EXAMINATION OF THE BRCA1-DEPENDENT DNA REPAIR PATHWAY IN BASAL-LIKE BREAST CANCERS

Hann-Hsiang Chao

A dissertation submitted to the faculty of the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Curriculum in Genetics and Molecular Biology

Chapel Hill
2012

Approved By:

Charles Perou, PhD

William Kaufmann, PhD

W. Kimryn Rathmell, MD, PhD

Kristy Richards, MD, PhD

Melissa Troester, PhD

ABSTRACT

HANN-HSIANG CHAO: Examination of the BRCA1-Dependent DNA Repair Pathway in Basal-like Breast Cancer
(Under the direction of Dr. Charles Perou)

Human breast cancer is a diverse disease, exhibiting variety in morphology, natural history, and therapeutic response. Multiple studies have shown that breast tumors can be segregated into distinct subtypes, characterized by similarities in the genes they express. One subtype, called basal-like breast tumors (BBT), represents 10-20% of breast cancer diagnoses and is typically associated with poor outcomes. Our work has shown that BBT occurs with significantly higher frequency in women with germline mutations in breast cancer 1 (BRCA1). Given the link between BRCA1-germline mutations and BBT, we proposed to determine whether the BRCA1-dependent DNA repair pathway is deficient during sporadic BBT formation.

Our initial step was to identify BBT specific regions of aberration and determine if they affected important genetic pathways. One region located on chromosome 5q contained multiple BRCA1-dependent repair pathway genes. These genes exhibited frequent co-associated loss with each other and with other cancer relevant genes. Exogenous disruption of these genes in normal breast epithelial cell lines increased sensitivity to DNA damage and impaired BRCA1 localization and function.

We further characterized the genomic instability aspect of BBT by examining a previously undetectable form of genomic aberration we termed micro-

aberrations. These small-scale genomic changes (<5kb in some cases) are detectable using a high-resolution tiling array. We found these events were functional, relevant for survival, preferentially located in the promoter regions of cell cycle genes, and appear most frequently in BBT.

Lastly, we examined BBT-specific loss of the tumor suppressor INPP4B. We found that DNA, RNA, and protein expression of INPP4B are highly correlated with BBT and that it functions as an excellent marker of this subtype as well as predicting survival and response to therapy.

It is critical to gain a greater understanding of BBT function as BBT poses a significant challenge to the US health care system; if BBT were to be treated as a unique disease separate from other breast cancers, it would represent the fourth leading cause of cancer deaths among women. My work here describes newly discovered functional basal-specific aberrations that explain much of the biology of BBT.

ACKNOWLEDGEMENTS

The work described herein was achieved with the help and encouragement of several people. I would like to thank my advisor, Charles Perou, for providing an ideal learning environment for my graduate training and teaching me the skills to independently design, perform, and critically analyze rigorous scientific experiments.

I also would like to thank the members of my thesis committee: William Kaufmann, Kim Rathmell, Kristy Richards, and Melissa Troester. Each person has been instrumental in guiding my graduate training and providing their unique expertise and a wealth of scientific knowledge.

The members of the Perou Lab, both past and present, also greatly enriched my training, and made it a pleasure to come to lab every day. In particular, I want to thank Aaron Thorner, Xiaping He, and Katherine Hoadley for their instrumental assistance at key points of my training.

Lastly, I would like to thank my family, Shu Jen Chao, Jing Wen Chao, Tony Chao, and Yvonne Chao, and my friends for their love and support.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

aCGH:	Array comparative genomic hybridization
ANOVA:	Analysis of variance
BRCA1:	Breast cancer 1, early onset
BRCA2:	Breast cancer 2, early onset
CK:	Cytokeratin
CNA:	Copy number aberrations
DNA:	Deoxyribonucleic acid
DSB:	DNA Double-Strand break
DWD:	Distance weighted discrimination
ER:	Estrogen receptor alpha
HD-94:	High-density tiling array sample set
HER2:	v-erb-b2 erythroblastic leukemia viral oncogene homolog 2
HR:	Homologous recombination
hTERT:	Human telomerase reverse transcriptase
IC50:	Inhibitory concentration 50%
INPP4B:	Inositol polyphosphate-4-phosphatase, type II
IHC:	Immunohistochemistry
HME-CC:	hTERT immortalized human mammary epithelial cell line
ME16C:	hTERT immortalized human mammary epithelial cell line
mRNA:	Messenger RNA
MTT:	3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide
NKI:	Netherlands Cancer Institute

NW:	Norway samples
OS:	Overall survival
PARP:	Poly (ADP-ribose) polymerase
pCR:	Pathologic complete response
PI3K:	Phosphoinositide 3-kinase
PTEN:	Phosphatase and tensin homolog
RB1:	Retinoblastoma 1
RFS:	Relapse-free survival
RD:	Residual disease
RNA:	Ribonucleic acid
SAM:	Significance analysis of microarrays
SMA:	Smooth muscle actin
SNP:	Single nucleotide polymorphism
SWITCHdna:	Sup Wald Identification of copy CHanges in *dna*
TCGA:	The Cancer Genome Atlas
T/FAC:	Taxane, fluorouracil, anthracycline, cyclophosphamide
TMA:	Tissue microarray
UMD:	UNC Microarray Database
UNC:	University of North Carolina, USA samples
UTR:	Untranslated region
TP53:	Tumor protein p53
RAP80:	Ubiquitin interaction motif containing 1
95% CI:	95% confidence interval

CHAPTER ONE

INTRODUCTION

1.1 General Introduction

Background

Breast cancer affects approximately one in seven women worldwide and is the second leading cause of cancer deaths. In recent years, the incidence of breast cancer has risen, owing to a combination of better awareness of the disease, more frequent screenings, and increased longevity. However, over this same time period, overall mortality has decreased thanks to early detection and improvements in patient care. Targeted therapies are a major advancement in the treatment of this disease, such as the use of selective estrogen receptor modulators, e.g. tamoxifen, and drugs that reduce circulating estrogen, e.g. aromatase inhibitors, in patients that have estrogen receptor positive disease. Another example of targeted therapy comes in the form of trastuzumab, which is a monoclonal antibody utilized against HER2 receptor positive breast cancer. Nonetheless, there still remains a significant portion of breast cancers that are unable to be treated by targeted therapy and must be managed by cytotoxic chemotherapy, which is associated with many adverse side effects. Despite current treatment advances, approximately 40,000 women were estimated to die of breast cancer in 2011 in the United States, indicating that additional studies to improve therapy are still needed.

One class of breast cancer in particular, the basal-like breast tumors (BBT), are currently devoid of targeted therapies due to a lack expression of the three current biomarkers for breast cancer that guide therapy, namely the estrogen receptor, progesterone receptor, and the growth factor receptor HER2/ERBB2. My research is focused on this subtype of breast cancer, specifically identifying regions of genomic instability within these tumors that may identify potential targets to develop for therapy. By expanding our understanding of the factors that lead to the progression of this disease, we hope to identify unique characteristics that can be exploited for both better disease prediction and patient treatment.

Normal Breast Development

In order to understand mammary carcinogenesis, one must first understand the normal development of mammary tissue. Normal human breast tissue is very diverse, with an architecture comprised of many different cell types, that also undergoes dramatic changes in shape, size, and function throughout a woman's lifetime (Howard and Gusterson 2000; Russo and Russo 2004). This diversity in mammary cells and their constant state of change contributes to the development of different types of breast tumors.

The mammary gland first develops prenatally as rudimentary duct structures that have two layers of epithelial cells, the basement membrane and the stromal layer (Wiseman and Werb 2002). At birth, the infant breast consists of a primitive, blunt-ended, ductal system comprised of one or two inner layers of luminal (epithelial) cells, and one layer of myoepithelial cells surrounding the lumen (Russo and Russo 2004). After birth, the mammary gland is no longer subject to maternal

hormones, and undergoes differentiation and involution, eventually leaving small tubular structures surrounded by fibroblastic stroma (Anbazhagan et al. 1991).

