasGAP domain, many of which are predicted to be deleterious (Tables S2 and S3). Because the mechanism by which Ras becomes activated in breast cancer is largely unknown, and because mutations in breast tumors were among the first to be identified, we began by investigating a potential role for *RASAL2* inactivation in breast cancer development.

Work from our laboratory and others have shown that the RasGAP genes *NF1* and *DAB2IP* are inactivated in cancer by genetic, epigenetic, and proteasomal mechanisms (Dote et al., 2004; McGillicuddy et al., 2009; Min et al., 2010). Moreover, in many instances the nongenetic mechanisms of inactivation of these tumor suppressors appear to be more prevalent than mutational events in sporadic tumors (McGillicuddy et al., 2009; Min et al., 2010; Maertens et al., 2012). Therefore, we began by examining *RASAL2* protein expression in a panel of breast cancer cell lines. In comparison to normal mammary epithelial cells, in at least 5 out of 15 breast cancer cell lines *RASAL2* was absent or minimally expressed, suggesting that *RASAL2* may be lost or suppressed in this tumor type (Figure 1C). *RASAL2* levels were high in MDA-MB-231 and SUM159PT cells, which are known to harbor mutations in *KRAS* and *HRAS*, respectively (Hollestelle et al., 2007). We also noted that *RASAL2* was frequently absent in cells derived from luminal cancers in this panel of lines. Cell sorting studies indicate that there are no inherent differences in *RASAL2* expression in any specific cell population within the mammary cell hierarchy: luminal progenitor, mature luminal, or mammary stem cell enriched, suggesting that the low *RASAL2* levels associated with luminal cancer cell lines are not inherently associated with a pre-existing reduction in *RASAL2* levels due to a specific cell of origin or fate, as has been suggested for other genes (Figure 1D) (Lim et al., 2009). When *RASAL2* was reconstituted in MCF7 cells, which express little to no endogenous *RASAL2*, Ras-GTP and phospho-ERK levels were suppressed (Figure 1E). Conversely, acute inactivation of *RASAL2* via shRNA sequences in immortalized mammary epithelial cells (MCF10A) increased Ras-GTP and phospho-ERK levels (Figure 1E). These data confirm that *RASAL2* is a functional RasGAP and that loss of *RASAL2* activates Ras and ERK in this tumor type.

### **RASAL2 Functions as a Tumor Suppressor in Breast Cancer**

We next investigated the biological consequences of reconstituting or suppressing *RASAL2* in breast cancer cell lines. When *RASAL2* was introduced into human breast cancer cells that lack endogenous *RASAL2*, proliferation was largely unaffected (Figure 2A); however, *RASAL2* reconstitution significantly inhibited anchorage-independent colony growth (Figure 2B). In contrast, *RASAL2* did not inhibit colony growth of SUM159PT cells, which retain *RASAL2* expression but harbor an activating *RAS* mutation (Figure 2B). *RASAL2* also potently suppressed the growth of *RASAL2*-deficient breast cancer xenografts in vivo but again had no effect on *RAS* mutant tumors (Figure 2C). Conversely, shRNA-mediated suppression of endogenous *RASAL2* in a breast cancer cell line that normally does not grow well as a xenograft promoted tumor growth in vivo (Fig-

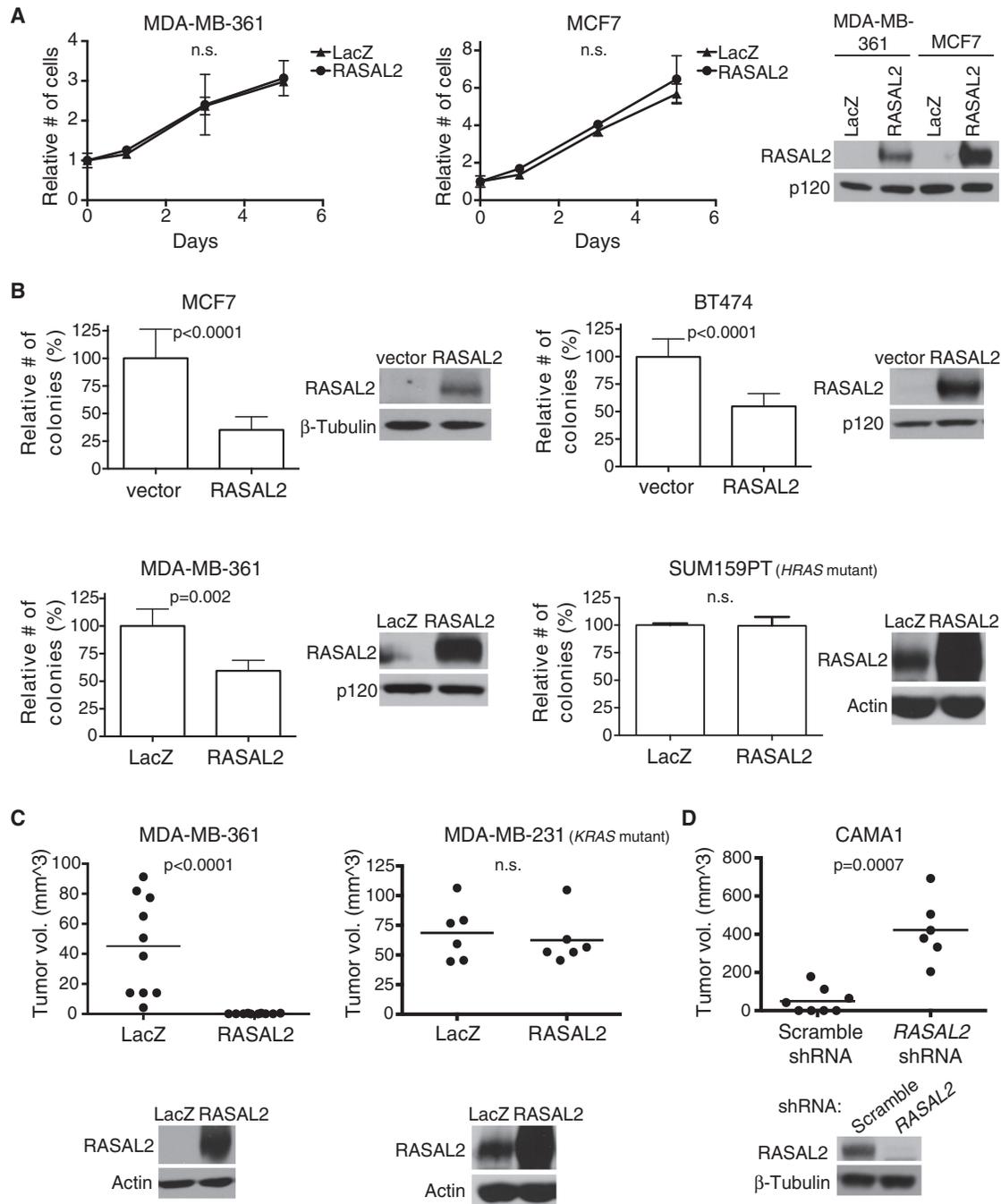
ure 2D). Together, these gain- and loss-of-function studies suggest that *RASAL2* can function as a tumor suppressor in the mammary epithelium and that inactivation or loss of *RASAL2* can contribute to mammary tumor development. Notably, like *NF1* and *DAB2IP*, *RASAL2* appears to restrict transformation and/or anchorage-independent growth, rather than generally suppressing cell proliferation in two-dimensional culture systems (Johannessen et al., 2005; Min et al., 2010).

### **RASAL2 Functions via Its Effects on Ras**

To determine whether the RasGAP domain of *RASAL2* and effects on Ras were critical for tumor suppression, we first evaluated the effects of *RASAL2* mutations identified in human breast cancer samples. Two of the three RasGAP domain mutants (K417E and K567X) failed to suppress anchorage-independent growth, demonstrating that these two mutations result in a clear loss of function (Figure 3A). The third mutation, which resulted in a more conservative amino acid change (E509D), still retained activity in this assay and therefore does not appear to be pathogenic; however, a number of additional nonconservative mutations have been detected in the RasGAP domain in other tumor types (Table S3). Consistent with these biological observations, both the K417E and the K567X mutations were defective in their ability to suppress activation of the Ras/ERK pathway (Figure 3B). Phospho-ERK levels in xenograft tumors further illustrate the difference in activity between pathogenic and nonpathogenic mutations (Figure 3C). To complement these studies, we investigated which Ras isoforms were activated in response to *RASAL2* suppression and found that both K-Ras and H-Ras-GTP levels were elevated (Figure 3D). Accordingly, ablation of either *KRAS* or *HRAS* suppressed colony growth by more than 50% (Figure 3E). Together, these results demonstrate that the RasGAP domain is essential for *RASAL2* tumor suppressor function and that both H- and K-Ras contribute to the pathogenesis caused by *RASAL2* inactivation.

### **RASAL2 Inactivation Promotes Migration, Invasion, and Tumor Progression**

To fully characterize the oncogenic effects of *RASAL2* loss, we investigated whether *RASAL2* suppression might also promote migration, invasion, and tumor progression. *RASAL2* suppression promoted the migration of MCF10A cells in a wound-healing assay (Figures 4A and 4B) and significantly enhanced invasion through Matrigel (Figure 4C,  $p = 0.002$ ). Similar to what was observed in colony assays shown in Figure 3E, ablation of *HRAS* or *KRAS* reduced invasiveness (Figure S1). We also utilized a xenograft model of breast cancer progression that mimics the progression of ductal carcinoma in situ (DCIS) to invasive carcinoma (Miller et al., 2000; Hu et al., 2008). Specifically, MCF10ADCIS cells, derivatives of MCF10A cells that are enriched for a progenitor population of cells (Miller et al., 2000), develop into DCIS-like lesions when grown as xenografts in mice. However, after a latency of approximately 8 weeks, they progress to invasive carcinoma, characterized by the loss of the myoepithelial cell layer and basement membrane (Hu et al., 2008). We acutely inactivated *RASAL2*, using two distinct shRNAs in these cells, and found that *RASAL2* inactivation accelerated tumor progression, resulting in a rapid disruption of the myoepithelium and basement membrane and the



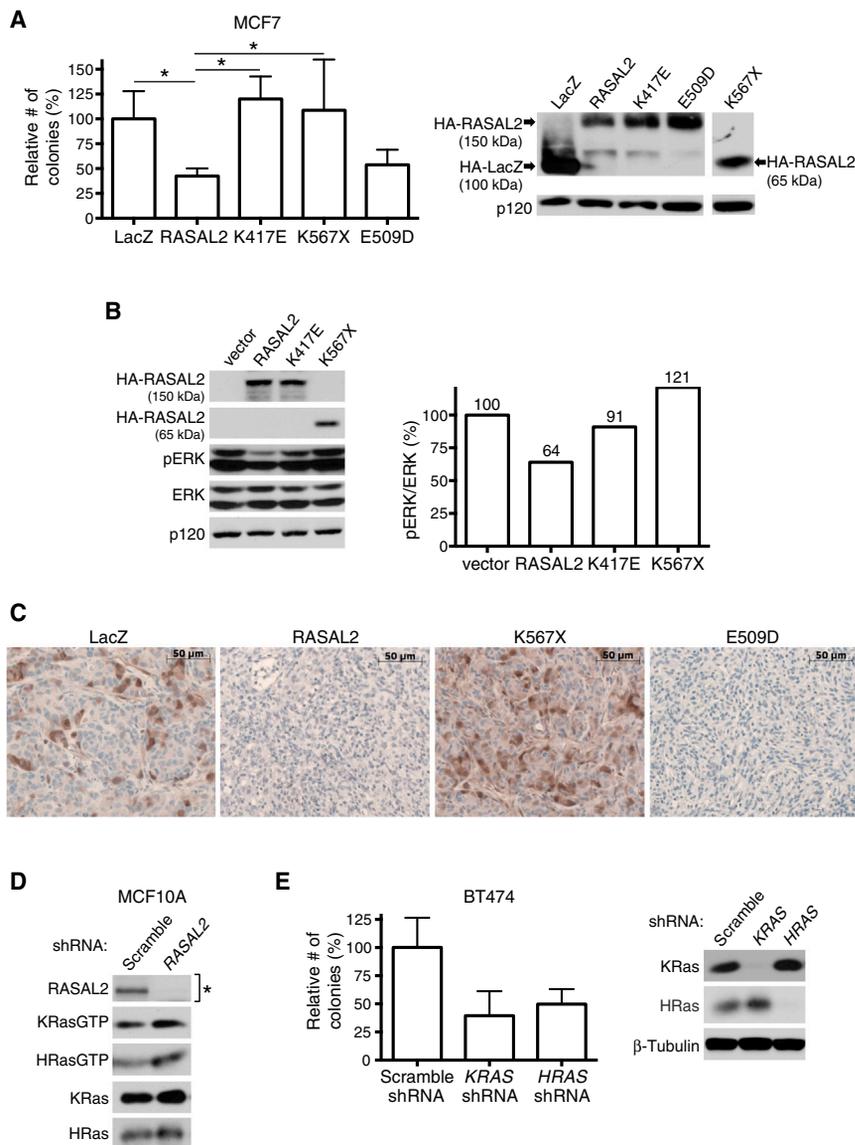
**Figure 2. RASAL2 Functions as a Tumor Suppressor in Breast Cancer**

(A) Growth curves of MDA-MB-361 and MCF7 cells expressing RASAL2 or LacZ. Data points show triplicate averages  $\pm$  SD. There were no statistically significant differences in proliferation. Western blot on right confirms ectopic RASAL2 expression.

(B) Soft agar colony formation of MCF7, BT474, MDA-MB-361, and SUM159PT cells expressing RASAL2 or LacZ. Data show relative number of colonies  $\pm$  SD. There was a statistically significant decrease in anchorage-independent growth upon ectopic RASAL2 expression in *RAS* wild-type cell lines (MCF7 and BT474 p < 0.0001; MDA-MB-361 p = 0.002), but not in the *HRAS* mutant cell line SUM159PT. Western blots confirm ectopic RASAL2 expression.

(C) Xenograft tumor formation of MDA-MB-361 and MDA-MB-231 cells expressing RASAL2 or LacZ. MDA-MB-361 cells were injected orthotopically into female NOD/SCID mice; MDA-MB-231 cells were injected subcutaneously into female nude mice. Horizontal bars indicate mean tumor volume. There was a statistically significant decrease in tumor growth upon ectopic RASAL2 expression (p < 0.0001) in the *RAS* wild-type cell line MDA-MB-361, but not in the *KRAS* mutant cell line MDA-MB-231. Western blots below confirm ectopic RASAL2 expression.

(D) Xenograft tumor formation of CAMA1 cells infected with shRNAs targeting RASAL2 or nontargeting control shRNA and injected subcutaneously into female NOD/SCID mice. Horizontal bars indicate mean tumor volume. There was a statistically significant increase in tumor growth upon RASAL2 inactivation (p = 0.0007). Western blot confirms RASAL2 knockdown.



**Figure 3. RASAL2 Functions as a Tumor Suppressor via Its Effects on Ras**

(A) Soft agar colony formation of MCF7 cells expressing HA-tagged LacZ, wild-type, or mutant RASAL2 (see also Table S3). Data show relative number of colonies  $\pm$  SD. \* $p \leq 0.05$ . Western blot confirms expression of constructs.

(B) Western blot reflecting the relative activation of the Ras/ERK pathway in the presence of HA-tagged wild-type or mutant RASAL2. The pERK/ERK ratio of each sample was calculated and normalized to the vector control.

(C) Phospho-ERK (pERK) expression in MDA-MB-361 xenograft tumors. LacZ, RASAL2, or mutant RASAL2 was expressed in MDA-MB-361 cells, and cells were injected orthotopically into female NOD/SCID mice. pERK levels were assessed by immunohistochemistry.

(D) Western blot showing HRas-GTP and KRas-GTP levels in MCF10A cells following shRNA-mediated inactivation of RASAL2 or control shRNA. As indicated by the asterisk, the blot confirming RASAL2 knockdown is a duplicate from Figure 1E, as these immunoblots were generated from the same samples.

(E) Soft agar colony formation of BT474 cells infected with an shRNA targeting HRAS or KRAS or a nontargeting control. Data show relative number of colonies  $\pm$  SD. Western blot confirms Ras isoform-specific knockdown.

development of invasive adenocarcinoma after just 3 weeks (Figure 4D). These results suggest that RASAL2 inactivation may also play a role in breast cancer progression. Notably, tumors in which RASAL2 had been depleted also exhibited a marked elevation of phospho-ERK as compared to control tumors (Figure 4E).

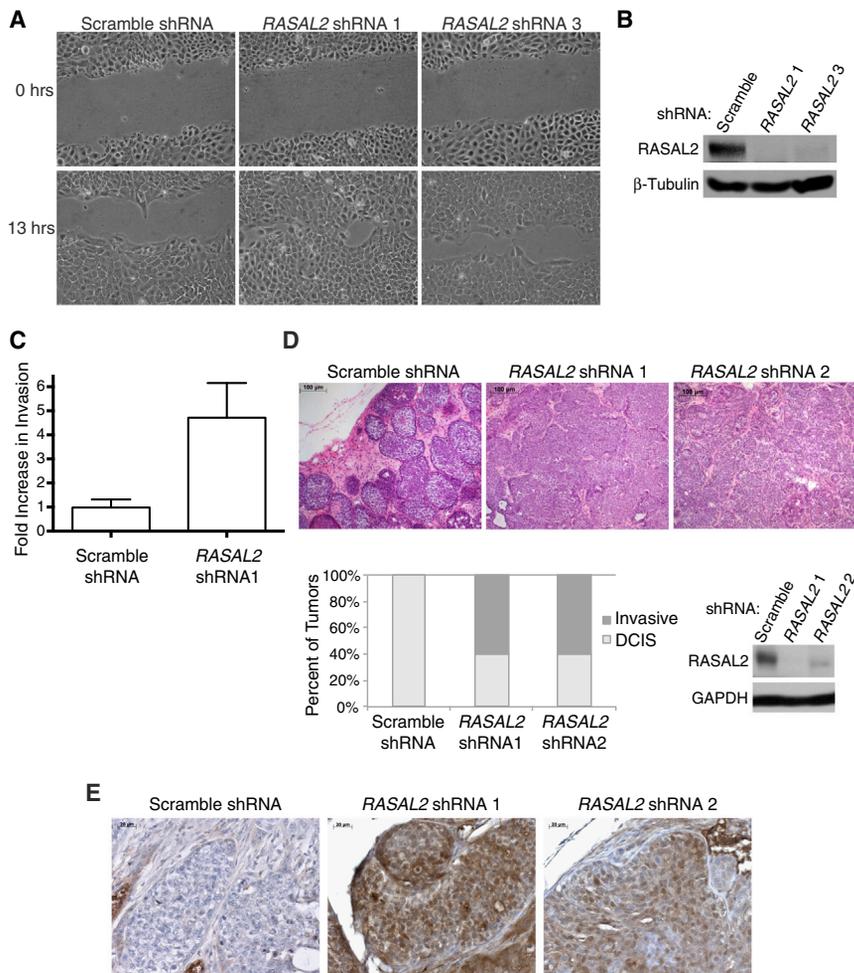
### Loss of *Rasal2* Promotes Metastasis and Ras Activation in a Genetically Engineered Mouse Model of Luminal Breast Cancer

As a rigorous and complementary means of investigating the biological consequences of RASAL2 inactivation in vivo, we generated genetically engineered mice that lack *Rasal2*. Mouse embryonic stem cells that contain a gene-trap cassette within the third intron of *Rasal2* were used to generate *Rasal2*-deficient mice (Figure 5A). Appropriate integration and loss of *Rasal2* expression were confirmed in heterozygous and homozygous mutant animals (Figures 5B and 5C). *Rasal2*<sup>-/-</sup> mice were viable,

fertile, and born at Mendelian ratios. We found that mutant animals did exhibit shorter overall survival as compared to control animals (77.8 compared to 95.6 weeks;  $p = 0.007$ , Figure S2A). However, there was no obvious difference in phenotype between wild-type and *Rasal2*<sup>-/-</sup> mice. A subset of animals from both cohorts developed tumors associated with old age. Whereas *Rasal2* mutant mice developed these tumors earlier, the tumor spectrum was similar

to wild-type animals, and they did not develop mammary lesions (Figure S2B). These results indicate that *Rasal2* loss is not sufficient to drive breast cancer in mice but may play a more general role in enhancing the development of other spontaneous tumors.