With the onset of puberty, the rudimentary duct structures elongate and branch, resulting in the terminal duct lobular units, the site of future β-casein (milk)-producing cells, or “alveolar buds" (Villadsen 2005). There is also a dramatic increase in stromal fat and fibroblastic tissue. Once adulthood is reached, the mammary gland is comprised of a main lactiferous duct attached to the nipple that branches into 15-25 milk ducts, which further branch into many sub-segmental ducts. Different cell types, luminal and myoepithelial, constitute the gland and these can be distinguished morphologically and through histological stains (Figure 1.1). The gland is constantly influenced by hormones, with each ovulation promoting a small amount of additional budding.

With pregnancy, a massive increase in proliferation and formation of acini in the alveolar buds occurs (Villadsen et al. 2007). After birth, milk is secreted by the alveolar cells, which flows through the ducts to the nipple. Once lactation is complete, the acini (clusters of milk-producing lobules) involute through a period of apoptosis (Schorr et al. 1999; Strange et al. 2001). Lastly, post menopause, the number of both ducts and lobules are reduced through involution (Howard and Gusterson 2000).

1.2 Breast Cancer Molecular Subtypes and Outcome

Overview

Breast cancer has been shown to be not a single disease, but instead is a very heterogeneous disease with a spectrum of distinct cellular origins, somatic changes, and etiologies. Over the years, clinical decision has expanded beyond factors such as stage, grade, and hormone receptor status to include the use of gene expression predictors as well (Fan et al. 2006). DNA microarray analyses have now shown repeatedly that breast cancer is composed of at least six distinct tumor subtypes, potentially arising from different cell lineages (Prat et al. ; Perou et al. 2000; Sorlie et al. 2001; Sorlie et al. 2003; Hu et al. 2006). These include luminal A, luminal B, normal breast-like, HER2-enriched, claudin-low, and basal-like subtypes. Each subtype has a unique gene expression profile that reflects the biology of the breast tumor and also predicts patient clinical outcomes (Figure 1.2 and Figure 1.3).

Estrogen Receptor-Positive Subtypes

The majority of breast cancer patients (60-80%) present with luminal/estrogen-receptor-alpha positive (ER+) tumors (luminal A and luminal B) (Perou et al. 2000). Although some luminal tumors are non-responsive or become resistant to therapy, they are typically excellent candidates for targeted endocrine therapies such as Tamoxifen, an estrogen receptor antagonist (Paik et al. 2004). Luminal A tumors are the most frequent (35-45%) and both luminal subtypes are characterized by the expression of transcription factors that include ER, GATA3, FOXA1 and XBP1 (Perou et al. 2000; Gruvberger et al. 2001; Sotiriou et al. 2002; van 't Veer et al. 2002; Huang et al. 2003). Luminal B tumors can be distinguished from

luminal A tumors through the "proliferation" signature (Perou et al. 1999; Whitfield et al. 2002), which is high in LumB tumors. The normal breast-like group also clusters with the ER-positive subtype and has a gene expression signature similar to that of normal breast. This group is usually comprised of true normal breast samples and a few tumors, and studies suggest that in most cases this classification is due to a high amount of contaminating normal breast tissue (Hu et al. 2006).

Estrogen Receptor-Negative Subtypes

In contrast to the estrogen receptor positive subtypes are the ER-negative tumors, which are comprised of the HER2-enriched, Basal-like breast tumors, and claudin-low tumors. Basal-like breast tumors will be described separately as they are the focus of this dissertation research. The large majority of ER-negative tumors are represented by the HER2-enriched and BBT subtypes. The HER2-enriched subtype is characterized by the high expression of HER2 and other genes on the HER2 amplicon (Pollack et al. 1999; Pollack et al. 2002). The HER2-enriched subtype classification is noteworthy because of the development of trastuzumab, which is a biologically-directed therapy that targets HER2 (Goldenberg 1999; Burris 2001; Slamon et al. 2001). Another ER-negative subtype is the recently discovered claudin-low subtype (Prat et al.). This relatively small subset of breast tumors (5%) is characterized by the lack of expression of tight junction and cell-cell adhesion proteins including Claudins 3, 4 and 7, occludin, and E-cadherin (Herschkowitz et al. 2007). Like basal-like breast tumors, which will be discussed below, these tumors are pathologically ER-negative, PR-negative, and HER2-not amplified, features

which make these "Triple-Negative" tumors; however, by expression they are a distinct subtype from basal-like breast tumors (Figure 1.2).

Basal-like Breast Tumors

In contrast to the luminal breast tumors are basal-like breast tumors, which are the focus of this work. BBT account for 10-20% of all breast cancers and are often clinically labeled "triple negative". BBT clearly poses a significant challenge to the US health care system, as evidenced by the fact that if BBT were to be treated as a unique disease separate from other breast cancers, it would represent the fourth leading cause of cancer deaths among women (Carey et al. 2006). These tumors do not express ER-alpha, PR, and HER2, have limited treatment options, and patients typically experience poor outcomes (Sorlie et al. 2001; Sorlie et al. 2003; Nielsen et al. 2004; Paik et al. 2004). They often exhibit high expression of cytokeratins 5 and 17, cell surface markers expressed in normal mammary basal/myoepithelial cells, and a transcriptional signature similar to skin basal epithelial cells. It is believed however, that this tumor type is derived from luminal epithelium and not myoepithelium (Sauer 2007). In addition, BBT also have high expression of the proliferation gene signature. It has also been shown that BBT were significantly more frequent in African Americans versus Caucasians and it has been seen in population based studies that BBT display a unique set of risk factors that are sometimes the opposite in effect of these same risk factors in ER+/Luminal tumors (Carey et al. 2006). Indeed, in the Carolina Breast Cancer Study, pregnancy was shown to be associated with an increased susceptibility to develop BBT, with increasing parity and younger age at first full term pregnancy also increasing risk,

which is the opposite of what is observed for luminal A tumors. Lack of breastfeeding was another risk factor for basal-like breast cancer, suggesting there are protective effects of lactation (Carey et al. 2006). BBT also occurs with significantly higher frequency in women with germline mutations in the breast cancer 1 gene (*BRCA1*) (Foulkes et al. 2003; Sorlie et al. 2003; Arnes et al. 2005), thus suggesting possible genetic susceptibilities as well.

1.3 Genetic Changes in Breast Cancer

Patterns of Genomic Instability in Breast Cancer Subtypes

A hallmark of many human cancers is genomic instability and cancer can be thought of as the result of aneuploidy. Specific patterns of copy number aberrations (CNA) can define genetic events associated with different breast cancer subtypes (Figure 1.4). Array comparative genome hybridization (aCGH) studies of tumor copy number states have demonstrated that tumors with similar gene expression subtypes may also share similar DNA Copy Number Aberrations (Russnes et al. ; Bergamaschi et al. 2006; Chin et al. 2006; Wood et al. 2007; Van Loo et al. 2010) and that CNA can be used to further sub-divide expression classes (Russnes et al.). In some cases, tumors have single gene-specific amplicons occurring in a large percentage of cases, such as amplification of *ERBB2* in 25-30% of breast tumors (Kauraniemi et al. 2004). In these cases, treatment options (e.g. trastuzumab) attacking the aberrant gene offers a positive response (Goldenberg 1999; Burris 2001). The luminal A tumors are notable for their distinct lack of genomic instability, often showing aberrations only in chromosome 1q and 16q (Figure 1.4B). In fact, they can be distinguished from luminal B tumors through the higher levels of genomic instability displayed by luminal B tumors (Figure 1.4C). The inherent heterogeneity of tumors is likely to demonstrate that the biological impact of several distinct CNA events initiates tumorigenesis through disruption of genes in specific pathways that are linked to specific cell types.

Basal-like Breast Tumor Genomic Alterations

Nowhere is the case of genomic instability-driven tumorigenesis more prevalent than in the Basal-like subtype, where the majority of its tumor genome exhibits aberrant copy number states (Ding et al. ; Russnes et al. ; Bergamaschi et al. 2006; Chin et al. 2006; Van Loo et al. 2010) (Figure 1.4a). Identifying the genes that contribute to this instability phenotype could be useful not only from a biological perspective, but also possibly as a clinical predictor of response. Previous studies (Loo et al. 2004; Bergamaschi 2006; Chin et al. 2006) have shown that BBT has greater genetic complexity compared to the other breast cancer subtypes (Figure 1.4).