To examine the effects of RASAL2 loss on mammary tumorigenesis, we crossed *Rasal2*<sup>-/-</sup> mice to animals that constitutively overexpress a wild-type *Her2* (*ErbB2*) transgene in the mammary epithelium (*MMTVneu* mice) (Guy et al., 1992): a mouse model of luminal tumors (Herschkowitz et al., 2007). These tumors exhibit some differences from human luminal cancers, as they do not express estrogen receptor; however, unsupervised hierarchical clustering analysis and tumor pathology demonstrate that lesions from these animals recapitulate many of the key features of human luminal tumors (Guy et al., 1992; Herschkowitz et al., 2007). As such, these animals are currently the best available genetically engineered mouse model for luminal cancer (Guy et al., 1992; Herschkowitz et al., 2007). Female *MMTVneu* mice develop focal luminal mammary tumors,



**Figure 4. RASAL2 Inactivation Promotes Migration, Invasion, and Tumor Progression**

(A) Cell migration of MCF10A cells infected with shRNAs targeting *RASAL2* or a nontargeting control.

(B) Western blot confirming *RASAL2* knockdown in MCF10A cells used in (A) and (C).

(C) Transwell invasion of MCF10A cells infected with an shRNA targeting *RASAL2* or a nontargeting control. Invasion was measured after 24 hr and reported as average  $\pm$  SD ( $p = 0.002$ ).

(D) Xenograft tumor progression of MCF10ADCIS cells infected with shRNAs targeting *RASAL2* or a nontargeting control. Top: H&E images of xenograft tumors. Bottom left: quantification of xenograft tumor progression. Bottom right: western blot confirming *RASAL2* knockdown.

(E) Phospho-ERK (pERK) expression in MCF10ADCIS xenograft tumors from (D) as assessed via immunohistochemistry. See also Figure S1.

and a fraction of tumor-bearing females develop lung metastases (Guy et al., 1992). As expected, *MMTVneu* and *MMTVneu; Rasal2<sup>-/-</sup>* compound mice developed mammary adenocarcinomas and did so at a similar high frequency (Figure 5D, top panels; Figure S2E). Strikingly, however, we found that *MMTVneu; Rasal2<sup>-/-</sup>* mice developed substantially more metastases than *MMTVneu* animals. First, a higher fraction of compound mutant mice developed lung metastases (Figure 5E, 74% versus 46%,  $p = 0.05$ ). Second, compound mutant mice developed more metastases per lung than *MMTVneu* animals (Figure 5E, 30 per mouse versus 8 per mouse,  $p = 0.04$ ). Finally, the metastases were significantly larger in *MMTVneu; Rasal2<sup>-/-</sup>* mice as compared to *MMTVneu* mice (Figure 5D, bottom panels; Figure 5E,  $p = 0.04$ ). Interestingly, a subset of compound mutant mice developed tumors that metastasized to other organs, including brain, kidney, ovary, and gastrointestinal tract, a phenomenon not observed in *MMTVneu* animals historically or in our cohort (Figure 5F). Moreover, in most autochthonous mouse models of mammary adenocarcinoma, metastasis is typically limited to the lung and occasionally lymph nodes (Kim and Baek, 2010). However, human breast cancers do frequently metastasize to the brain and these other distal sites, underscoring the significance of these observations and the potential utility of this mouse model (Weigelt et al., 2005). Despite the dramatic

*Rasal2<sup>-/-</sup>* lesions (Figure 5G). In addition, we found that *Rasal2* was spontaneously lost or suppressed in a subset of *MMTVneu* tumors, and this loss or suppression was accompanied by a substantial increase in phospho-ERK and phospho-AKT levels (Figure 5G). Finally, the primary tumor that spontaneously lost/suppressed *Rasal2* and exhibited the most robust activation of the Ras pathway was a metastatic outlier within the *MMTVneu* cohort (Tumor 6, Figure 5G; Figure S2G). Taken together, these findings indicate that *Rasal2* loss enhances Ras activity in mammary tumors and that it promotes tumor progression, invasion, and metastasis in both autochthonous mouse models of breast cancer and human xenografts.

#### RASAL2 in Primary Human Breast Cancers

Genomic analyses demonstrate that *RASAL2* mutations do occur in human breast cancer but are relatively rare (Bamford et al., 2004; Sjöblom et al., 2006; Shah et al., 2012; Table S1). However, the two other known RasGAP tumor suppressors appear to be more frequently inactivated in cancer via nongenetic mechanisms. To more accurately determine how frequently *RASAL2* is lost or suppressed in human breast cancers, we directly examined *RASAL2* protein levels in primary human tumors. Existing *RASAL2* antibodies cannot be used for immunohistochemistry; therefore, we obtained breast cancer

arrays comprised of 55 sets of protein lysates (in triplicate) from matched primary breast tumors and adjacent normal mammary tissue taken from naive patients (Mueller et al., 2010). These tumor samples were histologically verified to contain at least 80% cancer cells and the normal tissue is cancer cell free. We first validated our purified RASAL2 antibody in this assay and found that dot blots from RASAL2-expressing and nonexpressing human breast cancer cell lines exhibited the expected pattern of expression (Figure 6A, top). RASAL2-specific shRNA sequences also effectively ablated expression in this assay (Figure 6A, bottom). Using the tumor arrays, we found that RASAL2 expression was decreased by 75%–100% in 20% of human breast tumors as compared to adjacent normal mammary tissue (Figures 6B and 6C). These results confirm our findings in breast cancer cell lines, suggesting that RASAL2 expression is lost or suppressed in a significant fraction of human breast cancers at a frequency that is much greater than indicated by mutation analysis alone. More importantly, however, low RASAL2 protein levels were significantly associated with metastasis (Figures 6C and 6D,  $p = 0.006$ ).

Because the cell line analysis indicated that RASAL2 expression was low or undetectable in a subset of luminal breast cancer cell lines (Figure 1), and mouse modeling studies further demonstrated that RASAL2 loss promoted the metastasis of luminal tumors, we evaluated RASAL2 expression in different breast cancer subtypes. Human breast cancers can be molecularly classified into five distinct subtypes: basal-like, HER2-positive, luminal A, luminal B, and normal breast-like (Perou et al., 2000; Sorlie et al., 2001; Hu et al., 2006). Notably, molecular subtype association analysis of transcriptional profiles from primary breast cancers revealed that RASAL2 expression was low in luminal B breast cancers; 50% of luminal B tumors expressed the lowest levels of RASAL2, consistent with a potential role for RASAL2 loss in this subtype (Figures 6E and 6F). Moreover, low RASAL2 expression was also associated with both increased tumor recurrence (Figure 6G, log rank  $p = 0.0133$ ) and decreased overall survival (Figure 6H, log rank  $p = 0.0131$ ) in patients with luminal B cancers. Finally, using publicly available TCGA methylation 450 data, we found that two CpG sites in the RASAL2 promoter region are differentially methylated in primary breast tumors. Specifically, RASAL2 promoter methylation is enriched in luminal B tumors ( $p < 0.05$ , Mann-Whitney U test). These luminal B tumors also showed the lowest expression of RASAL2. Notably, both sites exhibit a significant increase in methylation when comparing luminal B samples with the lowest expression of RASAL2 (bottom 33%) to luminal B samples with highest expression (top 33%) ( $p < 0.05$ , t test). Taken together, cellular, xenograft, mouse modeling, and human tumor studies suggest that RASAL2 loss promotes breast cancer development and metastasis and may play a particularly important role in the progression of luminal B tumors. In this context, the general lack of KRAS amplifications in luminal tumors but their frequent occurrence in basal-like breast cancers is notable (Cancer Genome Atlas Network, 2012), suggesting that RASAL2 suppression may provide an alternative mechanism of Ras activation in the luminal subtype. Nevertheless, although the models used in this study provide functional evidence to support a role for RASAL2 inactivation in luminal B tumors, these data do not preclude its involvement in a fraction of other subtypes.

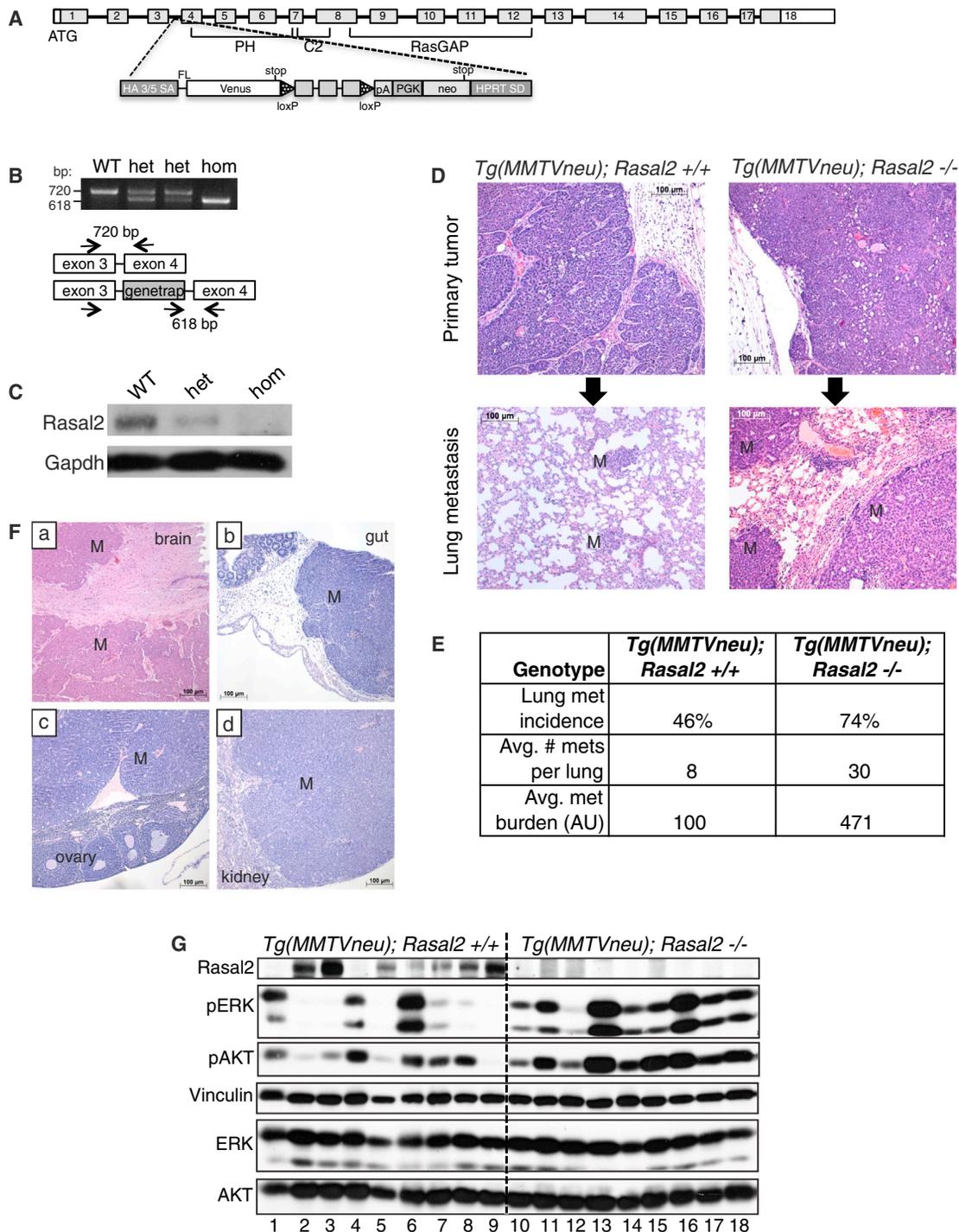
### Rasal2 Mutations Promote Tumor Development and Widespread Metastasis in p53 Mutant Mice

To determine whether RASAL2 inactivation might also contribute to the development of other sporadic tumors, *Rasal2*<sup>-/-</sup> mice were crossed to mice mutant for p53, one of the most commonly inactivated tumor suppressors in human cancer (Vousden and Lane, 2007). *Trp53* mutant mice develop a spectrum of lymphomas and sarcomas, and some carcinomas arise in heterozygotes (Donehower et al., 1992; Jacks et al., 1994). In addition to the classical tumors observed in *Trp53* mutant mice, *Rasal2/Trp53* compound mutant mice developed several other lesions that were not found in *Trp53* mutant controls, historically or in our cohort (Figure 7A). Specifically, *Rasal2/Trp53* mutant mice developed hepatocellular carcinomas and other liver tumors, colonic adenomas, and oral and stomach tumors (Figure 7A). These findings are of particular interest because RASAL2 mutations have been found in related human cancers, namely, hepatocellular carcinoma, colorectal carcinoma, head and neck squamous cell carcinoma, and stomach cancer (Table S2). Notably, the *Rasal2*<sup>-/-</sup>; *p53*<sup>-/-</sup> tumors exhibited higher levels of pERK as compared to *Rasal2*<sup>+/-</sup>; *p53*<sup>-/-</sup> tumors (Figure 7B).

However, the most striking phenotype in *Rasal2/Trp53* mutant mice was that *Rasal2* loss potentially promoted metastasis. The *Trp53*<sup>-/-</sup> mice, either with or without functional Rasal2, typically died from the primary tumor, which was frequently lymphoma. Nevertheless, 60% of the solid tumors that developed in *Rasal2*<sup>+/-</sup>; *Trp53*<sup>+/-</sup> mice and 83% of the solid tumors from *Rasal2*<sup>-/-</sup>; *Trp53*<sup>+/-</sup> mice were metastatic, as compared to 18% of tumors in *Trp53*<sup>+/-</sup> mice (Figure 7C,  $p = 0.003$ ). Specifically, *Rasal2/Trp53* mutant animals developed highly metastatic mammary adenocarcinomas, hepatocellular carcinomas, lung adenocarcinomas, and various sarcomas, again tumor types in which RASAL2 mutations have been detected in humans (Figure 7D). Thus, these findings further underscore the role of RASAL2 loss as a driver of metastasis and suggest that its inactivation may play a role in the progression of breast and other human cancers.

### DISCUSSION

The Ras pathway plays a well-established role in cancer (Downward, 2003). However, the primary mechanism(s) by which Ras becomes activated in breast cancers has remained elusive. Here, we report that RASAL2, which encodes a RasGAP, functions as a tumor and metastasis suppressor in breast and other cancers. Specifically, we have shown that loss-of-function mutations in RASAL2 are found in human breast cancers and other tumor types; however, like other RasGAP genes RASAL2 appears to be more frequently inactivated by nongenetic mechanisms, and it is substantially repressed in at least 20% of primary human breast cancers. We also showed that RASAL2 ablation promotes tumor growth and progression in two different human xenograft models, whereas RASAL2 reconstitution suppresses mammary tumor growth. Notably, RASAL2 mutations activate Ras and dramatically enhance metastasis in a genetically engineered mouse model of luminal mammary cancer. RASAL2 mutations also cooperate with p53 mutations to promote the development and metastasis of several tumor types, including mammary tumors, in a second mouse model. Finally, we show



**Figure 5. Loss of *Rasal2* Promotes Metastasis and Ras Activation in a Genetically Engineered Mouse Model of Breast Cancer**

(A) Schematic of *Rasal2* genomic locus and pNMDi4 genetrapp cassette. Unshaded regions in exons 1 and 18 mark 5' and 3' UTRs, respectively. Known domains of *Rasal2* are noted (PH, C2, and RasGAP). See the [Experimental Procedures](#) for detailed description of pNMDi4. The genetrapp cassette targets the third intron of *Rasal2*.

(B) Genotyping of *Rasal2* mice to distinguish wild-type (WT), heterozygous mutant (het), and homozygous mutant (hom).

(C) Western blot confirming loss of *Rasal2* protein in genetrapp animals (mammary gland tissue). WT, wild-type; het, heterozygous; hom, homozygous mutant.

(D) Top: H&E images of primary mammary adenocarcinomas from *MMTVneu; Rasal2<sup>+/+</sup>* and *MMTVneu; Rasal2<sup>-/-</sup>* animals. M, metastases. Bottom: H&E images of lung metastases from *MMTVneu; Rasal2<sup>+/+</sup>* and *MMTVneu; Rasal2<sup>-/-</sup>* animals. M, metastases.

(E) Lung metastasis burden in *MMTVneu; Rasal2<sup>+/+</sup>* and *MMTVneu; Rasal2<sup>-/-</sup>* animals. Lung metastasis incidence: percent of tumor-bearing females with lung metastases at sacrifice ( $p = 0.05$ ;  $n = 24$  *MMTVneu; Rasal2<sup>+/+</sup>*,  $n = 23$  *MMTVneu; Rasal2<sup>-/-</sup>*). Average number of lung metastases per animal: counted per

(legend continued on next page)

that low RASAL2 levels are associated with metastasis in human breast cancer.

Notably, the lowest RASAL2 messenger RNA (mRNA) expression levels are most frequently observed in luminal B human breast cancers and are associated with recurrence and reduced survival of patients with this tumor subtype. Collectively, these data suggest that RASAL2 loss plays a causal role in breast cancer pathogenesis. Whereas the breast cancer xenograft studies and the overall increase in tumor incidence in *Rasal2/p53* mice suggest that RASAL2 may play a role in primary tumor development, the dramatic metastatic phenotype in *Rasal2/MMTVneu* and *Rasal2/p53* mutant animals demonstrates a role for RASAL2 loss in metastasis. Similarly, the Ras pathway has been shown to play a role in both primary tumor development and metastasis, depending on context. As such, we hypothesize that RASAL2 inactivation may play a role in one or both processes, depending on the presence of other mutations in a given tumor. Nevertheless, the human breast cancer data presented in this study suggest that RASAL2 loss may play a more prominent role in progression and metastasis in this tumor type.

It should be noted that although RAS mutations are rare in breast cancer, they do occur. Moreover, amplifications of wild-type RAS are frequently observed in basal breast cancers, the most aggressive subtype of human breast cancer, underscoring the connection between Ras activation and breast cancer progression (Cancer Genome Atlas Network, 2012). Our data indicate that overall RASAL2 is suppressed or lost in at least 20% of human breast cancers. However, expression and DNA methylation analysis of different breast cancer subtypes suggests that RASAL2 loss may play a particularly important role in the progression of luminal B tumors. The observation that RASAL2 ablation promotes metastasis in a mouse model of luminal tumors provides important functional data to support this conclusion. In this respect, the fact that luminal B tumors have poorer outcomes than luminal A tumors is notable; however, the mechanism(s) that drive the progression of these tumors is largely unknown. Our data suggest that RASAL2 loss/suppression may play a causal role in the progression of this subtype, although these observations do not preclude its potential involvement in other subtypes.

Finally, whereas RASAL2 is mutated in breast cancer and in other human tumors, it appears to be more commonly inactivated via nonmutational mechanisms. Notably, the other Ras-GAP tumor suppressors, *NF1* and *DAB2IP*, are also inactivated by both genetic and several nongenetic mechanisms (McGill-Cuddy et al., 2009; Min et al., 2010). Similarly, *PTEN* and *INPP4B*, two other tumor suppressors that negatively regulate an overlapping set of signals, are also suppressed by multiple mechanisms in cancer, some of which have not yet been elucidated (Gewinner et al., 2009; Song et al., 2012). As such, loss of PTEN protein expression, rather than mutational status or copy number, is often evaluated in clinical samples during clinical trials and for pathological staging (Thomas et al., 2004). The observation

that RASAL2 loss plays a causal role in breast cancer progression and metastasis in animal models and that RASAL2 expression is lowest in primary human tumors that ultimately progress or recur, suggests that RASAL2 could be useful as a prognostic biomarker, in at least a subset of breast cancers, such as luminal B tumors. Regardless, these studies have identified an important tumor suppressor involved in breast cancer progression and have revealed an alternative mechanism by which Ras becomes activated in this disease.

## EXPERIMENTAL PROCEDURES

### Cell Culture and DNA Constructs

MEFs were immortalized as described previously (Johannessen et al., 2005). MCF7, MCF10A, and MDA-MB-361 cells were purchased from American Type Culture Collection. BT549, HS578T, MDA-MB-231, MDA-MB-453, MDA-MB-468, SKBR3, T47D, and ZR-75-1 cells were obtained from Dr. William Hahn (Dana-Farber Cancer Institute). SUM149PT, SUM159PT, SUM1315MO, and BT474 cells were obtained from Dr. Frank McCormick (University of California, San Francisco). CAMA1 cells were obtained from Dr. Marcia Haigis (Harvard Medical School). MCF10ADCIS were provided by Dr. Fred Miller (Karmanos Cancer Institute) and the Polyak laboratory.

shRNAs from the RNAi Consortium (Broad Institute, MIT) with the following sequences were utilized: *NF1* shRNA (5'-TTATAAATAGCCTGGAAAAGG-3'), RASAL2 shRNA1 (5'-CCCTCGTTCCTGCTGATAT-3'), RASAL2 shRNA2 (5'-GCCTCCACCTCTTCATAGTA-3'), *KRAS* shRNA (5'-CAGTTGAGACCTTCTAATTGG-3'), and *HRAS* shRNA (5'-GACGTGCCTGTTGGACATCCT-3'). A scrambled shRNA was purchased from Addgene (5'-CCTAAGGTAA GTCGCCCTCG-3'). A RASAL2-targeting shRNA was cloned into the pLKO vector (shRNA3) (5'-ATGGAGTGCAATAGGACATTG-3'). The Mammalian Gene Collection fully sequenced human RASAL2 complementary DNA (cDNA) was purchased from Open Biosystems (cat. number MHS4426-99623118) and was cloned into the pHAGE-N-Flag-HA lentiviral expression vector (Dr. J. Wade Harper, Harvard Medical School) for expression in cell lines. Infections and transfections were performed as described in the Supplemental Experimental Procedures.

Proliferation, soft agar, migration, and in vitro invasion assays are described in the Supplemental Experimental Procedures.

### Xenograft Assays

Female nude and NOD/SCID mice were purchased from Charles River Laboratories (cat. numbers 088 and 394, respectively) for subcutaneous xenograft experiments. Cells were injected with Matrigel (BD Biosciences, cat. number 354234) as follows: MCF10ADCIS ( $1 \times 10^5$  cells, 50% matrigel, nude mice), MDA-MB-231 ( $1 \times 10^6$  cells, 50% matrigel, nude mice), and CAMA1 ( $2 \times 10^6$  cells, 50% matrigel, NOD/SCID mice). For mammary fat pad orthotopic xenograft experiments,  $1 \times 10^6$  MDA-MB-361 cells were injected in 50% matrigel bilaterally into the fourth mammary glands of female NOD/SCID mice (Jackson Laboratories). Tumor size was measured by caliper, and tumor volume was calculated using the formula of volume = (length  $\times$  width<sup>2</sup>)  $\times$   $\pi$ /6.

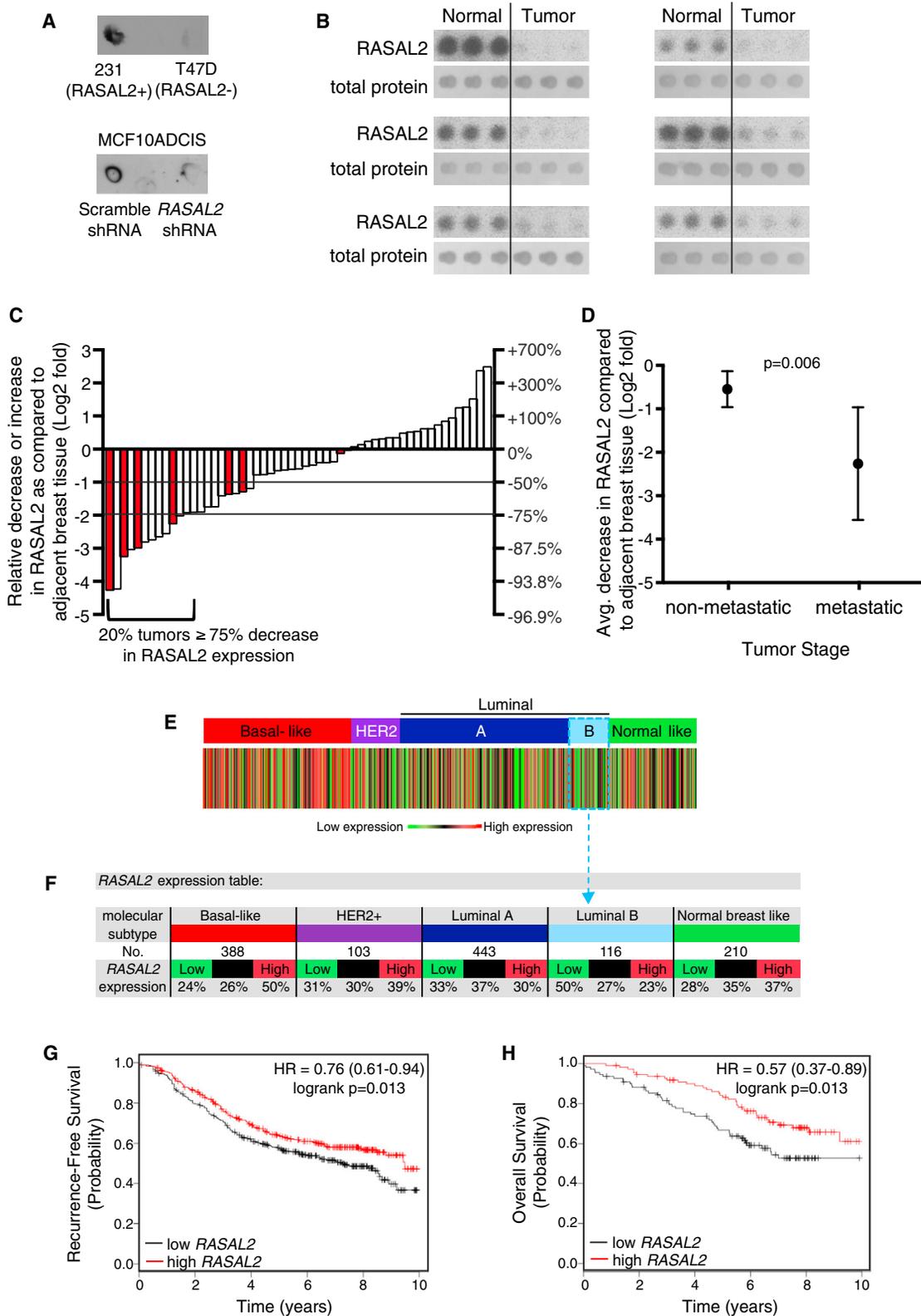
### MCF10ADCIS Invasion Assay

RASAL2 expression was ablated in MCF10ADCIS cells using two distinct shRNAs, and loss of RASAL2 expression was confirmed by immunoblot. Female nude (nu/nu) mice were injected subcutaneously with 100,000 shRASAL2 or shScramble control cells. Five or six tumors of each genotype (Scramble shRNA, RASAL2 shRNA1, or RASAL2 shRNA2) were harvested after 3 weeks, fixed, and stained with hematoxylin and eosin (H&E) to assess tumor morphology. A pathologist scored a tumor as an invasive carcinoma if it was

representative section of lungs for each tumor-bearing female ( $p = 0.04$ ). Average metastasis burden per animal: average total area of metastasis in a representative section of lung for each tumor-bearing female (arbitrary units;  $p = 0.04$ ). See also Figure S2.

(F) H&E images of metastases to brain (a), gut (b), ovary (c), and kidney (d) in compound tumor-bearing females. M indicates regions of metastasis.

(G) Western blot analysis of phospho-ERK (pERK) and phospho-AKT (pAKT) levels in primary mammary tumors from *MMTVneu*; *Rasal2*<sup>+/+</sup> animals (numbers 1–9) and *MMTVneu*; *Rasal2*<sup>-/-</sup> animals (numbers 10–18).



**Figure 6. RASAL2 Expression Is Lost/Low in Primary Human Breast Cancers, and Low Levels Are Associated with Metastasis and Recurrence**

(A) RASAL2 dot blot of whole-cell RIPA extracts from human breast cancer cell lines with high or low RASAL2 expression (MDA-MB-231 “231” and T47D, respectively) (top) or MCF10ADCIS cells infected with control or RASAL2-targeting shRNAs (bottom).

(legend continued on next page)

accompanied by an obvious loss in its well-circumscribed appearance and a disruption of the myoepithelium and basement membrane.

#### Rasal2 Mutant Mice

All animal procedures were approved by the Center for Animal and Comparative Medicine at Harvard Medical School in accordance with the National Institutes of Health Guidelines for the Care and Use of Laboratory Animals and the Animal Welfare Act.

A mouse embryonic stem cell line in which the pNMDi4 genetrap cassette targets *Rasal2* was purchased from the Toronto Centre for Phenogenomics/Canadian Mouse Mutant Repository (clone CMHD 463C12). The pNMDi4 genetrap cassette contains the following elements as depicted in Figure 4A: HA 3/5 SA (splice acceptor), FL (flexible linker), Venus (enhanced yellow fluorescent gene) with stop codon, loxP sites, pA (polyadenylation signal), PGK (promoter), neo (neomycin resistance gene) with stop codon, and HPRT SD (splice donor). The neomycin resistance gene stop codon prevents translation of 3' exons. The presence of the genetrap cassette within the third intron of *Rasal2* was confirmed using cDNA PCR. Chimeric mice were generated and crossed to C57BL/6-E animals (Charles River Laboratories), and pups were tested for presence of the genetrap. Two additional copies of the genetrap cassette elsewhere in the genome were discovered in the mouse ES cell line and genetrap mice. Genetrap-positive mice were crossed to wild-type animals, and Southern blotting was used to identify pups that had the genetrap cassette only within the *Rasal2* locus (data not shown). These animals were used as founders for all cohorts and subsequent crosses.

#### Rasal2 Genotyping

Primers for PCR of the genetrap cassette were NMD.F (5'-CATGGTCCTGCTGGAGTTC-3') and NMD.R (5'-TGCCTTTAGACCTTTTGTGG-3'). Total RNA was extracted from homogenized tails using QiaShredder and RNeasy kits (QIAGEN), and cDNA was synthesized using qScript cDNA Synthesis Kit (Quanta). PCR was performed on cDNA with primers NeoL (5'-GCTATCAGGACATAGCGTTGGCTAC-3'), GT463C12\_F3 (5'-TCGGATCCTTCTGGAGTCAG-3'), and GT463C12\_R1 (5'-CTCTCTCGGAGGCGAGAGCTA-3') to detect wild-type (F3/R1) and mutant (NeoL/R1) transcripts.

#### Compound Mutant Mice

*Rasal2* genetrap mice were crossed to FVB/N-Tg(MMTVneu)202Mul/J mice (Jackson Laboratories, cat. number 002376) (Guy et al., 1992) or to B6.129S2-*Trp53*<sup>Tm1Tyj</sup>/J mice (Jackson Laboratories, cat. number 002101) (Jacks et al., 1994). Cohorts of *Rasal2* mutant mice and controls were on a 129-enriched background (75% 129SvImJ, 25% C57BL6). Cohorts of *Trp53*; *Rasal2* compound mice and controls were on a mixed 129/B6 background (62.5% C57BL6, 37.5% 129SvImJ). Cohorts of *MMTVneu*; *Rasal2* compound mice and controls were on a background of 56% 129SvImJ, 25% FVB, and 19% C57BL6.

#### Protein Lysates and Western Blot Analyses

Protein extracts were isolated from cells or homogenized tissue in 1% SDS boiling lysis buffer. Ras-GTP levels were determined using a Ras Activation Assay Kit (EMD Millipore). The following antibodies were used for immunoblots: actin (Sigma, cat. number A2066), phospho-AKT (Ser473, Cell Signaling, cat. number 4060), AKT (Cell Signaling, cat. number 9272), ER (Thermo, cat. number R9101-SO), phospho-ERK (Thr202/Thr204, Cell Signaling, cat. num-

ber 4370), ERK (Cell Signaling, cat. number 9102), GAPDH (Cell Signaling, cat. number 2118), HA (Covance, cat. number MMS-101P), HER2 (Cell Signaling, cat. number 2242), NF1 (UP69 C-terminal polyclonal antibody) (McGillicuddy et al., 2009), p120RasGAP (BD Transduction Laboratories, cat. number 610040), HRas (Santa Cruz, cat. number SC-520), KRas (Santa Cruz, cat. number SC-30), panRas (Upstate, cat. number 05-516),  $\alpha$ -tubulin (Sigma, cat. number T5168),  $\beta$ -tubulin (Sigma, cat. number T4026), and Vinculin (Cell Signaling, cat. number 4650). A peptide antigen (NP\_773793 amino acids 1111–1130) was used to generate and affinity purify an anti-RASAL2 rabbit polyclonal antibody (Covance ImmunoTechnologies). For RASAL2, reconstitution studies cells were typically plated in 5% serum overnight 36 hr posttransfection. pERK and ERK levels were assessed by western blot and quantified using ImageJ software.

Immunohistochemistry was performed as described in the Supplemental Experimental Procedures.

#### Mouse Tumor and Tissue Analysis

Tumors and tissues were fixed in buffered formalin, stored in 70% ethanol, paraffin embedded, and sectioned. Sections were stained with hematoxylin and eosin.

#### Human Tumor Lysate Array Analysis

Qualitative breast cancer tumor lysate arrays were purchased from Protein Biotechnologies (cat. number PMA2-001-L). Samples were deidentified and are not considered human subject research. Arrays were probed with the affinity purified RASAL2 antibody. The RASAL2 antibody was validated for this assay using tumor lysate arrays by probing nitrocellulose membranes spotted with 1  $\mu$ g/ $\mu$ l RIPA lysates from human breast cancer cell lines with or without RASAL2. Developed film was scanned and quantified using ImageJ software. Arrays were stained with Colloidal Gold and scanned, and total protein was quantified using ImageJ. RASAL2 levels in each spot were normalized to the Colloidal Gold level in the same spot. Triplicate spots were averaged, and the ratio between the tumor normalized triplicate and normal normalized triplicate was calculated and reported as a Log<sub>2</sub> fold change value.

#### Molecular Subtype Association and Survival Analysis

Gene expression correlations targeted analysis was applied on published genomic data on patients classified in the same molecular subtype with the six molecular subtype predictors (Sørlie et al., 2001; Hu et al., 2006), using GenExMiner as previously described (Jézéquel et al., 2012). Samples were deidentified and are not considered human subject research. A gene expression map was determined by molecular subtype predictors (single sample predictors [SSPs] and/or subtype clustering models [SCMs]). A gene expression table was also provided for robust classifications, indicating for each subtype the proportion of patient with low, intermediate, and high gene expression; gene expression values were split in order to form three equal groups.

Gene expression data and relapse-free and overall survival information were analyzed as previously described (Györfy et al., 2010). Data were downloaded from GEO (Affymetrix HGU133A and HGU133+2 microarrays), EGA, and TCGA. The background database integrates gene expression and clinical data simultaneously. To analyze the prognostic value of RASAL2, the patient samples are split into two groups according to median expression of RASAL2.

(B) Dot blot images from human breast tumor lysate array. Six sets of RASAL2 and total protein stains are shown. Each set contains triplicate spots of tumor lysate (right) and triplicate spots of paired normal tissue lysate (left).

(C) Quantification of RASAL2 expression in tumor lysate arrays. Each bar depicts the change in RASAL2 expression in one sample as compared to the sample's matched normal control as described in the Experimental Procedures. Red bars indicate metastatic samples.

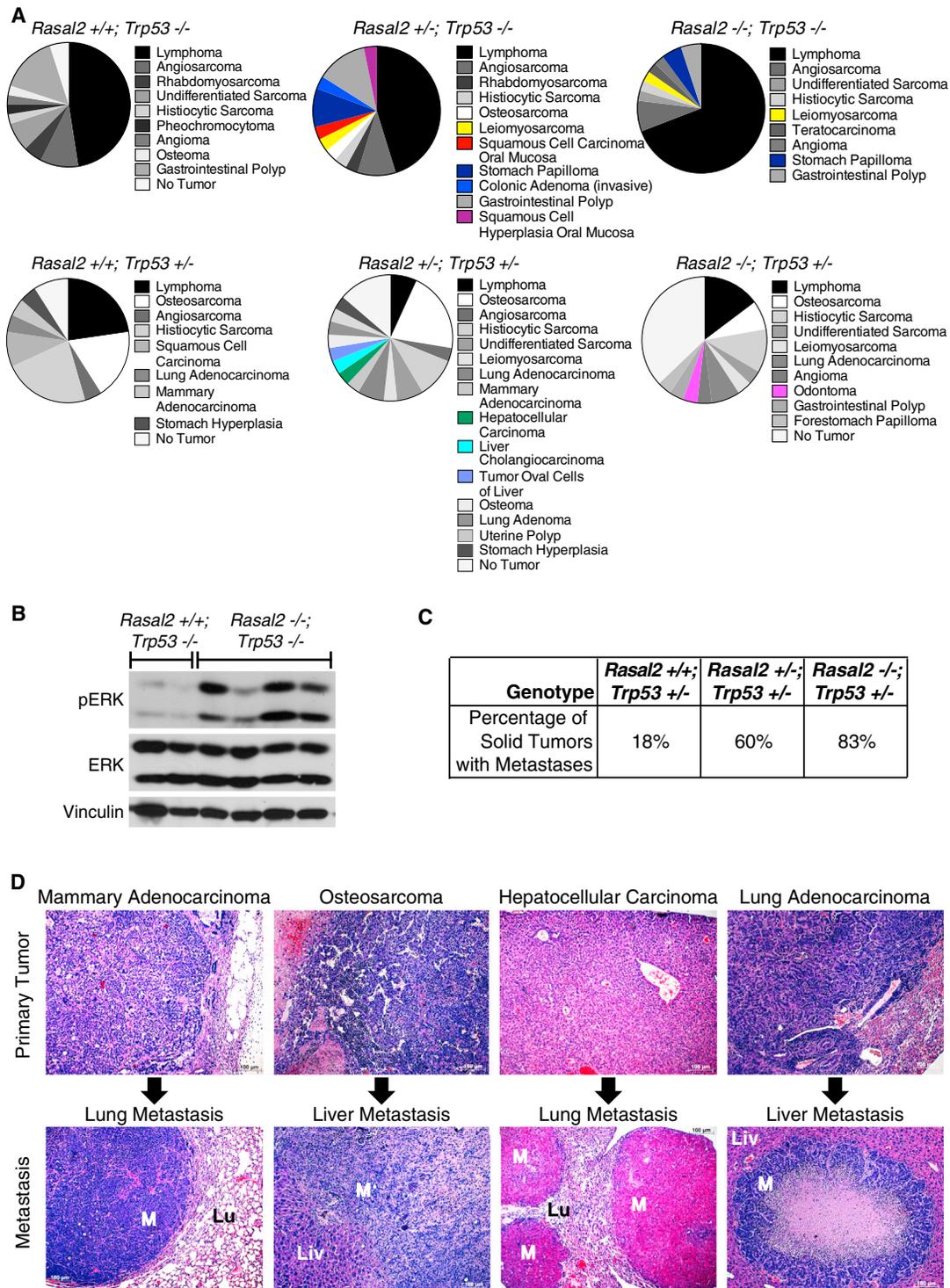
(D) RASAL2 protein expression in tumor versus normal in nonmetastatic (stages I, II, and III) versus metastatic (stage IV) tumors. Graph shows the Log<sub>2</sub> fold change in RASAL2 protein expression in tumor versus normal. Data are reported as average  $\pm$  95% CI.  $p = 0.006$ .

(E) Heatmap of RASAL2 gene expression as a function of robust molecular subtype predictor classification, which is based on patients classified in the same tumor subtype. Percentages of tumors with high, intermediate, and low RASAL2 expression per molecular subtype are given in the gene expression table.

(F) RASAL2 expression table. For each breast cancer subtype, the number of samples and percentage of samples with low, intermediate, or high RASAL2 mRNA expression are indicated.

(G) Kaplan-Meier curve showing recurrence-free survival of luminal B tumors with high or low RASAL2 expression (log rank  $p = 0.013$ ).

(H) Kaplan-Meier curve showing overall survival of luminal B tumors with high or low RASAL2 expression (log rank  $p = 0.013$ ).



**Figure 7. *Rasal2*, *Trp53* Compound Mutant Mice Develop Highly Metastatic Tumors**

(A) Phenotypes in *Rasal2/Trp53* compound mutant mice. Pie charts display the array of phenotypes in each genotype. Phenotypes observed in *Rasal2* mutant compound mice, but not observed in control cohorts, are shown in color.  $n = 21$  *Rasal2*<sup>+/+</sup>; *Trp53*<sup>-/-</sup>, 18 *Rasal2*<sup>+/-</sup>; *Trp53*<sup>-/-</sup>, 31 *Rasal2*<sup>-/-</sup>; *Trp53*<sup>-/-</sup>, 16 *Rasal2*<sup>+/+</sup>; *Trp53*<sup>+/-</sup>, 21 *Rasal2*<sup>+/-</sup>; *Trp53*<sup>+/-</sup>, 21 *Rasal2*<sup>-/-</sup>; *Trp53*<sup>+/-</sup>.

(B) Western blot analysis of phospho-ERK (pERK) levels in primary tumors from *Rasal2*<sup>+/+</sup>; *Trp53*<sup>-/-</sup> and *Rasal2*<sup>-/-</sup>; *Trp53*<sup>-/-</sup> compound mice.

(C) Percentage of metastatic solid tumors in *Rasal2*<sup>+/+</sup>, <sup>+/-</sup>, and <sup>-/-</sup>; *Trp53*<sup>+/-</sup> compound mice. Increased metastasis in compound animals is statistically significant ( $p = 0.003$ ). The paucity of metastatic solid tumors in *Trp53*<sup>+/-</sup> mice is supported by historical data.