Initial analyses of genetic and gene expression data indicate that various genomic regions are preferentially lost in BBT, suggesting that BBT is a unique entity not only on an expression level, but on a copy number level as well. Many interesting genes are found in these basal-specific regions encompassing a number of important cellular pathways. *INPP4B* is one gene that is specifically deleted in BBT and this gene has been shown to be a key player in the PI3K signaling pathway which is involved in controlling proliferation, cell survival, and genomic instability, among other functions (Fedele et al. ; Puc et al. 2005; Saal et al. 2005; Janzen and Scadden 2006; Baker 2007; Gewinner et al. 2009). Interestingly, some of these regions also contain genes involved specifically in the BRCA1-pathway or more generally in DNA repair. One hypothesis is that in a given BBT, a general deficiency in DNA repair occurs that can arise from deletion(s) in a subset of genes in the BRCA1/DNA repair pathway, such that different regions may be affected in different

tumors, but all affect the same pathway. This deficiency combined with ambient radiation damage or other mutagenic insults may induce BBT formation. BBT may be particularly prone to chromosomal instability, due to the fact that alterations in DNA repair genes/proteins likely contribute to its formation.

Despite the poor prognosis generally associated with BBTs, they are often initially responsive to chemotherapy treatments (Rouzier et al. 2005; Carey et al. 2007). Our explanation for this sensitivity is the proposed DNA repair deficiency in BBT caused by aberrations in BRCA1 and/or other DNA repair proteins, coupled with their loss of RB and TP53 function (Troester et al. 2006; Herschkowitz et al. 2008). In light of the association with BRCA1, this suggests a possible defect in double strand DNA break repair mechanisms, and some evidence of this is seen in the fact that basal-like tumor cell lines are more susceptible to DNA damaging agents as compared to normal human mammary epithelial cell lines (Figure 1.5).

Analyses using an Illumina SNP microarray platform identified numerous regions of copy number loss that were statistically correlated only with BBT, including 5q11-35, 13q12-34, 14q11-32, 15q11-26, 17q12-23, and the X-chromosome. Many of the genes mapped to these regions are part of the BRCA1-pathway and are necessary for DNA repair. Examples include *RAD17* (5q12), *RAD50* (5q23), *RAP80* (5q35), *RAD51* (15q15), *RAD51B* (14q12), *RAD51C* (17q23), and *BRIP1* (17q23). It is also worth noting that a BRCA1-modifier locus for hereditary breast cancer penetrance has been mapped to the 5q region (Nathanson et al. 2002). In addition, when examining the BRCA1-pathway on a gene expression level, the average expression of these highlighted genes is also found to be lowest in the basal-

like subtype (Figure 1.6). It stands to reason that there could be gene-environment interactions in these candidate genes in the form of germline variants associated with susceptibility to BBT. It is our hypothesis that sporadic BBTs do not necessarily have a defect in BRCA1 itself, but instead they may have a general defect in double strand DNA break repair caused by the sporadic loss of one or more DNA repair pathway genes.

There is some evidence suggesting that there is a coordinated loss of this specific pathway as many of the genes are highly co-correlated with respect to their CNA; for example *RAD17, RAD50, RAD51, RAD51L1*, and *RASGRF1* are often lost together in sporadic BBT. This finding can partly be explained by the fact that many of these genes are neighbors and found on the same chromosomal arm, but this is not universally the case. *RAD17* and *RAD50* are both on the 5q arm and have the highest concordance of loss, but *RAD51* and *RASGRF1* are located on a separate chromosome, on the 15q arm. On average, it appears that more than one repair gene is lost in a given BBT. Discovering other gene associations and the biology behind them should give insight into the etiology of these tumors.

1.4 The Role of BRCA1 in Breast Cancer

BRCA1 Function and Association with Basal-like Breast cancer

As previously mentioned, one clue in identifying BBT causative events was the discovery of an association between BBT and *BRCA1* (Sorlie et al. 2003). *BRCA1* is an ubiquitous tumor suppressor that has functions in the nucleus in transcription, gene silencing, chromatin remodeling, and several DNA repair mechanisms (Starita and Parvin 2003). It has been shown to be essential for normal development as *BRCA1* knockout mice show early embryonic lethality (Brodie and Deng 2001). In addition, *BRCA1* is thought to be a guardian of genomic stability because of its involvement in the cellular response pathway to DNA double-strand breaks (DSB) (Venkitaraman 2002). Mouse and human cells deficient for BRCA1 are sensitive to ionizing radiation (IR), indicating a role in DNA DSB repair (Deng and Wang 2003). Surprisingly, although human *BRCA1* mutation carriers develop BBT around 80% of the time, the BRCA1 gene/protein is shown to be normal in the great majority of sporadic BBT (>90%) (Richardson et al. 2006) and there is also, at present, no proven correlation between *BRCA1* methylation and BBT (Matros et al. 2005; Richardson et al. 2006; Turner et al. 2006). Overall, little conclusive data exists concerning BRCA1-pathway function in sporadic BBT. We hypothesized that defects in other members of the BRCA1-pathway, rather than disruption of BRCA1 itself, may contribute to sporadic BBT formation.

BRCA1 DNA Repair Pathway

A number of BRCA1 DNA repair pathway members were shown to be associated with BBT and have low gene expression within the subtype, including

RAP80, RAD50, RAD17, RAD51B, and *RAD51C* (Figure 1.6). RAP80 physically binds to BRCA1 and is required for the localization of BRCA1 to DNA-repair foci (Kim et al. 2007; Sobhian et al. 2007; Wang et al. 2007). RAD50 is part of a BRCA1-associated genome surveillance complex/BASC that is involved in BRCA1-dependent DNA repair (Zhong et al. 1999; Wang et al. 2000). RAD17 is phosphorylated by ATM/ATR in response to DNA damage (Rauen et al. 2000; Bao et al. 2001; Roos-Mattjus et al. 2002; Post et al. 2003). RAD51C is thought to be involved in homologous recombination mediated repair pathway (HR) of double-strand breaks and is part of a complex of DNA repair proteins (Dosanjh et al. 1998; Miller et al. 2002; French et al. 2003). RAD51B, also known as RAD51L1, is also thought to be involved in HR (Takata et al. 2000; Miller et al. 2002). Furthermore, haplo-insufficiency of RAD51B has been shown to lead to centrosome fragmentation, aneuploidy, and hypersensitivity to DNA-damaging agents (Takata et al. 2000; Miller et al. 2002; Date et al. 2006), of which the latter two are hallmarks of BBT.

Many studies have been performed focused on the link between *BRCA1* and breast cancer, but a broader analysis of the BRCA1 pathway as a whole has heretofore been lacking. A more far-reaching analysis looking at the component genes within the basal-like subtype may have the ability to uncover findings that have previously been masked either by examining all breast cancers as a homogenous entity, or using BRCA1 alone as the readout of a functioning repair pathway, when in fact many other disruptions can occur leading to the same result.

1.5 Research Introduction

This work will examine the contribution of the BRCA1-dependent DNA repair pathway and genomic instability to the initiation and progression of BBT, a breast cancer subtype with poor prognosis. My aim is to elucidate how the response to DNA damage contributes to the formation of some breast cancers and to investigate the role of BRCA1 pathway function in sporadic BBT. In addition, the data generated will provide information that will aid in the systematic classification of this complex tumor subtype. More broadly, these studies may offer insight into how the interaction between environmental exposures and pre-existing genetic alterations lead to disease.

Chapter 2 examines basal-like associated copy number events and cell-line functional studies that focus on disrupting two prominent members of the BRCA1-repair pathway in normal cell lines in order to imitate the basal-like tumor genetic environment, and generate a DNA damage sensitizing phenotype. Chapter 3 investigates a novel form of genomic instability that is associated with BBT and may be a hallmark of the defective mechanisms of DNA repair/genomic maintenance in this subtype. Lastly, in Chapter 4, *INPP4B*, a gene that is frequently and specifically lost in BBT and a key member of the PI3K signaling pathway is evaluated as a contributing factor involved in BBT formation.

There is currently a paucity of effective treatments for BBT, which is a leading cause of cancer deaths in women. We must more precisely understand how the BRCA1 pathway is dysregulated in order to take advantage of these defects to design effective treatment regimens for BBT. Greater understanding of the

molecular causes of BBT could be obtained by the eventual generation of mouse models with the same dysregulation events we are investigating. Further characterization of the DNA repair functionality of BBT will increase our understanding of their response to chemotherapeutics, and could identify points of intervention for the prevention and improved treatment of BBT.

1.6 Figures and Legends

Figure 1.1: Normal, human mammary gland stained for identifying markers.

Normal breast tissue stained for proteins that mark luminal epithelial cells (ER, GATA3, CK 8/18) or myoepithelial cells (SMA, CK17, CK5/6).

Figure 1.1

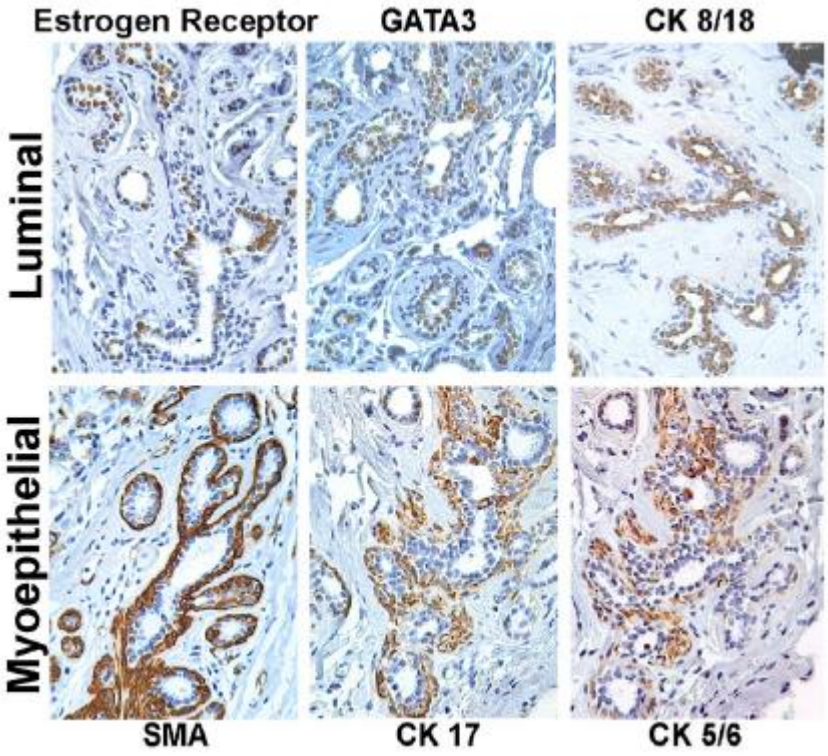

Estrogen Receptor
GATA3
CK 8/18
Luminal
Myoepithelial
SMA
CK 17
CK 5/6

Figure 1.2: Hierarchical clustering of human breast tumors.

A combined dataset of 324 tumor samples collected at UNC and 329 tumors from the NKI, was clustered using the intrinsic gene set (Hu et al. 2006). Clustering identified the six major intrinsic subtypes: luminal A, luminal B, normal breast-like, basal-like and HER2-enriched, and claudin-low.

Figure 1.2

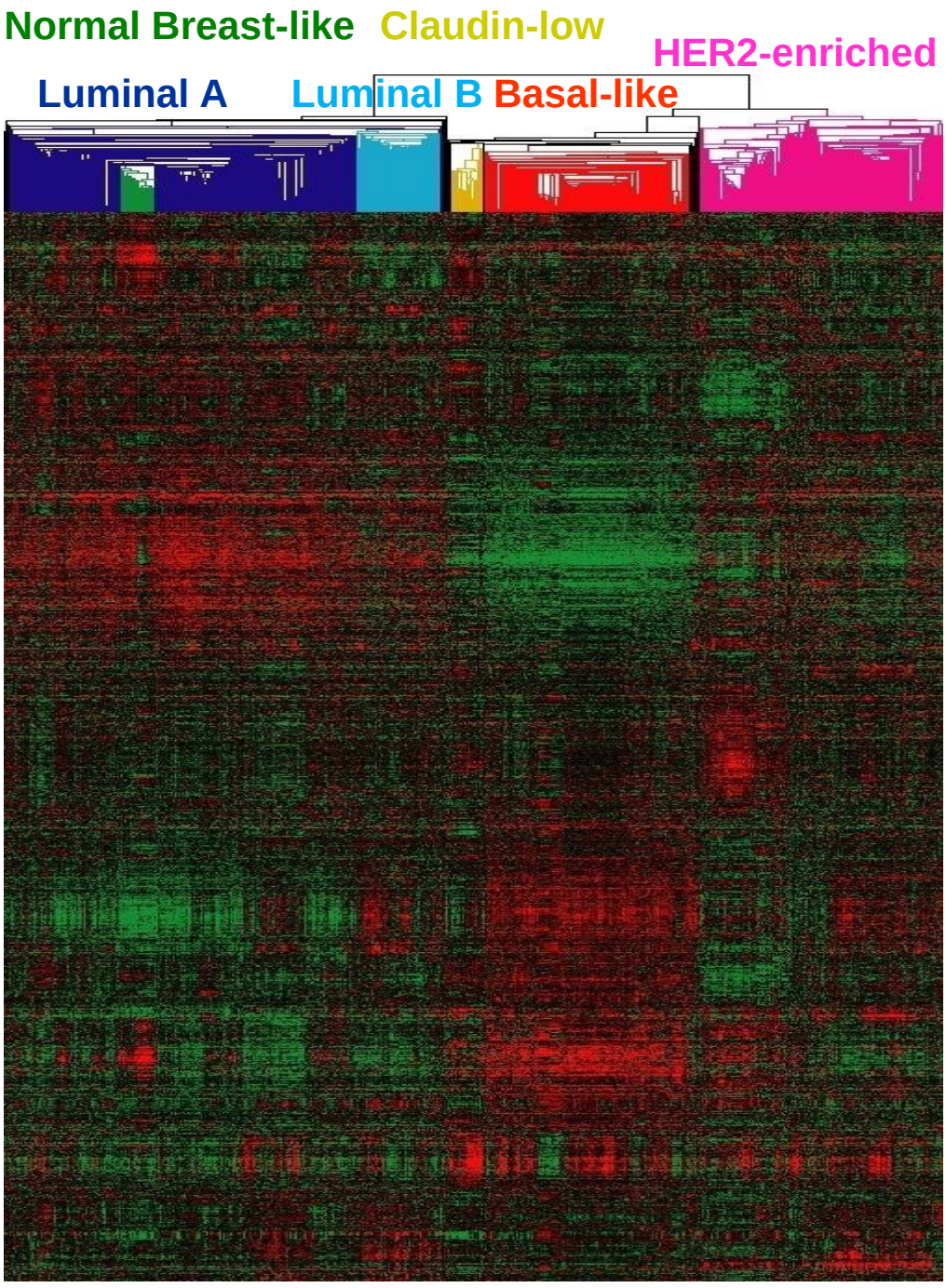

Normal Breast-like
Claudin-low
HER2-enriched
Luminal A
Luminal B
Basal-like

Figure 1.3: Kaplan-Meier plot depicting overall survival for each subtype.

Kaplan-Meier analyses examining overall survival outcomes according to tumor subtype.

Figure 1.3:

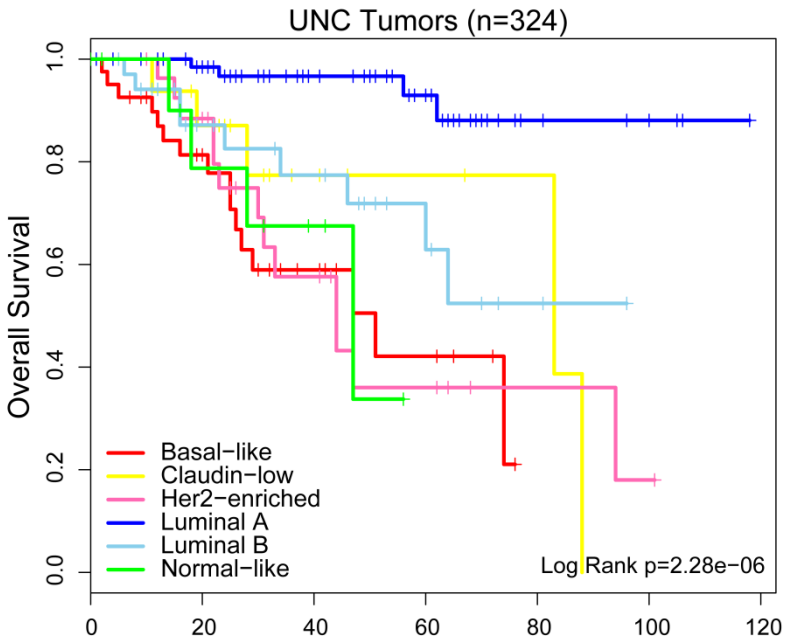

UNC Tumors (n=324)
Overall Survival
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
120
Basal−like
Claudin−low
Her2−enriched
Luminal A
Luminal B
Normal−like
Log Rank p=2.28e−06

Figure 1.4: Genomic landscapes of each breast cancer subtype.