(D) Images of metastatic lesions (bottom) and the corresponding primary tumors (top). From left to right: Mammary adenocarcinoma and lung metastasis, osteosarcoma and liver metastasis, hepatocellular carcinoma and lung metastasis, and lung adenocarcinoma and liver metastasis. M, metastasis; Lu, lung; Liv, liver.

## SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, two figures, and three tables and can be found with this article online at <http://dx.doi.org/10.1016/j.ccr.2013.08.004>.

## ACKNOWLEDGMENTS

This work was supported by grants from the Ludwig Center at Dana-Farber/Harvard Cancer Center (to K.C.), the National Cancer Institute (R01 CA111754 to K.C.), the Department of Defense (W81XWH-11-1-0140 to S.K.M.), and the National Institutes of Health National Heart, Lung, and Blood Institute (T32-HL07623 to S.K.M.).

Received: March 8, 2013

Revised: July 7, 2013

Accepted: August 6, 2013

Published: September 9, 2013

## REFERENCES

- Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R., and Wooster, R. (2004). The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br. J. Cancer* *91*, 355–358.
- Bernards, A. (2003). GAPs galore! A survey of putative Ras superfamily GTPase activating proteins in man and Drosophila. *Biochim. Biophys. Acta* *1603*, 47–82.
- Cancer Genome Atlas Network. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* *490*, 61–70.
- Cancer Genome Atlas Research Network. (2008). Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* *455*, 1061–1068.
- Cawthon, R.M., Weiss, R., Xu, G.F., Viskochil, D., Culver, M., Stevens, J., Robertson, M., Dunn, D., Gesteland, R., O'Connell, P., et al. (1990). A major segment of the neurofibromatosis type 1 gene: cDNA sequence, genomic structure, and point mutations. *Cell* *62*, 193–201.
- Ding, L., Getz, G., Wheeler, D.A., Mardis, E.R., McLellan, M.D., Cibulskis, K., Sougnez, C., Greulich, H., Muzny, D.M., Morgan, M.B., et al. (2008). Somatic mutations affect key pathways in lung adenocarcinoma. *Nature* *455*, 1069–1075.
- Donehower, L.A., Harvey, M., Slagle, B.L., McArthur, M.J., Montgomery, C.A., Jr., Butel, J.S., and Bradley, A. (1992). Mice deficient for p53 are developmentally normal but susceptible to spontaneous tumours. *Nature* *356*, 215–221.
- Dote, H., Toyooka, S., Tsukuda, K., Yano, M., Ouchida, M., Doihara, H., Suzuki, M., Chen, H., Hsieh, J.-T., Gazdar, A.F., and Shimizu, N. (2004). Aberrant promoter methylation in human DAB2 interactive protein (hDAB2IP) gene in breast cancer. *Clin. Cancer Res.* *10*, 2082–2089.
- Downward, J. (2003). Targeting RAS signalling pathways in cancer therapy. *Nat. Rev. Cancer* *3*, 11–22.
- Gewinner, C., Wang, Z.C., Richardson, A., Teruya-Feldstein, J., Etemadmoghadam, D., Bowtell, D., Barretina, J., Lin, W.M., Rameh, L., Salmena, L., et al. (2009). Evidence that inositol polyphosphate 4-phosphatase type II is a tumor suppressor that inhibits PI3K signaling. *Cancer Cell* *16*, 115–125.
- Györfy, B., Lanczky, A., Eklund, A.C., Denkert, C., Budczies, J., Li, Q., and Szallasi, Z. (2010). An online survival analysis tool to rapidly assess the effect of 22,277 genes on breast cancer prognosis using microarray data of 1,809 patients. *Breast Cancer Res. Treat.* *123*, 725–731.
- Guy, C.T., Webster, M.A., Schaller, M., Parsons, T.J., Cardiff, R.D., and Muller, W.J. (1992). Expression of the neu protooncogene in the mammary epithelium of transgenic mice induces metastatic disease. *Proc. Natl. Acad. Sci. USA* *89*, 10578–10582.
- Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., et al. (2007). Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol.* *8*, R76.
- Hollestelle, A., Elstrodt, F., Nagel, J.H.A., Kallemeijn, W.W., and Schutte, M. (2007). Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Mol. Cancer Res.* *5*, 195–201.
- Hölzel, M., Huang, S., Koster, J., Ora, I., Lakeman, A., Caron, H., Nijkamp, W., Xie, J., Callens, T., Asgharzadeh, S., et al. (2010). NF1 is a tumor suppressor in neuroblastoma that determines retinoic acid response and disease outcome. *Cell* *142*, 218–229.
- Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., et al. (2006). The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* *7*, 96.
- Hu, M., Yao, J., Carroll, D.K., Weremowicz, S., Chen, H., Carrasco, D., Richardson, A., Violette, S., Nikolskaya, T., Nikolsky, Y., et al. (2008). Regulation of in situ to invasive breast carcinoma transition. *Cancer Cell* *13*, 394–406.
- Jacks, T., Remington, L., Williams, B.O., Schmitt, E.M., Halachmi, S., Bronson, R.T., and Weinberg, R.A. (1994). Tumor spectrum analysis in p53-mutant mice. *Curr. Biol.* *4*, 1–7.
- Jézéquel, P., Campone, M., Gouraud, W., Guérin-Charbonnel, C., Leux, C., Ricolleau, G., and Campion, L. (2012). bc-GenExMiner: an easy-to-use online platform for gene prognostic analyses in breast cancer. *Breast Cancer Res. Treat.* *131*, 765–775.
- Johannessen, C.M., Reczek, E.E., James, M.F., Brems, H., Legius, E., and Cichowski, K. (2005). The NF1 tumor suppressor critically regulates TSC2 and mTOR. *Proc. Natl. Acad. Sci. USA* *102*, 8573–8578.
- Kamangar, F., Dores, G.M., and Anderson, W.F. (2006). Patterns of cancer incidence, mortality, and prevalence across five continents: defining priorities to reduce cancer disparities in different geographic regions of the world. *J. Clin. Oncol.* *24*, 2137–2150.
- Karnoub, A.E., and Weinberg, R.A. (2008). Ras oncogenes: split personalities. *Nat. Rev. Mol. Cell Biol.* *9*, 517–531.
- Kim, I.S., and Baek, S.H. (2010). Mouse models for breast cancer metastasis. *Biochem. Biophys. Res. Commun.* *394*, 443–447.
- Krauthammer, M., Kong, Y., Ha, B.H., Evans, P., Bacchiocchi, A., McCusker, J.P., Cheng, E., Davis, M.J., Goh, G., Choi, M., et al. (2012). Exome sequencing identifies recurrent somatic RAC1 mutations in melanoma. *Nat. Genet.* *44*, 1006–1014.
- 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.; kConFab. (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. *Nat. Med.* *15*, 907–913.
- Maertens, O., Johnson, B., Hollstein, P., Frederick, D.T., Cooper, Z.A., Messaien, L., Bronson, R.T., McMahon, M., Granter, S., Flaherty, K.T., et al. (2012). Elucidating distinct roles for NF1 in melanomagenesis. *Cancer Discov.* *3*, 338–349.
- McGillcuddy, L.T., Fromm, J.A., Hollstein, P.E., Kubek, S., Beroukhim, R., De Raedt, T., Johnson, B.W., Williams, S.M.G., Nghiemphu, P., Liao, L.M., et al. (2009). Proteasomal and genetic inactivation of the NF1 tumor suppressor in gliomagenesis. *Cancer Cell* *16*, 44–54.
- Miller, F.R., Santner, S.J., Tait, L., and Dawson, P.J. (2000). MCF10DCIS.com xenograft model of human comedo ductal carcinoma in situ. *J. Natl. Cancer Inst.* *92*, 1185–1186.
- Min, J., Zaslavsky, A., Fedele, G., McLaughlin, S.K., Reczek, E.E., De Raedt, T., Guney, I., Strohlic, D.E., Macconail, L.E., Beroukhim, R., et al. (2010). An oncogene-tumor suppressor cascade drives metastatic prostate cancer by coordinately activating Ras and nuclear factor-kappaB. *Nat. Med.* *16*, 286–294.
- Mueller, H., Flury, N., Eppenberger-Castori, S., Kueng, W., David, F., and Eppenberger, U. (2000). Potential prognostic value of mitogen-activated protein kinase activity for disease-free survival of primary breast cancer patients. *Int. J. Cancer* *89*, 384–388.
- Mueller, C., Liotta, L.A., and Espina, V. (2010). Reverse phase protein microarrays advance to use in clinical trials. *Mol. Oncol.* *4*, 461–481.

- Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., Gallia, G.L., et al. (2008). An integrated genomic analysis of human glioblastoma multiforme. *Science* *321*, 1807–1812.
- Perou, C.M., Sørlie, 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*, 747–752.
- Pylayeva-Gupta, Y., Grabocka, E., and Bar-Sagi, D. (2011). RAS oncogenes: weaving a tumorigenic web. *Nat. Rev. Cancer* *11*, 761–774.
- Shah, S.P., Roth, A., Goya, R., Oloumi, A., Ha, G., Zhao, Y., Turashvili, G., Ding, J., Tse, K., Haffari, G., et al. (2012). The clonal and mutational evolution spectrum of primary triple-negative breast cancers. *Nature* *486*, 395–399.
- Sivaraman, V.S., Wang, H., Nuovo, G.J., and Malbon, C.C. (1997). Hyperexpression of mitogen-activated protein kinase in human breast cancer. *J. Clin. Invest.* *99*, 1478–1483.
- Sjöblom, T., Jones, S., Wood, L.D., Parsons, D.W., Lin, J., Barber, T.D., Mandelker, D., Leary, R.J., Ptak, J., Silliman, N., et al. (2006). The consensus coding sequences of human breast and colorectal cancers. *Science* *314*, 268–274.
- Song, M.S., Salmena, L., and Pandolfi, P.P. (2012). The functions and regulation of the PTEN tumour suppressor. *Nat. Rev. Mol. Cell Biol.* *13*, 283–296.
- Sørlie, 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. USA* *98*, 10869–10874.
- Thomas, G.V., Horvath, S., Smith, B.L., Crosby, K., Lebel, L.A., Schrage, M., Said, J., De Kernion, J., Reiter, R.E., and Sawyers, C.L. (2004). Antibody-based profiling of the phosphoinositide 3-kinase pathway in clinical prostate cancer. *Clin. Cancer Res.* *10*, 8351–8356.
- von Lintig, F.C., Dreilinger, A.D., Varki, N.M., Wallace, A.M., Casteel, D.E., and Boss, G.R. (2000). Ras activation in human breast cancer. *Breast Cancer Res. Treat.* *62*, 51–62.
- Vousden, K.H., and Lane, D.P. (2007). p53 in health and disease. *Nat. Rev. Mol. Cell Biol.* *8*, 275–283.
- Weigelt, B., Peterse, J.L., and van 't Veer, L.J. (2005). Breast cancer metastasis: markers and models. *Nat. Rev. Cancer* *5*, 591–602.

## **VI.2 Using defined finger-finger interfaces as units of assembly for constructing zinc-finger nucleases.**

### **Abstract**

This paper published in *Nucleic Acids Research* was a collaboration with Scot Wolfe's lab. The paper identifies a new method of assembly for the genome editing zinc finger nuclease (ZFN) tool by using the structural similarities between adjacent zinc finger helices to contact target nucleotides. This increased the diversity of target sequences for ZFNs. The Wolfe lab provided a zinc finger that might be able to target the tumor suppressor BRCA1 and I cloned it into a eukaryotic expression vector and transfected it into 293T cells. I was able to show via a nucleic acid re-annealing bubble assay that approximately 3% of the transfected 293T cells showed base mismatch at the predicted cut site in exon 11 of the BRCA1 gene.

# Using defined finger–finger interfaces as units of assembly for constructing zinc-finger nucleases

Cong Zhu<sup>1</sup>, Ankit Gupta<sup>1,2</sup>, Victoria L. Hall<sup>1</sup>, Amy L. Rayla<sup>1</sup>, Ryan G. Christensen<sup>3</sup>, Benjamin Dake<sup>4</sup>, Abirami Lakshmanan<sup>1</sup>, Charlotte Kuperwasser<sup>4</sup>, Gary D. Stormo<sup>3</sup> and Scot A. Wolfe<sup>1,2,\*</sup>

<sup>1</sup>Program in Gene Function and Expression, University of Massachusetts Medical School, Worcester, MA, USA 01605, <sup>2</sup>Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, Worcester, MA, USA 01605, <sup>3</sup>Department of Genetics, Washington University School of Medicine, St Louis, MO, USA 63108 and <sup>4</sup>Molecular Oncology Research Institute (MORI), Tufts University School of Medicine, Boston, MA, USA 02111

Received November 7, 2012; Revised December 4, 2012; Accepted December 5, 2012

## ABSTRACT

**Zinc-finger nucleases (ZFNs) have been used for genome engineering in a wide variety of organisms; however, it remains challenging to design effective ZFNs for many genomic sequences using publicly available zinc-finger modules. This limitation is in part because of potential finger–finger incompatibility generated on assembly of modules into zinc-finger arrays (ZFAs). Herein, we describe the validation of a new set of two-finger modules that can be used for building ZFAs via conventional assembly methods or a new strategy—finger stitching—that increases the diversity of genomic sequences targetable by ZFNs. Instead of assembling ZFAs based on units of the zinc-finger structural domain, our finger stitching method uses units that span the finger–finger interface to ensure compatibility of neighbouring recognition helices. We tested this approach by generating and characterizing eight ZFAs, and we found their DNA-binding specificities reflected the specificities of the component modules used in their construction. Four pairs of ZFNs incorporating these ZFAs generated targeted lesions *in vivo*, demonstrating that stitching yields ZFAs with robust recognition properties.**

## INTRODUCTION

Zinc-finger nucleases (ZFNs) are chimeric fusions between a programmable zinc-finger array (ZFA) and the nuclease

domain of FokI (1). These artificial nucleases are powerful tools for genome modification, as they can generate a site-specific double-strand break (DSB) within the genome to promote a number of different types of genome editing (2,3). ZFNs can disrupt the function of a protein-coding gene when an imprecisely repaired DSB creates a frameshift in the coding sequence (2,3). These DSBs can also be used for the introduction of tailor-made changes to the genome by dramatically stimulating the rate of homologous recombination at a locus with an exogenously supplied donor DNA (2,3). ZFNs have been used in a variety of model and non-model organisms to facilitate reverse genetic approaches to study gene function or construct disease models for analysis (3–8). Engineered nucleases also have potential application as gene therapy-based therapeutics (9–15), where the first of these reagents are now in advanced clinical trials for treatment of AIDS (16).

The use of ZFNs has primarily been limited by the ease with which ZFAs can be created to selectively target a desired genomic region. Excluding purchase from commercial sources, selection-based approaches provide the most reliable method for creating ZFAs with novel DNA-binding specificity (17–23). Bacterial-based selection systems have somewhat simplified the process of creating ZFAs with novel specificity (24–26), but these systems still require effort on the part of end-users to generate functional constructs. Many zinc-finger proteins bind to DNA in an apparently modular fashion (27–32). Based on this supposition, a comprehensive archive of single-finger modules should enable the ready assembly of any multi-finger ZFA, where the resultant recognition site is a composite of the specificities of the incorporated finger modules (31,33–35). Using this approach, many

\*To whom correspondence should be addressed. Tel: +1 508 856 3953; Fax: +1 508 856 5460; Email: scot.wolfe@umassmed.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2013. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/3.0/>), which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

laboratories have generated ZFAs for incorporation into ZFNs (36–43). However, success rates with such modularly assembled ZFNs have typically been modest (<30%) (39–41). The inconsistency in these systems could reflect insufficient specificity or affinity of the modules within the published archives (44), or incompatibility between the assembled fingers, which has been dubbed ‘context-dependent effects’ (17,28,31,45). The primary source of this incompatibility likely resides with mismatched residues at the finger–finger interface that degrade or alter specificity (41) in some cases because of recognition overlap between neighbouring fingers (30,46–48).

The impact of interface incompatibility can be reduced by limiting the number of unproven finger–finger interfaces that are generated during assembly (2). The use of two-finger modules for ZFA assembly reduces, but does not eliminate, the number of unproven interfaces (49–51). Unproven interfaces can be eliminated entirely through the context-dependent assembly (CoDA) of two-finger modules, where three-finger ZFAs are constructed from two-finger units with a common overlapping central finger (52). CoDA-generated ZFNs have favourable success rates (~50%) *in vivo*; however, the vast majority of the modules that have been tested as ZFNs recognize NG-type (gnNGnn) dinucleotide junctions at the finger–finger interface, which are well understood and, therefore, are not the limiting interface for the expansion of ZFN targeting density. Moreover, some of the CoDA modules assigned to target non-NG junctions actually prefer NG junction sequences (50).

Recently, we described the selection of two-finger modules recognizing GRNNYG dinucleotide junctions using our bacterial one-hybrid (B1H) system (50). This approach focused on randomizing the recognition residues at the finger–finger interface in a library of two-finger modules to select optimal modules for each target sequence. The specificity of the selected modules was, subsequently, validated to identify modules with the most favourable recognition properties. Additional mutagenesis expanded the breadth of this archive to contain modules spanning 162 six base pair target sites that can be used together or in conjunction with other single-finger archives (41) for the assembly of ZFAs. This archive expands the current collection of publicly available two-finger modules, in particular those that recognize non-NG interfaces. However, their assembly into larger ZFAs remains predicated on the concatenation of modules at dinucleotide junctions that have well-defined sequence preferences (e.g. GK or AN) (50). Appropriate interface residues for many of the other dinucleotide junctions remain poorly defined. Moreover, based on our analysis of zinc-finger specificity, it seems that the appropriate choice of these recognition residues can be impacted by the identity of the residues at position +3 in each finger (Gupta and Wolfe, unpublished results), confirming the complexity in zinc-finger–DNA recognition that has been observed in other mutagenesis studies (45,53,54).

Most ZFA-assembly methods use finger archives composed of single- or multi-finger modules, where these units are delimited by the structural motif of the zinc

finger. However, if finger–finger interfaces represent critical grammar for the successful assembly of functional ZFAs, then construction units wherein this most dynamic feature of recognition is fixed may yield higher success rates. This type of assembly approach has been used in the construction of three-finger ZFAs from 1.5 finger modules by Choo and colleagues (19), where an intervening phage display selection step could be used to optimize the recognition properties of the assembled proteins. We extended this approach by choosing units of assembly with fixed elements of overlap, such that an additional selection is not required, and ZFAs containing any number of fingers could be assembled if complementary monomeric units were present in the archive. Moreover, this method allows the construction of hybrid fingers with novel specificity that is not present within the archive. To serve as an initial archive for this assembly approach, we selected a set of GANNAG two-finger modules that can be assembled into ZFAs either through the standard modular-assembly or via finger ‘stitching’. We demonstrated that ZFAs assembled through either method function robustly as ZFNs when assembled into nucleases. These results highlight the advantage of using defined finger–finger interfaces for the construction of artificial ZFAs.

## MATERIALS AND METHODS

### Animal husbandry

Zebrafish were handled according to established protocols (55) and in accordance with Institutional Animal Care and Use Committee (IACUC) guidelines of the University of Massachusetts Medical School.

### 2F-library construction

The two-finger (2F) library was designed with scheme RSDNLXX XXXNLTR, using codons VNS(+5)VNS(+6) NNW(–1)NNW(+1)NNW(+2) (V: A/C/G, N:A/C/G/T, S:G/C and W:A/T) for the five randomized residues. 2F-libraries were constructed as previously described (50). Briefly, individual F1 and F2 libraries were independently constructed via cassette mutagenesis of annealed randomized oligonucleotides into pBluescript vector containing the appropriate zinc finger backbone derived from Zif268. The 2F-library was constructed from these single-finger libraries by overlapping polymerase chain reaction (PCR) assembly. This 2F-library was then ligated into the B1H expression vector 1352-omega-UV2 between unique BssHII and Acc65I restriction enzyme sites, such that the  $\omega$ -subunit of the RNA polymerase is fused at the N-terminus of the two zinc fingers, and the Engrailed homeodomain follows the fingers at the C-terminus (Figure 1). After electroporation into bacterial cells,  $1 \times 10^8$  cells (five times the theoretical size of the library) were plated on ten 2xYT–carbenicillin plates (150 × 15 mm) and grown at 37°C for 14 h. 1352-omega-UV2 plasmids containing the 2F-library were isolated from pooled surviving colonies and used for selections.