Genomic copy number landscapes indicating the frequency of copy number gains and losses across the genome for A) Basal-like (N=40), B) LumA (N=52), C) LumB (N=40), D) Her2 (N=21), and E) Claudin-low (N=15).

Figure 1.4

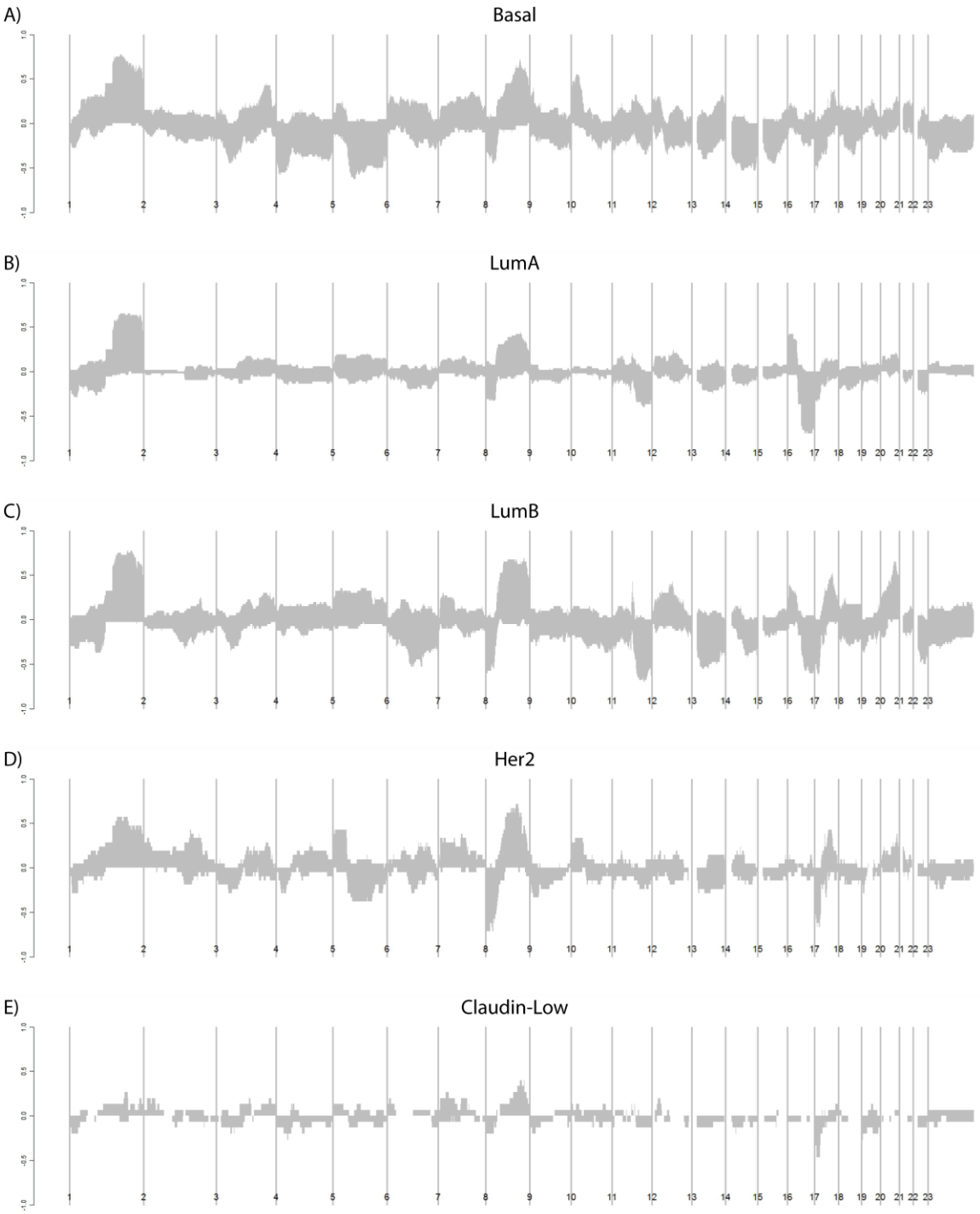


Figure 1.5: PARP inhibitor ABT-888 sensitivity in basal-like cell lines.

IC50 doses (72h) of ABT-888 on a normal, basal-like cell line (HME-CC) versus that of two basal-like tumor cell lines (SUM149, SUM102). Each MTT experiment was performed in triplicate and error bars represent 95% confidence intervals.