### Zinc-finger binding site cloning and 2F-module B1H selections

The 16 GANNAG zinc-finger binding sites (ggccTAATT ACCTGANNAGGagc) were cloned between the EcoRI and NotI sites in the pH3U3-mcs reporter vector (57). The homeodomain (Engrailed) binding site TAATTA (underlined) is present 3 bp away and on the strand opposite to the zinc-finger binding site to minimize any interference between the homeodomain and the zinc fingers. Selections for 2F-modules were performed as described previously (50). The zinc-finger library (20 ng) and the reporter vector (1  $\mu$ g) containing the zinc-finger target site were co-transformed via electroporation in the selection strain that lacks endogenous expression of the  $\omega$ -subunit of RNA polymerase (US0*AhisBApyrFArpoZ*). The  $2 \times 10^7$  co-transformed cells were plated on selective NM minimal medium plates containing different concentrations of competitive inhibitor (3-aminotriazole; 3-AT) and inducer (isopropyl- $\beta$ -D-thiogalactoside; IPTG) and grown at 37°C until a moderate number of colonies were visible. Selections were performed under different stringencies for each target site by varying 3-AT or IPTG concentration to achieve about surviving 1000 colonies per plate. In some instances, it was necessary to further increase stringency by adjusting the strength of the binding site. This was accomplished by altering either the engrailed binding site from canonical sequence TAATTA to mutant sequence TAATGC or the 6 bp 2-finger module target site from GANNAG to TANNAG. Post-selection, 2F-modules from eight surviving colonies of various sizes (typically four large, two medium and two small) were sequenced to identify functional amino acid sequences for further evaluation. The success of the selection was judged by the diversity of sequences obtained from these selections, with the expectation that successful selections will converge on a small number of functional residues at the critical recognition positions.

### Cloning B1H-selected 2F modules into 3F F1-GCG constructs

To determine the binding specificities of 2F-modules, a 'GCG' binding anchor zinc finger (recognition helix: RSDTLAR) was fused at the N-terminus of the 2F-module via overlapping PCR. After overlapping PCR, the 3F-ZFA was cloned into 1352-omega-UV2 vector between the Acc65I and BamHI sites for expression as an omega fusion.

### Constrained variation-B1H method

To determine binding site specificities of 2F-modules, the constrained variation-B1H assay was performed as described previously (56). After transformation into the selection strain,  $1 \times 10^6$  cells containing the zinc-finger plasmid (1352-omega-UV2-ZFP) and the 6 bp randomized binding site library plasmid (pH3U3) were plated on selective NM minimal medium plates (100  $\times$  15 mm) containing 50  $\mu$ M of IPTG and 1 or 2 mM of 3-AT and grown at 37°C for 22–30 h. The surviving colonies were pooled, and the binding site plasmid was isolated for identification of the

functional DNA sequences. The binding site region was PCR amplified, barcoded and sequenced via Illumina sequencing, and then binding specificities were determined from these data using the log-odds method (50,56,58).

### Creating multi-finger ZFPs

All stitching finger ZFPs and six-finger traditional-modular-assembly ZFPs were created by gene synthesis through Genscript USA (Piscataway, NJ, USA) or Invitrogen (Calsbad, CA, USA). In some cases, the specificity of the two-finger module was adjusted from GDNNMG to GDNNMA by altering the residues at positions -1, 1 and 2 in the N-terminal cap from RSD to QRG (50). The traditional-modular-assembly four- or five-finger ZFPs were created based on existing six-finger ZFPs by PCR amplification of the desired finger subsets. These ZFPs were flanked by Acc65I/BamHI sites to facilitate the cloning of these ZFPs into 1352-omega-UV2 vector for B1H-binding site selection or pCS2-DD or RR vectors for creating ZFNs for activity assay.

### B1H-binding site selections using the 28 bp library

The selections for 3F and 4F ZFAs were performed as previously described (41,58). The  $1-5 \times 10^7$  US0 selection strain cells co-transformed with the 1352-omega-UV2 ZFA expression plasmid and the 28 bp pH3U3 library plasmid were plated on NM minimal medium selective plates lacking uracil and containing 3-AT (2.5, 5 or 10 mM) as the competitor and grown at 37°C for 36–72 h. The number of surviving bacterial colonies on each plate was estimated, and then these colonies were pooled, and the population of recovered DNA sequences was determined via Illumina sequencing. Unique sequences were ranked based on the number of recovered reads. From this list, an overrepresented sequence motif was determined with MEME (59,60) using as input the number of unique sequences from the top of the list that correspond to the estimated number of colonies on the selection plate (typically >1000).

### ZFN injections and lesion analysis

For gene targeting in zebrafish, ZFAs were cloned in pCS2 vectors containing the DD/RR obligate heterodimer version of the *FokI* nuclease domain (61,62). pCS2-ZFN constructs were linearized with NotI, and mRNA was transcribed using the Message Machine SP6 kit from Ambion. ZFN mRNAs were injected into the blastomere of one-cell-stage zebrafish embryos as previously described (24). ZFN-injected embryos (8–30) with normal appearance and uninjected embryos were collected 24 h post-fertilization (h.p.f.) and incubated in 50 mM of NaOH (15  $\mu$ l/embryo) for 15 min at 95°C to isolate genomic DNA and then neutralized with 0.5 M of Tris-HCl (4  $\mu$ l/embryo). The DNA solution was centrifuged for 1 min at 13 000 r.p.m., and supernatant was taken for lesion analysis. For initial validation of ZFN activity, the region flanking the ZFN target site was amplified using the Phire Hot Start DNA Polymerase (New England Biolabs), and restriction fragment length polymorphism (RFLP) analysis or *T7 Endonuclease I*

(or T7E1, New England Biolabs) assay was performed (39,63). For T7E1 assay, The PCR products were denatured and re-annealed using the following program: 95°C for 180 s, 85°C for 20 s, 75°C for 20 s, 65°C for 20 s, 55°C for 20 s, 45°C for 20 s, 35°C for 20 s and 25°C for 20 s with a 0.1°C/s decrease rate in between steps. This allows the hybridization of mutant and wild-type DNA strands to form duplex DNA, which then can be detected by T7E1 nuclease assay. After re-annealing, 10 µl of the products were then treated with 2U of T7E1 for 1 h in the presence of NEB buffer 2. The T7E1 treated and untreated DNA was then subjected to electrophoresis on 3% agarose gel (Ultra pure 1000, Invitrogen). The gel images were analysed using ImageJ, and lesion rate was calculated as described previously (64).

### Transfection of ZFNs in HEK 293 T cells

Low passage HEK 293T cells were plated in 12-well plates overnight, such that they were 75% confluent the next day. Transfection was performed using TransIt-LT1 (Mirus Bio) according to the manufacturer's protocol with 1 µg of total DNA made up of 40% for each ZFN and 20% of a GFP transfection-efficiency reporter. Cells were passaged once to a larger plate and harvested ~3 days post-transfection by lysis in QuickExtract (EpiCentre) solution using 150 µl per million cells before PCR amplification of the ZFN cut sites.

### LacZα blue–white assay

To further determine the types of the mutations induced by the ZFNs in the zebrafish or human cell genome, we also cloned the target fragments, such that they generate a sequence lacking stop codons in-frame with LacZα gene on pBluescript-KS(-) vector between XbaI and KpnI sites. The lengths of the target fragments cloned were between 60 and 90 bp, such that it would have minimal impact on the function of the translated LacZ peptide in the α-complementation assay using XL1-Blue *Escherichia coli* cell (the amplicons were chosen, such that there are no stop codons in the reading frame). The small indels induced by the ZFN through non-homologous end joining (NHEJ) pathway would disrupt the reading frame of the LacZα on the pBluescript-KS (-) vector, so that there is no functional LacZα peptide produced and consequently no active β-galactosidase. *E. coli* colonies harbouring such fragments seem to be white on plates containing X-gal and IPTG. On the other hand, colonies containing wild-type sequence without indels would produce active LacZα peptide and appear to be blue colonies on these plates. The identities of the sequence (or the types of the lesions) are then identified through sequencing the inserts from white colonies.

## RESULTS

### Selection of two-finger modules recognizing GANNAG target sites

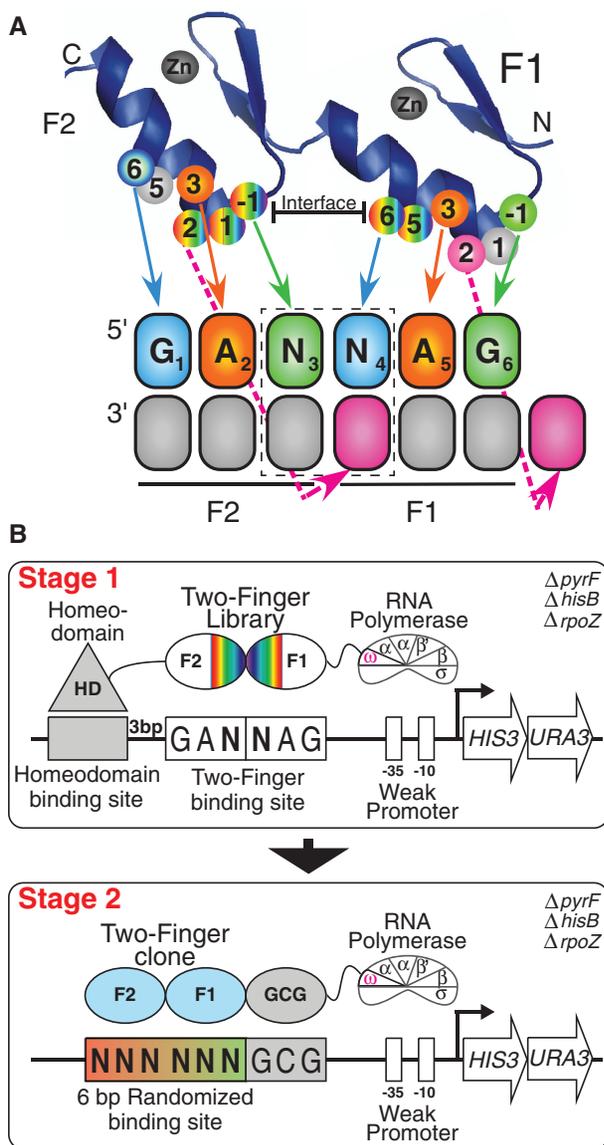
Leveraging our previous success in selecting functional two-finger modules recognizing GRNNYG sites using

the B1H system (50), we used a similar two-stage approach for the identification of modules targeting GANNAG sites (Figure 1). In the first stage, modules complementary to each of the 16 GANNAG interfaces were selected from a two-finger library that consists of  $\sim 2 \times 10^7$  variants, where positions +5 and +6 of the N-terminal finger and positions -1, +1 and +2 of the C-terminal finger were randomized. The other recognition positions within the fingers were fixed, where Asn is present at position +3 of each finger to mediate Adenine recognition. The stringency of the selection conditions for each target site was optimized to obtain a few hundred surviving colonies on each selection plate to restrict survival to the most favourable fingers for target recognition.

The resulting pools of selected modules for the sixteen gANNAG target sites trended towards a consensus sequence (Figure 2, Supplementary Table S1). Comparison of these recovered sequences with clones recovered from the corresponding gANNCg selections (50) reveals the complexity inherent in recognition at the finger–finger interface. Fundamentally, the two-finger libraries used in these selections differed at only a single position (either Thr or Asn present at position +3 of Finger 1), yet in some cases, this led to a dramatic difference in the recovered residues at the interface positions. Some of the dinucleotide junctions, in particular those for the AN-type junctions, result in the recovery of similar finger interface residues in both the ANNA and ANNC selections. However, others, in particular those for the CN-type junctions, result in the recovery of different residue sets at the finger–finger interface from each library. This disparity alludes to the presence of context-dependent effects at the finger–finger interface.

### Identification of two-finger modules preferring each dinucleotide junction

To identify two-finger modules with the most favourable recognition properties for each GANNAG dinucleotide junction, we analysed their DNA-binding specificity using the B1H system (41,57). We objectively selected a small number of clones from each recovered pool for analysis. The DNA-binding specificity of each module was characterized as a three-finger ZFA by appending a single finger recognizing a 'GCG' triplet to the N-terminus. In this context, each two-finger module was characterized using a reporter system containing a six base pair randomized binding site library flanking the finger 1 recognition sequence (50,56) (Figure 1B). Recovered binding sites from the randomized library for each two-finger module were pooled and then characterized by Illumina sequencing to determine the preferred recognition motif. Using this approach, we characterized 77 candidate two-finger modules to identify those with the strongest preference for each of the 16 GANNAG dinucleotide junctions (Supplementary Figure S1). Based on this analysis, we have identified modules that are compatible with each of the 16 dinucleotide junctions (Figure 3; Supplementary Table S2). For 14 of these modules, the desired dinucleotide junction is the most

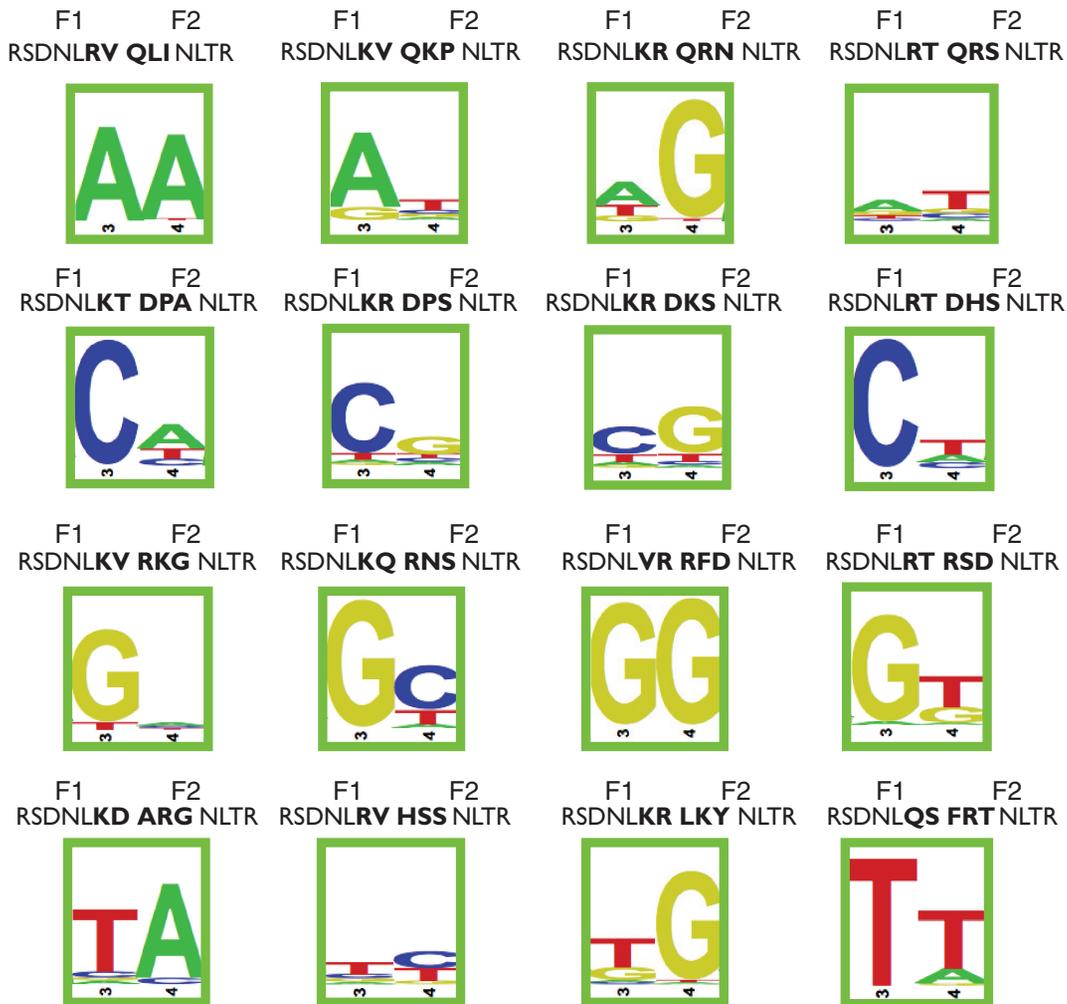


**Figure 1.** Selection of GANNAG finger sets. (A) Schematic of two-finger ZFP library used in these selections with the specificity determinants mapped to their recognition positions in their binding site. The dashed box indicates the position of the dinucleotide junction. This library contains randomized amino acids at the finger–finger interface at positions +5 and +6 of finger 1 (randomized with VNS codons) and positions –1, +1 and +2 of finger 2 (randomized with NNW codons), where the numbering scheme refers to the position of the residue relative to the start of the recognition helix. The finger 1 residues at positions –1, 1 and 2 (R, S and D) represent the N-terminal cap, and the finger 2 residues at positions 5 and 6 (T and R) represent the C-terminal cap. (B) Schematic representation of the two-stage process used to identify two-finger modules with the desired sequence preference. In Stage 1, the BIH system is used to select two-finger modules complementary to each target site. The randomized two-finger module library is fused between the DNA-binding domain of the Engrailed homeodomain and the  $\omega$ -subunit of the RNA polymerase. The fixed 6-bp GANNAG target site is present on the His3/Ura3 reporter plasmid between the homeodomain binding site and the –35 box. In Stage 2, the DNA binding specificity of candidate two-finger modules obtained from the first stage of the selection are interrogated. Each two-finger module is fused to an N-terminal finger (RSDTLAR) that binds to the ‘GCG’ triplet adjacent to the 6bp randomized zinc-finger binding region on the reporter plasmid. The recovered binding sites are determined by Illumina sequencing, and then a binding site motif is calculated from these sequences (56).



**Figure 2.** Comparison of the preferred residues recovered for the dinucleotide junctions in gANNCg (50) and gANNAG selections. In both selections, residues were selected at positions +5 and +6 of finger 1 and positions –1, +1 and +2 of finger 2. Amino acids at position +3 of the two fingers were fixed in the library. The recovered sequences are displayed as frequency logos. The only fixed difference between the two-finger libraries used in these selections occurs at position +3 of the finger 1, which is threonine (T) in the gANNCg selections and asparagine (N) in gANNAG selections.

prevalent sequence recovered in the binding site selections, although in some cases, the preference for this sequence is only modest. For the two remaining junctions (gaACag & gaCCag), we identified modules that recognize



**Figure 3.** DNA binding specificities of the two-finger GANNAG modules with the most favourable specificity for each of the 16 different dinucleotide junctions. The DNA binding specificities were determined using BIH system (Supplementary Figure S1), and the sequence logo at the 2-bp interface is shown. Amino acid residues at positions  $-1$  to  $+6$  of the recognition helix of each finger are shown. The five amino acid residues recovered in the ANNA interface selection are indicated in bold font.

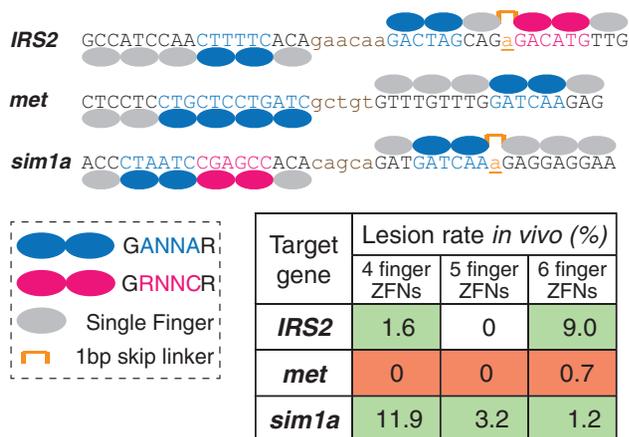
the intended target site as the second and third most preferred site, respectively. These modules constitute our validated GANNAG archive for ZFA assembly.

### Demonstrating the functionality of the GANNAG two-finger modules

The validated GANNAG two-finger modules expand our existing archives (41,50) for generating ZFAs via modular assembly. To demonstrate the functionality of these new modules, we generated ZFNs via a traditional modular assembly approach for three targets (*IRS2*, *met* and *sim1a*) in the zebrafish genome, where each ZFA incorporates at least one new GANNAG module (Figure 4). In some cases, the specificity of the two-finger module was adjusted from GANNAG to GANNAA by altering the residues (from RSD to QRG) in the N-terminal cap (50). These targets were chosen, such that six-finger ZFAs can be constructed for each half site providing the opportunity to compare the activity of four-, five- and six-finger

proteins (Supplementary Table S3). In two of these ZFAs, we used a recently described THPRAPIPKP linker between fingers from Sangamo BioSciences that allows a single base pair to be skipped between intervening modules (65).

ZFNs containing these ZFAs were tested as pairs of four-, five- or six-finger proteins for each target site in zebrafish embryos. Dose response curves were used to identify an optimal concentration of ZFN mRNA for each target. This initial survey revealed that all of the *met* ZFNs were highly toxic to embryos. Based on this analysis, each ZFN dose was calibrated to a level where  $\sim 50\%$  of the embryos developed normally at 24 h.p.f.. ZFN-injected normal embryos were subsequently analysed for lesions by T7 Endonuclease I (T7EI) analysis of PCR products spanning the target site (39,63) (Supplementary Figure S2). This revealed that the *IRS2* and *sim1a* ZFNs were active, whereas the *met1* ZFNs showed minimal activity (Figure 4). Interestingly the *IRS2* and *sim1a* ZFNs displayed opposite trends

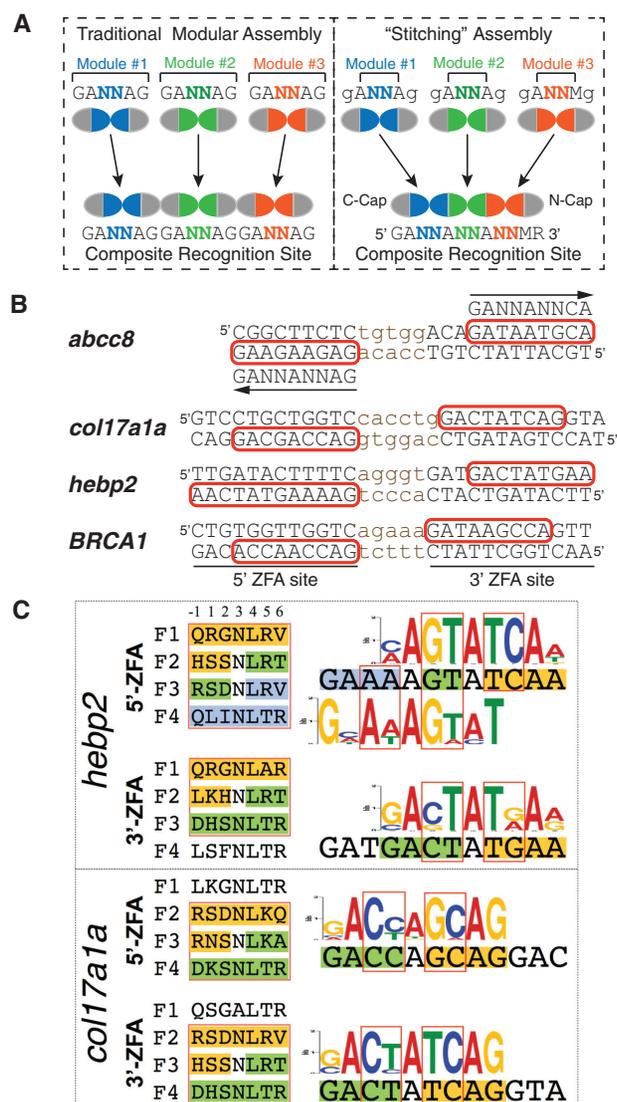


**Figure 4.** Schematic representation of the three pairs of ZFNs targeting *IRS2*, *met* and *sim1a*. Only the six-finger constructs are shown. Finger number was reduced by progressively removing fingers from the N-terminus of these constructs. The positions of two-finger modules (GANNAG) described herein or fingers from other archives [GRNNCG (50) and single fingers (41)] within these ZFAs are indicated, as is the position of the THPRAPIPKP linker that allows a single base pair (underlined lowercase base) to be skipped between intervening modules (65). The efficiency of lesion generation in zebrafish embryos by these ZFAs as a function of the number of fingers is indicated in the chart, where red highlights indicate the high toxicity of the *met* ZFNs.

with regards to activity and finger number. The *IRS2* ZFNs had the greatest activity with the longer ZFAs, whereas the *sim1a* ZFNs displayed the opposite trend.

### Stitching together ZFAs with novel DNA binding specificity

These newly characterized two-finger modules expand the archive of modules for generating ZFNs via traditional modular assembly (50). In addition these newly identified two-finger modules can serve as building blocks for a novel ZFA assembly method: finger stitching (Figure 5A). This new strategy takes advantage of a common feature of the two-finger modules targeting GANNAG interfaces. Each finger contains an asparagine (Asn) at position +3 of the recognition helix that recognizes A<sub>2</sub> and A<sub>5</sub> of the 'G<sub>1</sub>A<sub>2</sub>N<sub>3</sub>N<sub>4</sub>A<sub>5</sub>G<sub>6</sub>' 6-bp target site (Figure 1A). We envisioned that the two Asn residues might serve as bookends for the selected interface residues recognizing the dinucleotide junction N<sub>3</sub>N<sub>4</sub> and thereby preserve their DNA binding specificity on incorporation into multi-finger proteins. Thus, in this approach, ZFAs are assembled by joining interfaces from compatible two-finger modules that share identical residues at the +3 position to create large arrays. For example, a three-finger protein recognizing a sequence GAN<sub>3</sub>N<sub>4</sub>AN<sub>6</sub>N<sub>7</sub>AG would be constructed from two two-finger modules recognizing GAN<sub>3</sub>N<sub>4</sub>AG and GAN<sub>6</sub>N<sub>7</sub>AG, where the bold A indicates the position of recognition overlap. Because our previously selected GANNCG modules also contain Asn at position +3 within the C-terminal finger of the two-finger module, these can be incorporated to generate ZFAs recognizing GAN<sub>3</sub>N<sub>4</sub>AN<sub>6</sub>N<sub>7</sub>CG target sites. In addition,



**Figure 5.** 'Finger stitching' for ZFA assembly using GANNAG modules. (A) Schematic comparison of the traditional modular assembly approach (left) with our stitching approach (right). Instead of assembling whole finger units, stitching assembles segments between the +3 positions of the neighbouring recognition helices and then caps these at the N- and C- terminus (N-cap and C-cap). Two additional components extend the targetable sequences: the ability to incorporate GANNCG modules as the last unit in a stitched array, and the use of an alternate N-cap specific for a 3' adenine (50), which allows the final specified base to be either G or A depending on the choice of the N-cap. (B) Target sites for the four pairs of 'stitched' ZFNs, where the binding site for each monomer is indicated in capital letters and the recognition element of the three or four finger stitched ZFA is boxed in red on the primary recognition strand. In some cases, a ZFA contains three stitched fingers and one additional single-finger module. For the *abcc8* target site, the composite recognition site for the stitched portion of the array is indicated above or below the primary recognition sequence, where the arrow denotes the 5'–3' orientation. (C) To assess the quality of ZFAs generated through this approach, we assembled three-finger stitched ZFAs spanning portions of the target site and determined their DNA binding specificities using the BIH system. The recognition helices for these fingers are indicated to the left of the target sites, where the positions of the stitched fingers are boxed in red. The segments of the stitched fingers that arise from a common 2F-module share a common colour. Likewise, the positions of the dinucleotide junctions between fingers in the recognition motifs for these fingers are boxed in red, and the subsites recognized by the stitched finger segments are differentially coloured.

we have developed an alternate N-terminal cap that is specific for adenine (50); consequently, the 3'-most base recognized by a stitched array can be G or A depending on the N-cap that is used. Similarly, the finger stitching can be used to create ZFPs of any desired length by extending the array through the overlap of additional modules off either terminus.

To demonstrate that this novel assembly approach can create functional ZFAs and ZFNs, we chose three target genes (*abcc8*, *coll17a1a* and *hebp2*) in the zebrafish genome containing ZFN sites that could be targeted using 'stitched' ZFAs, and where these target sites contain non-NG interfaces within the stitched fingers (Figure 5B). We also designed a pair of ZFNs for a human target gene, *BRCA1*, applying the same criteria. Genes encoding these ZFAs were generated by gene synthesis, where canonical TG(E/Q)KP linkers were used to connect all fingers in these arrays. As a first step in validating this approach, we determined the DNA binding specificities of these ZFAs in the BIH system using a randomized 28 bp library (41,50). Because our 28 bp library ( $\sim 10^8$  unique members) can more effectively sample all possible recognition sequences for a three-finger than four-finger ZFA, we determined the DNA binding specificities of three-finger subsets for many of these ZFAs to provide a clearer assessment of their specificity. The determined DNA-binding specificities demonstrate in many cases that ZFAs assembled using the stitching method recognize their intended target sites with reasonable fidelity both at the dinucleotide junction sequences and the neighbouring adenines (Figure 5C and Supplementary Figure S3). However, there are instances where the dinucleotide preference is more degenerate in the stitched fingers than observed in the parent modules (e.g. *abcc8*-3p, Supplementary Figure S3), suggesting that in some cases, these assemblies are influenced by context-dependent effects.

### Stitched ZFAs yield functional ZFNs

Given the favourable specificity of the stitched ZFAs, we evaluated their functionality as ZFNs. For each of the ZFNs targeting the zebrafish genomic sites, the optimal mRNA concentration was determined via a dose response curve as previously described, and ZFN activity was assessed at the optimal dose in healthy embryos at 24 h.p.f. The activity of the *BRCA1* ZFNs was assessed by transfection of expression plasmids encoding these ZFNs into HEK 293T cells. Genomic DNA was harvested 64 h after transfection for lesion analysis. For all samples, lesion rates were determined by enzymatic digestion of PCR products spanning the target region (either T7EI or site-specific restriction enzyme) relative to untreated control samples. All four ZFN pairs induced lesions at frequencies between 1 and 11.4% (Table 1 and Supplementary Figures S4–S7), where the lesion sequences were consistent with the types of mutations expected for ZFN activity (Supplementary Figure S8).

**Table 1.** Lesion rates for 'stitched' ZFNs

Target gene	Lesion rate <i>in vivo</i> (%)
<i>abcc8</i>	1.1
<i>coll17a1a</i>	11.4
<i>hebp2</i>	2.8
<i>brca1</i> *	2.7

\*In 293T cells.

## DISCUSSION

Although ZFN technology has been successfully used in a multitude of systems for genomic modification (2,3), one of the major barriers in adoption is the need for a simple approach to generate functional ZFNs for nearly any target site. Traditional-modular-assembly of ZFAs, although becoming more facile as the quality of the finger archives improves, still suffers from either moderate success rates (39–41,51) or moderate targeting density (50). The functional assembly of these units is complicated by the influence of context-dependent effects at the finger–finger interface (17,28,31,41,45). The CoDA method described by the Zinc Finger Consortium bypasses this problem by using two separate archives of two-finger modules that share common N-terminal or C-terminal fingers, which permits the assembly of three-finger proteins through overlap at these common units (52). Although straightforward, this system is inherently limited to the creation of three-finger ZFAs, which can restrict the precision of these ZFNs in complex metazoan genomes, and assessment of its modules has focused on ZFAs recognizing NG-type junctions (Supplementary Table S4). We have sought to bypass this limitation through the development of a new assembly method wherein finger–finger interface units provide the grammar for assembly, which ensures that the finger–finger interface is always compatible. We assembled four ZFN pairs using this stitching approach focused on ZFAs that contain non-NG junctions at the finger recognition sequences. Remarkably, the specificity of the modules comprising these ZFAs on assembly was generally preserved when compared with the determined specificities of the primary two-finger modules that compose the archive. Moreover, all of these ZFNs were functional when tested in zebrafish or in human cells, demonstrating that this approach can produce ZFAs that can function in the context of a complex genome.

We believe that the success of this stitching approach stems primarily from using the +3 positions to demarcate the units of assembly. The recognition preference of residues is probably best understood for the +3 position in canonically binding fingers (54). Obviously, our stitching approach ignores potential context-dependent effects along each recognition helix that is splinted together from two different modules. For example, if one considers position +3 as a pivot point for the docking of the zinc finger within the major groove, the length and bulk of the residues at the flanking recognition positions (–1 and +6) may influence the geometry of

finger binding, which could lead to sub-optimal recognition in some instances for stitched fingers. Thus, we do not expect that this new methodology will be completely free of complications, but we anticipate that it will perform favourably when compared with the traditional modular assembly approach by minimizing the incompatibility of fingers that are joined together.

Our strategy is adapted from the 1.5-finger assembly method described by Isalan *et al.* (19); however, their strategy focused solely on the assembly of three-finger ZFAs and used selection in many cases to generate a functional three-finger protein. Our approach does not require selection and can be adapted for the creation of ZFAs of any desired length. One limitation of our approach is the requirement for a suitable archive of two-finger modules that can be used for targeting the desired DNA sequence. Currently, our archive is limited to the GANNAG and GANNCG two-finger modules, but this set can be readily expanded through the selection of additional archives of modules or by co-opting two-finger modules with good specificity from archives that have been generated for other systems (39,52). Although the available archive for this assembly is currently somewhat sparse, subsets of stitched modules can be combined through standard modular assembly with one- and two-finger units from existing archives to broaden the sequences that can be readily targeted.

To facilitate the discovery of ZFN target sites that are accessible using this approach, we have modified our existing website for the identification of ZFN target sites within a user-input sequence element (<http://pgfe.umassmed.edu/ZFPmodularchV2.html>) to include the incorporation of stitched finger sets or subsets within a ZFA. The Web interface ranks a set of target sites based on the quality of the ZFA that can be assembled and outputs the amino acid and DNA sequence for the ZFAs to facilitate their creation through gene synthesis. This new assembly method coupled with the standard modular assembly approach increases the density of ZFN target sites in the zebrafish genome to approximately one every 110 bp, where 98% of the protein coding genes have a ZFN target site (Supplementary Table S5). The number of target sites that are accessible could be greatly expanded through the creation of additional two-finger module archives, where it should be readily feasible to generate a validated set of all 256 possible GNNNG units allowing virtually any site to be targeted by varying the number of fingers and the spacer between ZFN binding sites.

Although our new archive of modules and our new assembly method increases the density of ZFN target sites, our zinc finger-based systems do not have the flexibility in targeting that has recently been demonstrated with the Transcription Activator-Like Effector Nuclease (TALEN)-based platform (63,66–69). However, ZFNs remain an important platform for targeted genomic editing that may have advantages over TALENs for certain applications, in particular therapeutics. Because each zinc finger recognizes three base pairs as opposed to one base pair for each TALE module (70–73), ZFNs are inherently more compact than TALENs. Thus, for

nuclease-based gene therapy applications using viral delivery systems (74), ZFNs constitute a more compact cargo than TALENs, and as such, they may prove to be more amenable to use in certain settings.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online: Supplementary Tables 1–5 and Supplementary Figures 1–8.

## ACKNOWLEDGEMENTS

The authors gratefully acknowledge the other members of the Wolfe laboratory for insightful comments and discussions. They thank N. Lawson and his laboratory for their insightful advice and zebrafish husbandry training. They thank M. Noyes for his advice for improving the stringency of the BIH selections. They thank J. Zhu for her assistance with the website construction.

## FUNDING

U.S. National Institutes of Health (NIH) [GM068110 to S.A.W., HL093766 to N. Lawson and S.A.W.], [HG000249 to G.D.S.]; Breast Cancer Research Foundation and Silvian Foundation (to C.K.). Funding for open access charge: NIH [GM068110].

*Conflict of interest statement.* None declared.

## REFERENCES

- Kim, Y.G., Cha, J. and Chandrasegaran, S. (1996) Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain. *Proc. Natl. Acad. Sci. USA*, **93**, 1156–1160.
- Urnov, F.D., Rebar, E.J., Holmes, M.C., Zhang, H.S. and Gregory, P.D. (2010) Genome editing with engineered zinc finger nucleases. *Nat. Rev. Genet.*, **11**, 636–646.
- Carroll, D. (2011) Genome engineering with zinc-finger nucleases. *Genetics*, **188**, 773–782.
- Ochiai, H., Sakamoto, N., Fujita, K., Nishikawa, M., Suzuki, K., Matsuura, S., Miyamoto, T., Sakuma, T., Shibata, T. and Yamamoto, T. (2012) Zinc-finger nuclease-mediated targeted insertion of reporter genes for quantitative imaging of gene expression in sea urchin embryos. *Proc. Natl. Acad. Sci. USA*, **109**, 10915–10920.
- Watanabe, T., Ochiai, H., Sakuma, T., Horch, H.W., Hamaguchi, N., Nakamura, T., Bando, T., Ohuchi, H., Yamamoto, T., Noji, S. *et al.* (2012) Non-transgenic genome modifications in a hemimetabolous insect using zinc-finger and TAL effector nucleases. *Nat. Commun.*, **3**, 1017.
- Young, J.J., Cherone, J.M., Doyon, Y., Ankoudinova, I., Faraji, F.M., Lee, A.H., Ngo, C., Guschin, D.Y., Paschon, D.E., Miller, J.C. *et al.* (2011) Efficient targeted gene disruption in the soma and germ line of the frog *Xenopus tropicalis* using engineered zinc-finger nucleases. *Proc. Natl. Acad. Sci. USA*, **108**, 7052–7057.
- Merlin, C., Beaver, L.E., Taylor, O.R., Wolfe, S.A. and Reppert, S.M. (2013) Efficient targeted mutagenesis in the monarch butterfly using zinc finger nucleases. *Genome Res.*, **23**, 159–168.
- Straimer, J., Lee, M.C., Lee, A.H., Zeitler, B., Williams, A.E., Pearl, J.R., Zhang, L., Rebar, E.J., Gregory, P.D., Llinas, M. *et al.* (2012) Site-specific genome editing in *Plasmodium falciparum* using engineered zinc-finger nucleases. *Nat. Methods*, **9**, 993–998.