Figure 1.5

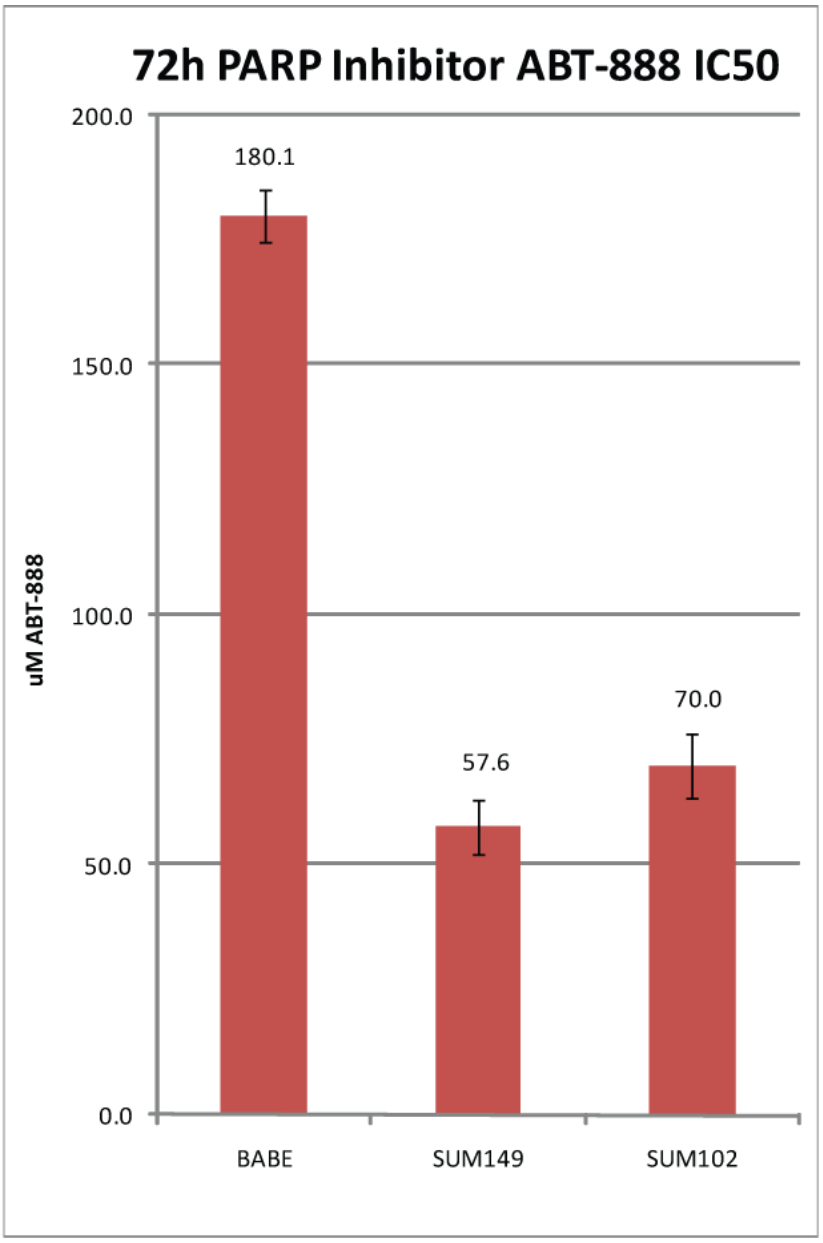
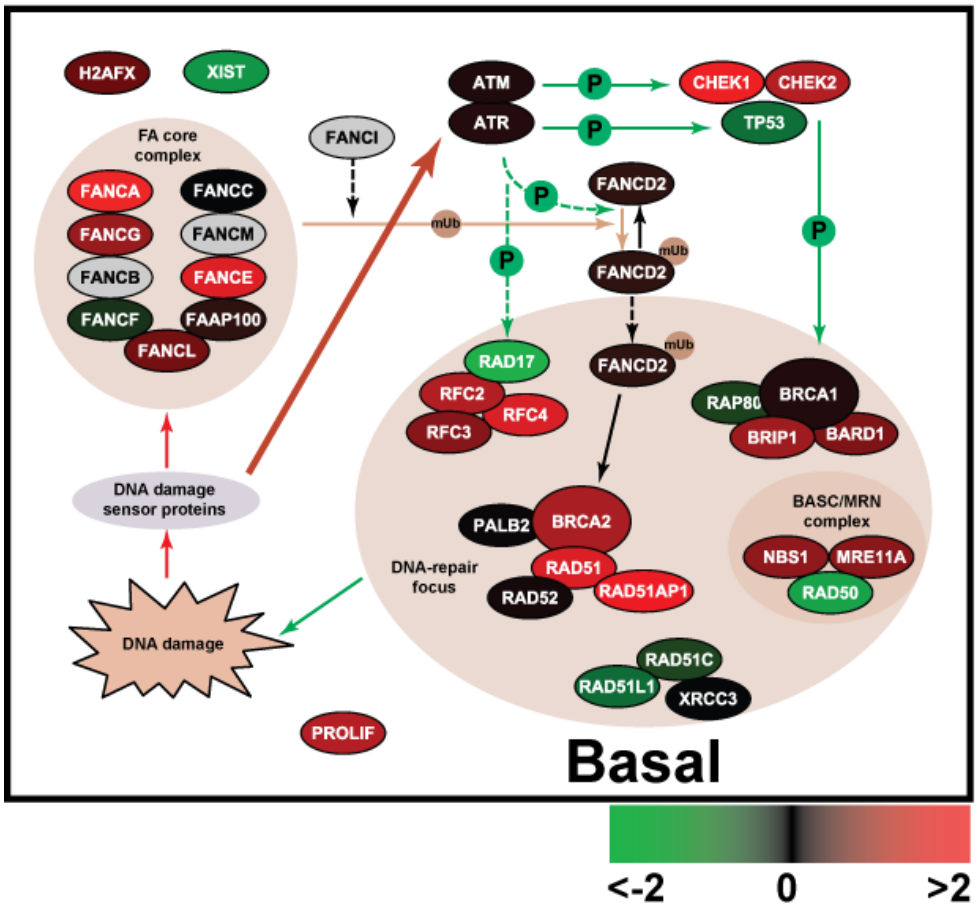

72h PARP Inhibitor ABT-888 IC50
uM ABT-888
200.0
150.0
100.0
50.0
0.0
180.1
57.6
70.0
BABE
SUM149
SUM102

Figure 1.6: Expression of genes in the BRCA1/DNA repair pathway in basal-like breast cancer.

Data is displayed using Cytoscape 2.61, a pathway modeling program. Average gene expression for basal-like breast cancer is shown relative to the Luminal A subtype from a dataset of 250 tumors with known subtype, in which the average gene expression value for each gene was calculated for each subtype.

Figure 1.6


H2AFX
XIST
FA core complex
FANCA
FANCC
FANCG
FANCM
FANCB
FANCE
FANCF
FAAP100
FANCL
FANCI
ATM
ATR
P
CHEK1
CHEK2
TP53
FANCD2
mUb
RAD17
RFC2
RFC4
RFC3
RAP80
BRCA1
BRIP1
BARD1
DNA damage sensor proteins
PALB2
BRCA2
RAD51
RAD52
RAD51AP1
DNA-repair focus
BASC/MRN complex
NBS1
MRE11A
RAD50
DNA damage
RAD51C
RAD51L1
XRCC3
PROLIF
Basal
<-2
0
>2

CHAPTER TWO

BASAL-LIKE BREAST CANCER DNA COPY NUMBER LOSSES IDENTIFY GENES INVOLVED IN GENOMIC INSTABILITY, RESPONSE TO THERAPY, AND PATIENT SURVIVAL

2.1 Introduction

Breast cancer is a heterogeneous disease with known expression defined tumor subtypes. DNA copy number studies have suggested that tumors within gene expression subtypes share similar DNA Copy Number Aberrations (CNA) and that CNA can be used to further sub-divide expression classes. To gain further insights into the etiologies of the intrinsic subtypes, we classified tumors according to gene expression subtype and next identified subtype-associated CNA using a novel method called SWITCHdna, using a training set of 180 tumors and a validation set of 359 tumors. Fisher's Exact tests, Chi-square approximations, and Wilcoxon rank-sum tests were performed to evaluate differences in CNA by subtype. To assess the functional significance of loss of a specific chromosomal region, individual genes were knocked down by shRNA and drug sensitivity and DNA repair foci assays performed. Most tumor subtypes exhibited specific CNA. The Basal-like subtype was the most distinct with common losses of the regions containing *RB1*, *BRCA1*, *INPP4B*, and the greatest overall genomic instability. One Basal-like subtype associated CNA was loss of 5q11-35, which contains at least three genes important

---

Victor J. Weigman*, Hann-Hsiang Chao*, Andrey A. Shabalin, Xiaping He, Joel S. Parker, Silje Nordgard, Tatyana Grushko, Dezheng Huo, Chika Nwachukwu, Andrew Nobel, Vessela N. Kristensen, Anne-Lise Børresen-Dale, Olufunmilayo I Olopade, and Charles M. Perou. Basal-like Breast Cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Research and Treatment.* DOI: 10.1007/s10549-011-1846-y.

for BRCA1-dependent DNA repair (*RAD17*, *RAD50*, and *RAP80*); these genes were predominantly lost as a pair, or all three simultaneously. Loss of two or three of these genes was associated with significantly increased genomic instability and poor patient survival. RNAi knockdown of RAD17, or RAD17/RAD50, in immortalized human mammary epithelial cell lines caused increased sensitivity to a PARP inhibitor and carboplatin, and inhibited BRCA1 foci formation in response to DNA damage. These data suggest a possible genetic cause for genomic instability in Basal-like breast cancers and a biological rationale for the use of DNA repair inhibitor related therapeutics in this breast cancer subtype.

Previous gene expression profiling studies of human breast tumors have shaped our understanding that breast cancer is not one disease, but is in fact many biologically separate diseases. A classification of tumors by expression profiling into five distinct groups (Luminal A, Luminal B, HER2-enriched, Basal-like, and Claudin-low subtypes) has added prognostic and predictive value to the existing repertoire of biomarkers for breast cancer (Prat et al. ; Perou et al. 2000; Sorlie et al. 2001; Sorlie et al. 2003; Hu et al. 2006; Parker et al. 2009). For many cancers, improper maintenance of genome stability is a major cause of tumorigenesis and thus, the characterization of the tumor genomic DNA landscape is an important avenue of investigation (Pinkel and Albertson 2005). Array comparative genome hybridization (aCGH) studies of tumor copy number states have demonstrated that tumors with similar gene expression subtypes may also share similar DNA Copy Number Aberrations (CNA) (Russnes et al. ; Bergamaschi et al. 2006; Chin et al. 2006; Wood et al. 2007; Van Loo et al. 2010) and that CNA can be used to further

sub-divide expression classes (Russnes et al.). In breast cancers, genomic instability-driven tumorigenesis is most prevalent in the Basal-like subtype (also referred to as Triple-Negative Breast Cancers), where the majority of tumors exhibit many CNA (Ding et al. ; Russnes et al. ; Bergamaschi et al. 2006; Chin et al. 2006; Van Loo et al. 2010). Identifying the genes that contribute to this instability phenotype would be useful not only from a biological perspective, but also possibly as a clinical predictor of therapeutic response.