9. Perez, E.E., Wang, J., Miller, J.C., Jouvenot, Y., Kim, K.A., Liu, O., Wang, N., Lee, G., Bartsevich, V.V., Lee, Y.L. *et al.* (2008) Establishment of HIV-1 resistance in CD4+ T cells by genome editing using zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 808–816.
10. Yusa, K., Rashid, S.T., Strick-Marchand, H., Varela, I., Liu, P.Q., Paschon, D.E., Miranda, E., Ordóñez, A., Hannan, N.R., Rouhani, F.J. *et al.* (2011) Targeted gene correction of alpha1-antitrypsin deficiency in induced pluripotent stem cells. *Nature*, **478**, 391–394.
11. Sebastiano, V., Maeder, M.L., Angstman, J.F., Haddad, B., Khayter, C., Yeo, D.T., Goodwin, M.J., Hawkins, J.S., Ramirez, C.L., Batista, L.F. *et al.* (2011) In situ genetic correction of the sickle cell anemia mutation in human induced pluripotent stem cells using engineered zinc finger nucleases. *Stem Cells*, **29**, 1717–1726.
12. Li, H., Haurigot, V., Doyon, Y., Li, T., Wong, S.Y., Bhagwat, A.S., Malani, N., Anguela, X.M., Sharma, R., Ivanciu, L. *et al.* (2011) In vivo genome editing restores haemostasis in a mouse model of haemophilia. *Nature*, **475**, 217–221.
13. Urnov, F.D., Miller, J.C., Lee, Y.L., Beausejour, C.M., Rock, J.M., Augustus, S., Jamieson, A.C., Porteus, M.H., Gregory, P.D. and Holmes, M.C. (2005) Highly efficient endogenous human gene correction using designed zinc-finger nucleases. *Nature*, **435**, 646–651.
14. Handel, E.M., Gellhaus, K., Khan, K., Bednarski, C., Cornu, T.I., Muller-Lerch, F., Kotin, R.M., Heilbronn, R. and Cathomen, T. (2012) Versatile and efficient genome editing in human cells by combining zinc-finger nucleases with adeno-associated viral vectors. *Hum. Gene Ther.*, **23**, 321–329.
15. Connelly, J.P., Barker, J.C., Pruett-Miller, S. and Porteus, M.H. (2010) Gene correction by homologous recombination with zinc finger nucleases in primary cells from a mouse model of a generic recessive genetic disease. *Mol. Ther.*, **18**, 1103–1110.
16. Cannon, P. and June, C. (2011) Chemokine receptor 5 knockout strategies. *Current Opin. HIV AIDS*, **6**, 74–79.
17. Greisman, H.A. and Pabo, C.O. (1997) A general strategy for selecting high-affinity zinc finger proteins for diverse DNA target sites. *Science*, **275**, 657–661.
18. Rebar, E.J. and Pabo, C.O. (1994) Zinc finger phage: affinity selection of fingers with new DNA-binding specificities. *Science*, **263**, 671–673.
19. Isalan, M., Klug, A. and Choo, Y. (2001) A rapid, generally applicable method to engineer zinc fingers illustrated by targeting the HIV-1 promoter. *Nature Biotechnol.*, **19**, 656–660.
20. Segal, D.J., Dreier, B., Beerli, R.R. and Barbas, C.F. (1999) Toward controlling gene expression at will: Selection and design of zinc finger domains recognizing each of the 5'-GNN-3' DNA target sequences. *Proc. Natl. Acad. Sci. USA*, **96**, 2758–2763.
21. Sepp, A. and Choo, Y. (2005) Cell-free Selection of Zinc Finger DNA-binding proteins using in vitro compartmentalization. *J. Mol. Biol.*, **354**, 212–219.
22. Herrmann, F., Garriga-Canut, M., Baumstark, R., Fajardo-Sanchez, E., Cotterell, J., Minoche, A., Himmelbauer, H. and Isalan, M. (2011) p53 Gene repair with zinc finger nucleases optimized by yeast 1-hybrid and validated by Solexa sequencing. *PLoS One*, **6**, e20913.
23. Jamieson, A.C., Kim, S.H. and Wells, J.A. (1994) In vitro selection of zinc fingers with altered DNA-binding specificity. *Biochemistry*, **33**, 5689–5695.
24. Meng, X.D., Noyes, M.B., Zhu, L.H.J., Lawson, N.D. and Wolfe, S.A. (2008) Targeted gene inactivation in zebrafish using engineered zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 695–701.
25. Maeder, M.L., Thibodeau-Beganny, S., Osiak, A., Wright, D.A., Anthony, R.M., Eichinger, M., Jiang, T., Foley, J.E., Winfrey, R.J., Townsend, J.A. *et al.* (2008) Rapid Open-source engineering of customized zinc-finger nucleases for highly efficient gene modification. *Mol. Cell*, **31**, 294–301.
26. Durai, S., Bosley, A., Abulencia, A.B., Chandrasegaran, S. and Ostermeier, M. (2006) A bacterial one-hybrid selection system for interrogating zinc finger-DNA interactions. *Comb. Chem. High Throughput Screen.*, **9**, 301–311.
27. Pavletich, N.P. and Pabo, C.O. (1991) Zinc Finger DNA Recognition—Crystal-Structure of a Zif268-DNA Complex at 2.1-Å. *Science*, **252**, 809–817.
28. Desjarlais, J.R. and Berg, J.M. (1993) Use of a zinc-finger consensus sequence framework and specificity rules to design specific DNA binding proteins. *Proc. Natl. Acad. Sci. USA*, **90**, 2256–2260.
29. Choo, Y. and Klug, A. (1997) Physical basis of a protein-DNA recognition code. *Curr. Opin. Struct. Biol.*, **7**, 117–125.
30. Elrod-Erickson, M., Rould, M.A., Neklodova, L. and Pabo, C.O. (1996) Zif268 protein-DNA complex refined at 1.6Å: a model system for understanding zinc finger-DNA interactions. *Structure*, **4**, 1171–1180.
31. Segal, D.J., Beerli, R.R., Blancafort, P., Dreier, B., Effertz, K., Huber, A., Koksche, B., Lund, C.V., Magnenat, L., Valente, D. *et al.* (2003) Evaluation of a modular strategy for the construction of novel polydactyl zinc finger DNA-binding proteins. *Biochemistry*, **42**, 2137–2148.
32. Carroll, D., Morton, J.J., Beumer, K.J. and Segal, D.J. (2006) Design, construction and in vitro testing of zinc finger nucleases. *Nat. Protocols*, **1**, 1329–1341.
33. Liu, Q., Xia, Z. and Case, C.C. (2002) Validated zinc finger protein designs for all 16 GNN DNA triplet targets. *J. Biol. Chem.*, **277**, 3850–3856.
34. Bae, K.H., Kwon, Y.D., Shin, H.C., Hwang, M.S., Ryu, E.H., Park, K.S., Yang, H.Y., Lee, D.K., Lee, Y., Park, J. *et al.* (2003) Human zinc fingers as building blocks in the construction of artificial transcription factors. *Nat. Biotechnol.*, **21**, 275–280.
35. Blancafort, P., Segal, D.J. and Barbas, C.F. (2004) Designing transcription factor architectures for drug discovery. *Mol. Pharmacol.*, **66**, 1361–1371.
36. Bibikova, M., Beumer, K., Trautman, J.K. and Carroll, D. (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science*, **300**, 764.
37. Bibikova, M., Carroll, D., Segal, D.J., Trautman, J.K., Smith, J., Kim, Y.-G. and Chandrasegaran, S. (2001) Stimulation of homologous recombination through targeted cleavage by chimeric nucleases. *Mol. Cell Biol.*, **21**, 289–297.
38. Porteus, M.H. and Baltimore, D. (2003) Chimeric nucleases stimulate gene targeting in human cells. *Science*, **300**, 763.
39. Kim, H.J., Lee, H.J., Kim, H., Cho, S.W. and Kim, J.S. (2009) Targeted genome editing in human cells with zinc finger nucleases constructed via modular assembly. *Genome Res.*, **19**, 1279–1288.
40. Ramirez, C.L., Foley, J.E., Wright, D.A., Muller-Lerch, F., Rahman, S.H., Cornu, T.I., Winfrey, R.J., Sander, J.D., Fu, F., Townsend, J.A. *et al.* (2008) Unexpected failure rates for modular assembly of engineered zinc fingers. *Nat. Methods*, **5**, 374–375.
41. Zhu, C., Smith, T., McNulty, J., Rayla, A.L., Lakshmanan, A., Siekmann, A.F., Buffardi, M., Meng, X., Shin, J., Padmanabhan, A. *et al.* (2011) Evaluation and application of modularly assembled zinc-finger nucleases in zebrafish. *Development*, **138**, 4555–4564.
42. Shimizu, Y., Sollu, C., Meckler, J.F., Adriaenssens, A., Zykovich, A., Cathomen, T. and Segal, D.J. (2011) Adding fingers to an engineered zinc finger nuclease can reduce activity. *Biochemistry*, **50**, 5033–5041.
43. Mandell, J.G. and Barbas, C.F. 3rd. (2006) Zinc finger tools: custom DNA-binding domains for transcription factors and nucleases. *Nucleic Acids Res.*, **34**, W516–W523.
44. Sander, J.D., Zaback, P., Joung, J.K., Voytas, D.F. and Dobbs, D. (2009) An affinity-based scoring scheme for predicting DNA-binding activities of modularly assembled zinc-finger proteins. *Nucleic Acids Res.*, **37**, 506–515.
45. Wolfe, S.A., Greisman, H.A., Ramm, E.I. and Pabo, C.O. (1999) Analysis of zinc fingers optimized via phage display: evaluating the utility of a recognition code. *J. Mol. Biol.*, **285**, 1917–1934.
46. Isalan, M., Choo, Y. and Klug, A. (1997) Synergy between adjacent zinc fingers in sequence-specific DNA recognition. *Proc. Natl. Acad. Sci. USA*, **94**, 5617–5621.
47. Dreier, B., Beerli, R.R., Segal, D.J., Flippin, J.D. and Barbas, C.F. 3rd. (2001) Development of zinc finger domains for recognition of the 5'-ANN-3' family of DNA sequences and their use in the construction of artificial transcription factors. *J. Biol. Chem.*, **276**, 29466–29478.
48. Wolfe, S.A., Grant, R.A., Elrod-Erickson, M. and Pabo, C.O. (2001) Beyond the “recognition code”: structures of two Cys2His2 zinc finger/TATA box complexes. *Structure*, **9**, 717–723.

49. Doyon, Y., McCammon, J.M., Miller, J.C., Faraji, F., Ngo, C., Katibah, G.E., Amora, R., Hocking, T.D., Zhang, L., Rebar, E.J. *et al.* (2008) Heritable targeted gene disruption in zebrafish using designed zinc-finger nucleases. *Nat. Biotechnol.*, **26**, 702–708.
50. Gupta, A., Christensen, R.G., Rayla, A.L., Lakshmanan, A., Stormo, G.D. and Wolfe, S.A. (2012) An optimized two-finger archive for ZFN-mediated gene targeting. *Nat. Methods*, **9**, 588–590.
51. Kim, S., Lee, M.J., Kim, H., Kang, M. and Kim, J.S. (2011) Preassembled zinc-finger arrays for rapid construction of ZFNs. *Nat. Methods*, **8**, 7.
52. Sander, J.D., Dahlborg, E.J., Goodwin, M.J., Cade, L., Zhang, F., Cifuentes, D., Curtin, S.J., Blackburn, J.S., Thibodeau-Beganny, S., Qi, Y. *et al.* (2011) Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA). *Nat. Methods*, **8**, 67–69.
53. Cheng, C. and Young, E.T. (1995) A single amino acid substitution in zinc finger 2 of Adr1p changes its binding specificity at two positions in UAS1. *J. Mol. Biol.*, **251**, 1–8.
54. Miller, J.C. and Pabo, C.O. (2001) Rearrangement of side-chains in a zif268 mutant highlights the complexities of zinc finger-DNA recognition. *J. Mol. Biol.*, **313**, 309–315.
55. Westerfield, M. (1993) *The Zebrafish Book*. University of Oregon Press, Eugene, OR.
56. Christensen, R.G., Gupta, A., Zuo, Z., Schrieffer, L.A., Wolfe, S.A. and Stormo, G.D. (2011) A modified bacterial one-hybrid system yields improved quantitative models of transcription factor specificity. *Nucleic Acids Res.*, **39**, e83.
57. Noyes, M.B., Meng, X., Wakabayashi, A., Sinha, S., Brodsky, M.H. and Wolfe, S.A. (2008) A systematic characterization of factors that regulate *Drosophila* segmentation via a bacterial one-hybrid system. *Nucleic Acids Res.*, **36**, 2547–2560.
58. Gupta, A., Meng, X., Zhu, L.J., Lawson, N.D. and Wolfe, S.A. (2011) Zinc finger protein-dependent and -independent contributions to the in vivo off-target activity of zinc finger nucleases. *Nucleic Acids Res.*, **39**, 381–392.
59. Bailey, T.L., Boden, M., Buske, F.A., Frith, M., Grant, C.E., Clementi, L., Ren, J., Li, W.W. and Noble, W.S. (2009) MEME SUITE: tools for motif discovery and searching. *Nucleic Acids Res.*, **37**, W202–W208.
60. Bailey, T.L., Williams, N., Misleh, C. and Li, W.W. (2006) MEME: discovering and analyzing DNA and protein sequence motifs. *Nucleic Acids Res.*, **34**, W369–W373.
61. Miller, J.C. (2007) An improved zinc-finger nuclease architecture for highly specific genome editing. *Nature Biotechnol.*, **25**, 778–785.
62. Szczepek, M., Brondani, V., Buchel, J., Serrano, L., Segal, D.J. and Cathomen, T. (2007) Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases. *Nature Biotechnol.*, **25**, 786–793.
63. Reyon, D., Tsai, S.Q., Khayter, C., Foden, J.A., Sander, J.D. and Joung, J.K. (2012) FLASH assembly of TALENs for high-throughput genome editing. *Nat. Biotechnol.*, **30**, 460–465.
64. Guschin, D.Y., Waite, A.J., Katibah, G.E., Miller, J.C., Holmes, M.C. and Rebar, E.J. (2010) *Engineered Zinc Finger Proteins*, **649**, pp. 247–256.
65. Soldner, F., Laganier, J., Cheng, A.W., Hockemeyer, D., Gao, Q., Alagappan, R., Khurana, V., Golbe, L.I., Myers, R.H., Lindquist, S. *et al.* (2011) Generation of isogenic pluripotent stem cells differing exclusively at two early onset Parkinson point mutations. *Cell*, **146**, 318–331.
66. Christian, M., Cermak, T., Doyle, E.L., Schmidt, C., Zhang, F., Hummel, A., Bogdanove, A.J. and Voytas, D.F. (2010) Targeting DNA double-strand breaks with TAL effector nucleases. *Genetics*, **186**, 757–761.
67. Miller, J.C., Tan, S., Qiao, G., Barlow, K.A., Wang, J., Xia, D.F., Meng, X., Paschon, D.E., Leung, E., Hinkley, S.J. *et al.* (2011) A TALE nuclease architecture for efficient genome editing. *Nat. Biotechnol.*, **29**, 143–148.
68. Bedell, V.M., Wang, Y., Campbell, J.M., Poshusta, T.L., Starker, C.G., Krug II, R.G., Tan, W., Penheiter, S.G., Ma, A.C. *et al.* (2012) In vivo genome editing using a high-efficiency TALEN system. *Nature*, **491**, 114–118.
69. Mussolino, C., Morbitzer, R., Lutge, F., Dannemann, N., Lahaye, T. and Cathomen, T. (2011) A novel TALE nuclease scaffold enables high genome editing activity in combination with low toxicity. *Nucleic Acids Res.*, **39**, 9283–9293.
70. Boch, J., Scholze, H., Schornack, S., Landgraf, A., Hahn, S., Kay, S., Lahaye, T., Nickstadt, A. and Bonas, U. (2009) Breaking the code of DNA binding specificity of TAL-type III effectors. *Science*, **326**, 1509–1512.
71. Moscou, M.J. and Bogdanove, A.J. (2009) A simple cipher governs DNA recognition by TAL effectors. *Science*, **326**, 1501.
72. Mak, A.N., Bradley, P., Cernadas, R.A., Bogdanove, A.J. and Stoddard, B.L. (2012) The crystal structure of TAL effector PthXo1 bound to its DNA target. *Science*, **335**, 716–719.
73. Deng, D., Yan, C., Pan, X., Mahfouz, M., Wang, J., Zhu, J.K., Shi, Y. and Yan, N. (2012) Structural basis for sequence-specific recognition of DNA by TAL effectors. *Science*, **335**, 720–723.
74. Atkinson, H. and Chalmers, R. (2010) Delivering the goods: viral and non-viral gene therapy systems and the inherent limits on cargo DNA and internal sequences. *Genetica*, **138**, 485–498.

# Chapter VII:

# References

- Aguirre-Ghiso, J.A. (2007). Models, mechanisms and clinical evidence for cancer dormancy. *Nat. Rev. Cancer* 7, 834–846.
- Alhazzazi, T.Y., Kamarajan, P., Verdin, E., and Kapila, Y.L. (2013). Sirtuin-3 (SIRT3) and the Hallmarks of Cancer. *Genes Cancer* 4, 164–171.
- Baccelli, I., Schneeweiss, A., Riethdorf, S., Stenzinger, A., Schillert, A., Vogel, V., Klein, C., Saini, M., Bäuerle, T., Wallwiener, M., et al. (2013). Identification of a population of blood circulating tumor cells from breast cancer patients that initiates metastasis in a xenograft assay. *Nat. Biotechnol.* 31, 539–544.
- Bannister, A.J., Zegerman, P., Partridge, J.F., Miska, E.A., Thomas, J.O., Allshire, R.C., and Kouzarides, T. (2001). Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* 410, 120–124.
- Beaulieu, J.-M., and Gainetdinov, R.R. (2011). The Physiology, Signaling, and Pharmacology of Dopamine Receptors. *Pharmacol. Rev.* 63, 182–217.
- Berg, J.M., Tymoczko, J.L., Stryer, L., Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). *Biochemistry* (W H Freeman).
- Bigarella, C.L., Liang, R., and Ghaffari, S. (2014). Stem cells and the impact of ROS signaling. *Development* 141, 4206–4218.
- Bilodeau, S., Kagey, M.H., Frampton, G.M., Rahl, P.B., and Young, R.A. (2009). SetDB1 contributes to repression of genes encoding developmental regulators and maintenance of ES cell state. *Genes Dev.* 23, 2484–2489.
- Bird, A. (2002). DNA methylation patterns and epigenetic memory. *Genes Dev.* 16, 6–21.
- Bird, A., Taggart, M., Frommer, M., Miller, O.J., and Macleod, D. (1985). A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40, 91–99.
- Bos, P.D., Zhang, X.H.-F., Nadal, C., Shu, W., Gomis, R.R., Nguyen, D.X., Minn, A.J., van de Vijver, M.J., Gerald, W.L., Foekens, J.A., et al. (2009). Genes that mediate breast cancer metastasis to the brain. *Nature* 459, 1005–1009.
- Bozic, I., Antal, T., Ohtsuki, H., Carter, H., Kim, D., Chen, S., Karchin, R., Kinzler, K.W., Vogelstein, B., and Nowak, M.A. (2010). Accumulation of driver and passenger mutations during tumor progression. *Proc. Natl. Acad. Sci.* 107, 18545–18550.
- Braun, S., Pantel, K., Müller, P., Janni, W., Hepp, F., Kentenich, C.R.M., Gastroph, S., Wischnik, A., Dimpfl, T., Kindermann, G., et al. (2000). Cytokeratin-Positive Cells in the Bone Marrow and Survival of Patients with Stage I, II, or III Breast Cancer. *N. Engl. J. Med.* 342, 525–533.

- Breslow, M.J. (1992). Regulation of adrenal medullary and cortical blood flow. *Am. J. Physiol. - Heart Circ. Physiol.* 262, H1317–H1330.
- Cailleau, R., Olivé, M., and Cruciger, Q.V.J. (1978). Long-Term Human Breast Carcinoma Cell Lines of Metastatic Origin: Preliminary Characterization. *In vitro* 14, 911–915.
- Campbell, P.J., Yachida, S., Mudie, L.J., Stephens, P.J., Pleasance, E.D., Stebbings, L.A., Morsberger, L.A., Latimer, C., McLaren, S., Lin, M.-L., et al. (2010). The patterns and dynamics of genomic instability in metastatic pancreatic cancer. *Nature* 467, 1109–1113.
- Ceol, C.J., Houvras, Y., Jane-Valbuena, J., Bilodeau, S., Orlando, D.A., Battisti, V., Fritsch, L., Lin, W.M., Hollmann, T.J., Ferré, F., et al. (2011). The histone methyltransferase SETDB1 is recurrently amplified in melanoma and accelerates its onset. *Nature* 471, 513–517.
- Chaffer, C.L., and Weinberg, R.A. (2011). A Perspective on Cancer Cell Metastasis. *Science* 331, 1559–1564.
- Chaffer, C.L., Marjanovic, N.D., Lee, T., Bell, G., Kleer, C.G., Reinhardt, F., D’Alessio, A.C., Young, R.A., and Weinberg, R.A. (2013). Poised chromatin at the ZEB1 promoter enables breast cancer cell plasticity and enhances tumorigenicity. *Cell* 154, 61–74.
- Chang, H.-W., Huang, C.-Y., Yang, S.-Y., Wu, V.-C., Chu, T.-S., Chen, Y.-M., Hsieh, B.-S., and Wu, K.-D. (2014). Role of D2 dopamine receptor in adrenal cortical cell proliferation and aldosterone-producing adenoma tumorigenesis. *J. Mol. Endocrinol.* 52, 87–96.
- Chen, Q., Zhang, X.H.-F., and Massagué, J. (2011). Macrophage Binding to Receptor VCAM-1 Transmits Survival Signals in Breast Cancer Cells that Invade the Lungs. *Cancer Cell* 20, 538–549.
- Chen, Y., Sprung, R., Tang, Y., Ball, H., Sangras, B., Kim, S.C., Falck, J.R., Peng, J., Gu, W., and Zhao, Y. (2007). Lysine propionylation and butyrylation are novel post-translational modifications in histones. *Mol. Cell. Proteomics MCP* 6, 812–819.
- Collins, S. (2008). *The Hunger Games* (New York: Scholastic Press).
- Dang, L., White, D.W., Gross, S., Bennett, B.D., Bittinger, M.A., Driggers, E.M., Fantin, V.R., Jang, H.G., Jin, S., Keenan, M.C., et al. (2009). Cancer-associated IDH1 mutations produce 2-hydroxyglutarate. *Nature* 462, 739–744.
- Davies, H., Bignell, G.R., Cox, C., Stephens, P., Edkins, S., Clegg, S., Teague, J., Woffendin, H., Garnett, M.J., Bottomley, W., et al. (2002). Mutations of the BRAF gene in human cancer. *Nature* 417, 949–954.
- Demicheli, R., Retsky, M.W., Hrushesky, W.J., and Baum, M. (2007). Tumor dormancy and surgery-driven interruption of dormancy in breast cancer: learning from failures. *Nat. Rev. Clin. Oncol.* 4, 699–710.

- Ding, L., Ellis, M.J., Li, S., Larson, D.E., Chen, K., Wallis, J.W., Harris, C.C., McLellan, M.D., Fulton, R.S., Fulton, L.L., et al. (2010). Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* *464*, 999–1005.
- diSibio, G., and French, S.W. (2009). *Metastatic Patterns of Cancers: Results From a Large Autopsy Study*.
- Feinberg, A.P., and Vogelstein, B. (1983). Hypomethylation distinguishes genes of some human cancers from their normal counterparts. *Nature* *301*, 89–92.
- Fidler, I.J., Gersten, D.M., and Hart, I.R. (1978). The biology of cancer invasion and metastasis. *Adv. Cancer Res.* *28*, 149–250.
- Figuroa, M.E., Abdel-Wahab, O., Lu, C., Ward, P.S., Patel, J., Shih, A., Li, Y., Bhagwat, N., Vasanthakumar, A., Fernandez, H.F., et al. (2010). Leukemic IDH1 and IDH2 Mutations Result in a Hypermethylation Phenotype, Disrupt TET2 Function, and Impair Hematopoietic Differentiation. *Cancer Cell* *18*, 553–567.
- Finley, L.W.S., Carracedo, A., Lee, J., Souza, A., Egia, A., Zhang, J., Teruya-Feldstein, J., Moreira, P.I., Cardoso, S.M., Clish, C.B., et al. (2011). SIRT3 opposes reprogramming of cancer cell metabolism through HIF1 $\alpha$  destabilization. *Cancer Cell* *19*, 416–428.
- Garraway, L.A., and Lander, E.S. (2013). Lessons from the Cancer Genome. *Cell* *153*, 17–37.
- Garraway, L.A., Widlund, H.R., Rubin, M.A., Getz, G., Berger, A.J., Ramaswamy, S., Beroukhi, R., Milner, D.A., Granter, S.R., Du, J., et al. (2005). Integrative genomic analyses identify MTF1 as a lineage survival oncogene amplified in malignant melanoma. *Nature* *436*, 117–122.
- Gilbert, S.F. (2014). *Developmental biology* (Sunderland, MA, USA: Sinauer Associates, Inc., Publishers).
- Greger, V., Passarge, E., Höpping, W., Messmer, E., and Horsthemke, B. (1989). Epigenetic changes may contribute to the formation and spontaneous regression of retinoblastoma. *Hum. Genet.* *83*, 155–158.
- Gupta, G.P., Nguyen, D.X., Chiang, A.C., Bos, P.D., Kim, J.Y., Nadal, C., Gomis, R.R., Manova-Todorova, K., and Massagué, J. (2007). Mediators of vascular remodelling co-opted for sequential steps in lung metastasis. *Nature* *446*, 765–770.
- Hammond, C. (2015). *Cellular and molecular neurophysiology* (Boston, MA: Elsevier).
- Hawkes, C.A., Jayakody, N., Johnston, D.A., Bechmann, I., and Carare, R.O. (2014). Failure of Perivascular Drainage of  $\beta$ -amyloid in Cerebral Amyloid Angiopathy: Failure of Perivascular Drainage of  $\beta$ -amyloid. *Brain Pathol.* *24*, 396–403.

- Hevir, N., Trošt, N., Debeljak, N., and Lanišnik Rižner, T. (2011). Expression of estrogen and progesterone receptors and estrogen metabolizing enzymes in different breast cancer cell lines. *Chem. Biol. Interact.* *191*, 206–216.
- Hotchkiss, R.D. (1948). The quantitative separation of purines, pyrimidines, and nucleosides by paper chromatography. *J. Biol. Chem.* *175*, 315–332.
- Jacob, L.S., Vanharanta, S., Obenauf, A.C., Pirun, M., Viale, A., Socci, N.D., and Massague, J. (2015). Metastatic competence can emerge with selection of pre-existing oncogenic alleles without a need of new mutations. *Cancer Res.*
- Jenuwein, T. (2001). Translating the Histone Code. *Science* *293*, 1074–1080.
- Kang, Y., and Pantel, K. (2013). Tumor Cell Dissemination: Emerging Biological Insights from Animal Models and Cancer Patients. *Cancer Cell* *23*, 573–581.
- Kang, Y., Siegel, P.M., Shu, W., Drobnjak, M., Kakonen, S.M., Cordon-Cardo, C., Guise, T.A., and Massagué, J. (2003). A multigenic program mediating breast cancer metastasis to bone. *Cancer Cell* *3*, 537–549.
- Karmodiya, K., Krebs, A.R., Oulad-Abdelghani, M., Kimura, H., and Tora, L. (2012). H3K9 and H3K14 acetylation co-occur at many gene regulatory elements, while H3K14ac marks a subset of inactive inducible promoters in mouse embryonic stem cells. *BMC Genomics* *13*, 424.
- Keller, P.J., Lin, A.F., Arendt, L.M., Klebba, I., Jones, A.D., Rudnick, J.A., DiMeo, T.A., Gilmore, H., Jefferson, D.M., Graham, R.A., et al. (2010). Mapping the cellular and molecular heterogeneity of normal and malignant breast tissues and cultured cell lines. *Breast Cancer Res.* *12*, R87.
- Kennecke, H., Yerushalmi, R., Woods, R., Cheang, M.C.U., Voduc, D., Speers, C.H., Nielsen, T.O., and Gelmon, K. (2010). Metastatic Behavior of Breast Cancer Subtypes. *J. Clin. Oncol.* *28*, 3271–3277.
- Kim, H.-S., Patel, K., Muldoon-Jacobs, K., Bisht, K.S., Aykin-Burns, N., Pennington, J.D., van der Meer, R., Nguyen, P., Savage, J., Owens, K.M., et al. (2010). SIRT3 is a Mitochondrial Localized Tumor Suppressor Required for Maintenance of Mitochondrial Integrity and Metabolism During Stress. *Cancer Cell* *17*, 41–52.
- Kung, A.W.C., Pun, K.K., Lam, K., Wang, C., and Leung, C.Y. (1990). Addisonian crisis as presenting feature in malignancies. *Cancer* *65*, 177–179.
- Lam, K.-Y., and Lo, C.-Y. (2002). Metastatic tumours of the adrenal glands: a 30-year experience in a teaching hospital. *Clin. Endocrinol. (Oxf.)* *56*, 95–101.
- Langmead, B., Trapnell, C., Pop, M., and Salzberg, S.L. (2009). Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol.* *10*, R25.

- 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*, 2113–2126.
- Liu, L., Kimball, S., Liu, H., Holowatyj, A., and Yang, Z.-Q. (2015). Genetic alterations of histone lysine methyltransferases and their significance in breast cancer. *Oncotarget* *6*, 2466–2482.
- Louveau, A., Smirnov, I., Keyes, T.J., Eccles, J.D., Rouhani, S.J., Peske, J.D., Derecki, N.C., Castle, D., Mandell, J.W., Lee, K.S., et al. (2015). Structural and functional features of central nervous system lymphatic vessels. *Nature* *523*, 337–341.
- Lu, C., Ward, P.S., Kapoor, G.S., Rohle, D., Turcan, S., Abdel-Wahab, O., Edwards, C.R., Khanin, R., Figueroa, M.E., Melnick, A., et al. (2012). IDH mutation impairs histone demethylation and results in a block to cell differentiation. *Nature* *483*, 474–478.
- Luzzi, K.J., MacDonald, I.C., Schmidt, E.E., Kerkvliet, N., Morris, V.L., Chambers, A.F., and Groom, A.C. (1998). Multistep Nature of Metastatic Inefficiency: Dormancy of Solitary Cells after Successful Extravasation and Limited Survival of Early Micrometastases. *Am. J. Pathol.* *153*, 865–873.
- Macchiarelli, G., Jiang, J., Nottola, S.A., and Sato, E. (2006). Morphological patterns of angiogenesis in ovarian follicle capillary networks. A scanning electron microscopy study of corrosion cast. *Microsc. Res. Tech.* *69*, 459–468.
- Marmorstein, R., and Zhou, M.-M. (2014). Writers and readers of histone acetylation: structure, mechanism, and inhibition. *Cold Spring Harb. Perspect. Biol.* *6*, a018762.
- McKay, L.I., and Cidlowski, J.A. (2003). Corticosteroids in the Treatment of Neoplasms.
- Miller, A.J., and Mihm, M.C. (2006). Melanoma. *N. Engl. J. Med.* *355*, 51–65.
- Miller, G., and Ogilvie, G. (1985). *Mad Max Beyond Thunderdome*.
- 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 Massagué, J. (2005). Genes that mediate breast cancer metastasis to lung. *Nature* *436*, 518–524.
- Mullier, A., Bouret, S.G., Prevot, V., and Dehouck, B. (2010). Differential distribution of tight junction proteins suggests a role for tanycytes in blood-hypothalamus barrier regulation in the adult mouse brain. *J. Comp. Neurol.* *518*, 943–962.
- Neman, J., Termini, J., Wilczynski, S., Vaidehi, N., Choy, C., Kowolik, C.M., Li, H., Hambrecht, A.C., Roberts, E., and Jandial, R. (2014). Human breast cancer metastases to the brain display GABAergic properties in the neural niche. *Proc. Natl. Acad. Sci.* *111*, 984–989.

- Network, T.C.G.A. (2012). Comprehensive molecular portraits of human breast tumours. *Nature* *490*, 61–70.
- Neve, K.A., Ford, C.P., Buck, D.C., Grandy, D.K., Neve, R.L., and Phillips, T.J. (2013). Normalizing dopamine D2 receptor-mediated responses in D2 null mutant mice by virus-mediated receptor restoration: Comparing D2L and D2S. *Neuroscience* *248*, 479–487.
- Nonaka, H., Akima, M., Nagayama, T., Hatori, T., Zhang, Z., and Ihara, F. (2003). Microvasculature of the human cerebral meninges. *Neuropathology* *23*, 129–135.
- O’Keefe, G.C., Tyers, P., Aarsland, D., Dalley, J.W., Barker, R.A., and Caldwell, M.A. (2009). Dopamine-induced proliferation of adult neural precursor cells in the mammalian subventricular zone is mediated through EGF. *Proc. Natl. Acad. Sci.* *106*, 8754–8759.
- Oskarsson, T., Acharyya, S., Zhang, X.H.-F., Vanharanta, S., Tavazoie, S.F., Morris, P.G., Downey, R.J., Manova-Todorova, K., Brogi, E., and Massagué, J. (2011). Breast cancer cells produce tenascin C as a metastatic niche component to colonize the lungs. *Nat. Med.* *17*, 867–874.
- Paget, S. (1889). THE DISTRIBUTION OF SECONDARY GROWTHS IN CANCER OF THE BREAST. *The Lancet* *133*, 571–573.
- Parsons, D.W., Jones, S., Zhang, X., Lin, J.C.-H., Leary, R.J., Angenendt, P., Mankoo, P., Carter, H., Siu, I.-M., Gallia, G.L., et al. (2008). An Integrated Genomic Analysis of Human Glioblastoma Multiforme. *Science* *321*, 1807–1812.
- Perou, C.M., Sørlie, 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*, 747–752.
- Podsypanina, K., Du, Y.-C.N., Jechlinger, M., Beverly, L.J., Hambarzumyan, D., and Varmus, H. (2008). Seeding and propagation of untransformed mouse mammary cells in the lung. *Science* *321*, 1841–1844.
- Pothos, E.N., Davila, V., and Sulzer, D. (1998). Presynaptic Recording of Quanta from Midbrain Dopamine Neurons and Modulation of the Quantal Size. *J. Neurosci.* *18*, 4106–4118.
- Rangarajan, A., and Weinberg, R.A. (2003). Opinion: Comparative biology of mouse versus human cells: modelling human cancer in mice. *Nat. Rev. Cancer* *3*, 952–959.
- Reiner, A., Reiner, G., Spona, J., Schemper, M., and Holzner, J.H. (1988). Histopathologic characterization of human breast cancer in correlation with estrogen receptor status. A comparison of immunocytochemical and biochemical analysis. *Cancer* *61*, 1149–1154.
- Rodriguez-Paredes, M., Martinez de Paz, A., Simó-Riudalbas, L., Sayols, S., Moutinho, C., Moran, S., Villanueva, A., Vázquez-Cedeira, M., Lazo, P.A., Carneiro, F., et al. (2014).

Gene amplification of the histone methyltransferase SETDB1 contributes to human lung tumorigenesis. *Oncogene* 33, 2807–2813.

Van Roosmalen, W., Le Dévédec, S.E., Golani, O., Smid, M., Pulyakhina, I., Timmermans, A.M., Look, M.P., Zi, D., Pont, C., de Graauw, M., et al. (2015). Tumor cell migration screen identifies SRPK1 as breast cancer metastasis determinant. *J. Clin. Invest.* 125, 1648–1664.

Sampson, J.H., Carter, J.H., Friedman, A.H., and Seigler, H.F. (1998). Demographics, prognosis, and therapy in 702 patients with brain metastases from malignant melanoma. *J. Neurosurg.* 88, 11–20.

Sänger, N., Effenberger, K.E., Riethdorf, S., Van Haasteren, V., Gauwerky, J., Wiegratz, I., Strebhardt, K., Kaufmann, M., and Pantel, K. (2011). Disseminated tumor cells in the bone marrow of patients with ductal carcinoma in situ. *Int. J. Cancer* 129, 2522–2526.

Santos-Rosa, H., Schneider, R., Bannister, A.J., Sherriff, J., Bernstein, B.E., Emre, N.C.T., Schreiber, S.L., Mellor, J., and Kouzarides, T. (2002). Active genes are tri-methylated at K4 of histone H3. *Nature* 419, 407–411.

Saxonov, S., Berg, P., and Brutlag, D.L. (2006). A genome-wide analysis of CpG dinucleotides in the human genome distinguishes two distinct classes of promoters. *Proc. Natl. Acad. Sci.* 103, 1412–1417.

Schuettengruber, B., Chourrout, D., Vervoort, M., Leblanc, B., and Cavalli, G. (2007). Genome Regulation by Polycomb and Trithorax Proteins. *Cell* 128, 735–745.

Seidenwurm, D.J., Elmer, E.B., Kaplan, L.M., Williams, E.K., Morris, D.G., and Hoffman, A.R. (1984). Metastases to the adrenal glands and the development of Addison's disease. *Cancer* 54, 552–557.

Sethi, N., Dai, X., Winter, C.G., and Kang, Y. (2011). Tumor-Derived Jagged1 Promotes Osteolytic Bone Metastasis of Breast Cancer by Engaging Notch Signaling in Bone Cells. *Cancer Cell* 19, 192–205.

Siegel, R.L., Miller, K.D., and Jemal, A. (2015). Cancer statistics, 2015. *CA. Cancer J. Clin.* 65, 5–29.

Smid, M., Wang, Y., Zhang, Y., Sieuwerts, A.M., Yu, J., Klijn, J.G.M., Foekens, J.A., and Martens, J.W.M. (2008). Subtypes of Breast Cancer Show Preferential Site of Relapse. *Cancer Res.* 68, 3108–3114.

Soni, A., Ren, Z., Hameed, O., Chanda, D., Morgan, C.J., Siegal, G.P., and Wei, S. (2015). Breast Cancer Subtypes Predispose the Site of Distant Metastases. *Am. J. Clin. Pathol.* 143, 471–478.

- Strahl, B.D., Ohba, R., Cook, R.G., and Allis, C.D. (1999). Methylation of histone H3 at lysine 4 is highly conserved and correlates with transcriptionally active nuclei in *Tetrahymena*. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 14967–14972.
- Su, B., Gao, L., Baranowski, C., Gillard, B., Wang, J., Ransom, R., Ko, H.-K., and Gelman, I.H. (2014). A Genome-Wide RNAi Screen Identifies FOXO4 as a Metastasis-Suppressor through Counteracting PI3K/AKT Signal Pathway in Prostate Cancer. *PLoS ONE* *9*.
- Suzuki, H., Yamamoto, E., Maruyama, R., Niinuma, T., and Kai, M. (2014). Biological significance of the CpG island methylator phenotype. *Biochem. Biophys. Res. Commun.* *455*, 35–42.
- Tan, M., Luo, H., Lee, S., Jin, F., Yang, J.S., Montellier, E., Buchou, T., Cheng, Z., Rousseaux, S., Rajagopal, N., et al. (2011). Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* *146*, 1016–1028.
- Teicher, B.A., and Fricker, S.P. (2010). CXCL12 (SDF-1)/CXCR4 Pathway in Cancer. *Clin. Cancer Res.* *16*, 2927–2931.
- Thiery, J.P., Acloque, H., Huang, R.Y.J., and Nieto, M.A. (2009). Epithelial-Mesenchymal Transitions in Development and Disease. *Cell* *139*, 871–890.
- Tobin, N.P., Harrell, J.C., Lötvot, J., Egyhazi Brage, S., Frostvik Stolt, M., Carlsson, L., Einbeigi, Z., Linderholm, B., Loman, N., Malmberg, M., et al., TEX Trialists Group (2015). Molecular subtype and tumor characteristics of breast cancer metastases as assessed by gene expression significantly influence patient post-relapse survival. *Ann. Oncol. Off. J. Eur. Soc. Med. Oncol. ESMO* *26*, 81–88.
- Torre, L.A., Bray, F., Siegel, R.L., Ferlay, J., Lortet-Tieulent, J., and Jemal, A. (2015). Global cancer statistics, 2012. *CA. Cancer J. Clin.* *65*, 87–108.
- Toyota, M., Ahuja, N., Ohe-Toyota, M., Herman, J.G., Baylin, S.B., and Issa, J.P. (1999). CpG island methylator phenotype in colorectal cancer. *Proc. Natl. Acad. Sci. U. S. A.* *96*, 8681–8686.
- Toyota, M., Ohe-Toyota, M., Ahuja, N., and Issa, J.P. (2000). Distinct genetic profiles in colorectal tumors with or without the CpG island methylator phenotype. *Proc. Natl. Acad. Sci. U. S. A.* *97*, 710–715.
- Turcan, S., Rohle, D., Goenka, A., Walsh, L.A., Fang, F., Yilmaz, E., Campos, C., Fabius, A.W.M., Lu, C., Ward, P.S., et al. (2012). IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype. *Nature* *483*, 479–483.
- Valster, A., Tran, N.L., Nakada, M., Berens, M.E., Chan, A.Y., and Symons, M. (2005). Cell migration and invasion assays. *Methods* *37*, 208–215.
- Vermeulen, M., and Timmers, H.T.M. (2010). Grasping trimethylation of histone H3 at lysine 4. *Epigenomics* *2*, 395–406.

- Vinson, G.P., Pudney, J.A., and Whitehouse, B.J. (1985). The mammalian adrenal circulation and the relationship between adrenal blood flow and steroidogenesis. *J. Endocrinol.* *105*, 285–294.
- Visvader, J.E. (2009). Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. *Genes Dev.* *23*, 2563–2577.
- Weihe, E., Schäfer, M.K., Erickson, J.D., and Eiden, L.E. (1994). Localization of vesicular monoamine transporter isoforms (VMAT1 and VMAT2) to endocrine cells and neurons in rat. *J. Mol. Neurosci.* *MN 5*, 149–164.
- Weisenberger, D.J. (2014). Characterizing DNA methylation alterations from The Cancer Genome Atlas. *J. Clin. Invest.* *124*, 17–23.
- Weiss, G.J., and Korn, R.L. (2012). Metastatic basal cell carcinoma in the era of hedgehog signaling pathway inhibitors. *Cancer* *118*, 5310–5319.
- Whittle, J.R., Lewis, M.T., Lindeman, G.J., and Visvader, J.E. (2015). Patient-derived xenograft models of breast cancer and their predictive power. *Breast Cancer Res.* *17*, 17.
- Xie, Z., Dai, J., Dai, L., Tan, M., Cheng, Z., Wu, Y., Boeke, J.D., and Zhao, Y. (2012). Lysine succinylation and lysine malonylation in histones. *Mol. Cell. Proteomics MCP* *11*, 100–107.
- Xu, W., Yang, H., Liu, Y., Yang, Y., Wang, P., Kim, S.-H., Ito, S., Yang, C., Wang, P., Xiao, M.-T., et al. (2011). Oncometabolite 2-Hydroxyglutarate Is a Competitive Inhibitor of  $\alpha$ -Ketoglutarate-Dependent Dioxygenases. *Cancer Cell* *19*, 17–30.
- Yachida, S., Jones, S., Bozic, I., Antal, T., Leary, R., Fu, B., Kamiyama, M., Hruban, R.H., Eshleman, J.R., Nowak, M.A., et al. (2010). Distant metastasis occurs late during the genetic evolution of pancreatic cancer. *Nature* *467*, 1114–1117.
- Yates, L.R., and Campbell, P.J. (2012). Evolution of the cancer genome. *Nat. Rev. Genet.* *13*, 795–806.
- Yen, R.W., Vertino, P.M., Nelkin, B.D., Yu, J.J., el-Deiry, W., Cumaraswamy, A., Lennon, G.G., Trask, B.J., Celano, P., and Baylin, S.B. (1992). Isolation and characterization of the cDNA encoding human DNA methyltransferase. *Nucleic Acids Res.* *20*, 2287–2291.
- Zhang, R.D., Fidler, I.J., and Price, J.E. (1991). Relative malignant potential of human breast carcinoma cell lines established from pleural effusions and a brain metastasis. *Invasion Metastasis* *11*, 204–215.
- Zhang, X.H.-F., Jin, X., Malladi, S., Zou, Y., Wen, Y.H., Brogi, E., Smid, M., Foekens, J.A., and Massagué, J. (2013). Selection of Bone Metastasis Seeds by Mesenchymal Signals in the Primary Tumor Stroma. *Cell* *154*, 1060–1073.
- (2011). Williams textbook of endocrinology (Philadelphia, Pa: Saunders Elsevier).

(2014a). *Molecular biology of the gene* (Boston: Pearson).

(2014b). *Yen & Jaffe's reproductive endocrinology: physiology, pathophysiology, and clinical management* (Philadelphia, PA: Elsevier/Saunders).