2.2 Results and Discussion

Identifying Subtype-Specific Regions of Copy Number Aberration

To identify CNA that might be causative of Basal-like breast cancers, we assembled a dataset of 180 tumors with Agilent gene expression microarrays and Illumina 109,000 SNP marker DNA copy number microarrays (UNC-NW). We classified each tumor into one of five previously defined expression subtypes using the published intrinsic subtypes (i.e. PAM50) and Claudin-low subtype predictors (Prat et al. ; Parker et al. 2009). To identify regions of copy number gain/loss, we developed a new segmenting method called "SWITCHdna" (Sup Wald Identification of copy CHanges in *dna*). Specifics of the SWITCHdna method can be found at https://genome.unc.edu/pubsup/SWITCHdna/.

SWITCHdna identified regions/segments of copy number gains and losses in each tumor, which were then aggregated based on subtype to look at the frequency of each copy number event in each subtype and identify regions specific to each subtype (Figure 2.1 and Supplementary Table 2.1). A heatmap display of the copy number data is provided in Supplementary Figure 2.1. A number of new findings were observed including the first aCGH characterization of the Claudin-low subtype (Figure 2.1B). Despite its high grade and similarity to Basal-like tumors (Prat et al. ; Parker et al. 2009), Claudin-low tumors showed few copy number changes, which may correspond to the previously described ER-negative and copy number neutral tumor subtype reported in Chin et al (Chin et al. 2007). In addition, human Claudin-low cell lines, which are often called "Basal B" lines, also have a similar flat copy number profile of showing very few chromosomal abnormalities (Kao et al. 2009).

We next searched for CNA occurring specifically within each subtype (Figure 2.1A-F, black shading). The Basal-like subtype had the most subtype-specific events (Figure 2.1A, G) including the previously described amplicon at 10p containing *MAP3K8*, *ZEB1*, and *FAM107B* (Ding et al. ; Adelaide et al. 2007; Bergamaschi et al. 2008), 16q loss (Haverty et al. 2008), deletion of 5q11-35 (Bergamaschi et al. 2006), and deletion of 4q. This last region contains *INPP4B*, which has recently been identified as a potential tumor suppressor involved in the inhibition of PI3K signaling (Gewinner et al. 2009) and that is selectively lost in Basal-like/Triple-negative breast cancers (Fedele et al.).

Basal-like tumors have previously been observed to have copy number loss and/or low expression of genes involved in BRCA1 DNA damage repair (Natrajan et al.), and we noted that loss of 5q11-5q35 would delete several genes involved in BRCA1-dependent DNA repair including *RAD17*, *RAD50* (Johannsdottir et al. 2006), and *RAP80* (Figure 2.1H). Closer examination of the pattern of loss of these genes revealed that each gene was rarely lost as an individual event, but predominantly lost as a pair or triplet (Table 2.1A). These doublet or triplet losses occurred at the highest rates in the Basal-like subtype, but also occurred less frequently in the HER2-enriched subtype. These paired or triplet losses were not simply due to loss of the entire chromosomal arm as >65% of the analyzed tumors did not show a loss pattern indicative of such an event and several samples had intervening regions of normal copy number. Loss of 5q11-35 was also found to statistically co-occur with CNA of other regions including 10p amplification (~50%), *INPP4B*/4q31.21 loss (~40%), *PTEN*/10q23.31 loss (~40%), *BRCA1*/17q21 loss (~50%), and most

frequently loss of *RB1*/13q14.2 (~80%) (Table 2.1E), which are genes/regions that have all been previously shown to be associated with Basal-like Breast Cancers.

In order to validate these subtype-specific findings observed in the UNC+NW dataset, we classified the samples in Jonsson et al. (Jonsson et al.) according to PAM50 and Claudin-low subtype predictors and performed similar supervised analyses using their BAC-based DNA copy number data; very similar associations between CNA and subtypes were observed (Table 2.2). Jonsson et al. identified 6 unique tumor subtypes based upon CNA landscapes, which we determined were highly correlated with our expression defined intrinsic subtypes (p-value < 0.001, Table 2.3); importantly, there was high overlap between our Basal-like subtype and their Basal-Complex phenotype, both of which showed the frequent loss of 5q11-35 and amplification of 10p.

Increased Genomic Instability of Tumors Associated with Loss of Specific Regions/Genes

To objectively assess "genomic instability", we calculated a loss/normal/gain value for every gene using the SWITCHdna assigned copy number states, and calculated the levels of genomic instability by subtype using the average number of gains/losses per sample on a gene by gene basis. The Basal-like subtype was the most prone to aberrations, while the Claudin-low and Luminal A subtypes showed the lowest number of gene-based CNA (Table 2.1C). To control for a large number of genes being gained or lost by a large single genomic aberration event (i.e. whole chromosome loss), we also calculated the average number of SWITCHdna-defined segments and their length for each subtype, as more genomic breaks will result in

more segments. The subtypes that had greater numbers of gene aberrations were also the same ones that had more SWITCHdna segments of shorter average length (Table 2.1C). Thus, the increased number of aberrant gene-based events in the copy number unstable subtypes was due to more frequent aberrations in the genome, rather than as a large number of genes gained or lost by a few large-in-size aberration events.

Tumors with loss of *PTEN*/10q23.31, *RB1*/13q14.2, or *TP53*/17p13.1, or amplification of the 10p region were also found to have high rates of total gene-based CNA compared to tumors without loss of these genes (Table 2.4). Loss of 5q11-35 was also associated with the highest numbers of CNA, with the greatest instability seen when all three DNA repair genes were lost.

Low Expression of Genes Residing in Basal-like Regions Correlates with Poor Survival and Predicts Therapeutic Response

To determine if these DNA loss events also impacted gene function, we determined whether the mRNA levels of candidate genes contained within these regions correlated with DNA loss. The expression of 10 genes selected based on their associations with the basal-like subtype, or breast cancer in general, was evaluated. Most of these genes showed significantly lower mRNA expression when the genomic DNA was lost including *RAD17, RAD50, RAP80, MSH3*, *RB1*, *PTEN*, *BRCA1*, and *INPP4B* (Figure 2.2); these data suggest that these losses have functional consequences (noting that only *TP53* and *BRCA2* did not show in *cis* correlation between expression and copy number). It is also of note that *MSH3* (a gene involved in DNA mismatch repair), is located within the 5q11-35 loss region (between *RAD17*

and *RAD50*, Figure 2.1H), and it also showed reduced mRNA expression when lost and low expression within Basal-like tumors in general (Figure 2.2, 2.3E). In addition, the mRNA expression levels of *RAD17*, *RAD50*, *MSH3*, *RAP80, INPP4B*, and *PTEN* were lowest in the Basal-like subtype (Figure 2.3, UNC337 expression dataset (Prat et al.)); thus loss of 5q11-35 likely affects multiple aspects of DNA repair.

Using patient survival data from two additional data sets containing gene expression data (UNC337 (Prat et al.) and NKI295 (van de Vijver et al. 2002)), Kaplan-Meier analysis showed that the low average expression of *RAD17+RAD50* was associated with worse outcomes compared to high expression (Figure 2.4A). A similar trend was observed with *INPP4B*, mirroring previous observations (Figure 2.4B) (Gewinner et al. 2009). *RAD17+RAD50* expression was also examined for treatment effects using the Hess et al. dataset, which examined T/FAC neoadjuvant chemotherapy responsiveness across 130 breast cancer patients (Hess et al. 2006). Low expression of *RAD17+RAD50* was correlated with pathological complete response (pCR) (ANOVA p-value <0.0001). This finding may be due to the association between low expression of RAD17+RAD50 and Basal-like tumors, as Basal-like tumors have also been shown to have high neoadjuvant chemotherapy pCR rates (Rouzier et al. 2005; Carey et al. 2007).

Knockdown of RAD17 +/- RAD50 Affects Sensitivity to Chemotherapeutics and BRCA1 Foci Formation

Given the involvement of *RAD17*, *RAD50* and *RAP80* in the BRCA1-DNA repair pathway, we determined whether disruption of these genes via RNAi knockdown would lead to changes in sensitivity to drugs whose mechanism of action has

already been linked to BRCA1 loss like carboplatin/cisplatin (Chang et al. ; Silver et al.) and PARP inhibitors (Huang et al. 2003; Donawho et al. 2007). RAD17 was stably knocked down with shRNA in the HME-CC cell line (an hTERT immortalized Human Mammary Epithelial Cell) (Troester et al. 2004) and knockdown was confirmed by Western blotting (Figure 2.5A). HME-CC cells with RAD17 knockdown exhibited increased sensitivity to ABT-888 (PARPi) and carboplatin (Figure 2.5C). No difference in paclitaxel sensitivity was observed, which was used as a non-DNA damaging agent control. A RAD50 knockdown line did not exhibit any change in sensitivity to ABT-888 and had a paradoxical increase in resistance to carboplatin. We next emulated the most common *in vivo* co-occurring loss by generating a double knockdown of RAD17 and RAD50, which showed the greatest increased sensitivity to ABT-888 and carboplatin (Figure 2.5C). Similar results were observed when this experiment was repeated in ME16C cells, a second hTERT-immortalized human mammary epithelial cell line (Supplementary Figure 2.2).

In order to assess the effects of RAD17/RAD50 loss on BRCA1-dependent DNA repair, we performed a DNA repair foci formation assay on the control and RAD17+RAD50 double knockdown line. Using anti-BRCA1 protein immunofluorescence, and automated foci counting within geminin-positive cells, we observed a significant decrease in the number of BRCA1-containing DNA repair foci in the double knockdown line when treated with ionizing radiation or ABT888 versus control (Figure 2.6); cells were simultaneously stained for geminin in order to control for differences in proliferation as described by Graeser et al. (Graeser et al. ; Gonzalez et al. 2004). These data suggest that loss of RAD17 and/or RAD50 may

impair BRCA1 function, and could contribute to increased sensitivity to DNA damaging agents.

Discussion

The presence of distinct breast cancer expression subtypes suggests different underlying genetic events may be driving each subtype. To address this hypothesis, we used 180 diverse tumors and performed supervised analyses of their tumor DNA copy number landscape and identified subtype-specific copy number events. Many studies have identified numerous regions of gain and loss in human breast tumors (Jonsson et al. ; Bergamaschi et al. 2006; Chin et al. 2006; Fridlyand et al. 2006; Haverty et al. 2008); however, most did not specifically search for regions uniquely associated with specific intrinsic subtypes. Some previous attempts were made to identify basal-like specific CNA (Bergamaschi et al. 2006; Adelaide et al. 2007) and we observed a number of the same findings. We take these previous findings as validation of our identified regions, and we build and expand upon these here, along with the addition of functional studies.

Overall, we identified many subtype-specific CNA and validated these findings on a second, independent dataset. Here we have focused on the Basal-like subtype, which showed by far the greatest number of subtype specific CNA and were the most genomically unstable as determined by the sheer number of CNA, a feature which has been observed in the past (Chin et al. 2006). Basal-like tumors also showed consistent loss of 4q (which harbors *INPP4B* and *FBXW7*), and 5q11-35, which contains many DNA repair genes. Basal-like tumors are known to be associated with BRCA1-pathway dysfunction in that 80-90% of BRCA1 mutation

carriers, if and when they develop breast cancer, develop Basal-like tumors (Foulkes et al. 2003; Sorlie et al. 2003; Arnes et al. 2005); however, in most sporadic Basal-like tumors, the *BRCA1* gene appears normal in sequence (Richardson et al. 2006). The loss of 5q11-35 may provide an alternative means to impair BRCA1-pathway function and explain why despite many Basal-like patients having normal BRCA1 gene/protein, high levels of genomic instability and a "BRCAness" phenotype are observed in Basal-like tumors. Previous evidence indicates a link between genes involved in BRCA1 DNA damage control and genes that are deleted and downregulated in Basal-like cancers, lending further credence to our hypothesis (Natrajan et al.).

In order to expand our understanding of the relationship between the Basal-like subtype and impaired BRCA1-pathway function, we pursued functional studies by RNAi mediated knockdown of two members of the pathway, RAD17 and RAD50, in order to emulate the genomic losses observed in tumors. Besides being members of the BRCA1-pathway, others have highlighted these genes for their possible Basal-like association, but without functional studies (Bergamaschi et al. 2006; Johannsdottir et al. 2006). We show here that genetic ablation of these genes results in impaired DNA repair and increased drug sensitivity, and furthermore, deletion of RAD17 and RAD50 in yeast has also been shown to result in increased sensitivity to DNA-damaging agents including platinum drugs (http://fitdb.stanford.edu) (Hillenmeyer et al. 2008); these data highlight that there is an evolutionarily conserved role for these genes in DNA repair.

By building upon the discovery of the subtype association and the deletion phenotypes in yeast, we propose a role in DNA repair function for the 5q11-35 region. The drug sensitivity assays show the importance of these genes in DNA damage sensitivity and the foci formation experiments show that their function is mediated through BRCA1. Additionally, from the combination of our genomic analyses and functional data, it is our hypothesis that the somatic loss of *RAD17*, *RAD50* and/or *RAP80* leads to impaired BRCA1-pathway function, impaired homologous recombination mediated DNA repair, and thus, contributes to overall genomic instability.

There are, however, two caveats to these analyses and our hypothesis. First, the 5q11-35 loss is a large region that typically involves >100 genes, therefore, we cannot definitively say that loss of these 3 genes is the target of this deletion, or that these 3 genes are the most important targeted genes of this region. Second, a high frequency of co-occurrence with other DNA chromosomal losses happens in tumors with 5q11-35 loss; for example, in ~80% of tumors with 5q11-35 loss, *RB1*/13q14.2 DNA loss also occurs (and by itself is associated with increased genomic instability). In addition, ~60% of these tumors show *TP53*/17p13.1 loss (Table 2.1 and 2.2). The co-occurrence of 5q11-35 loss with *RB1* and *TP53* loss are likely causative events in Basal-like carcinogenesis (the latter two being corroborated by mouse studies) (Cressman et al. 1999; Herschkowitz et al. 2007; Jiang et al. 2010). Given the high co-occurrence of chromosome region losses that are not physically linked, it is impossible to say which one is the cause of the genomic instability. However, our hypothesis is that each of these regions harbors genes needed for maintenance of

the genome and that the combinatorial loss of 2-3 of these regions is what results in the genomic instability phenotype seen in Basal-like breast cancers. In this paper, we examine DNA losses, but do note that it is possible that loss of these same genes could also occur via methylation, altered microRNA regulation, and/or somatic mutation, although the latter has yet to be found when searching current somatic mutation databases for *RAD17/RAD50/RAP80*. Preliminary sequence analysis of *RAD17* and *RAD50* (data not shown), as well as evaluation of previous breast cancer sequencing efforts (Sjoblom et al. 2006) and the COSMIC database (Bamford et al. 2004), revealed few, if any, somatic variants/mutations in these two genes, which is consistent with the finding that loss of any one gene is rarely seen; thus, if loss of two or more genes is the target of this CNA, then somatic mutation of any one gene would not impart a selective tumorigenic advantage. Therefore, these data suggest that the target of 5q11-35 loss is two or more genes in this region, with loss of *RAD17* and *RAD50* likely contributing to genomic instability.

Conclusions

The gene expression-defined intrinsic subtypes of breast cancer are mirrored by DNA copy number changes. The Basal-like subtype is the most distinct in the copy number landscape world, and these subtype associated CNA have clinical implications. If 5q11-35 loss results in impaired homologous recombination mediated DNA repair, as was suggested by our *in vitro* studies and *in vivo* correlates, then the loss of this region may sensitize tumors to specific classes of DNA damaging agents. Based upon BRCA1 studies *in vitro* (Quinn et al. 2003; Farmer et al. 2005) and *in vivo* (Silver et al. ; Donawho et al. 2007), these drugs could include PARP

inhibitors and cis/carboplatin. Loss of *RAD17+RAD50* (mRNA and/or genomic DNA) may thus be a biomarker of chemotherapy responsiveness, which is supported by our finding of an association for predicting a likelihood of achieving a pathological complete response. We hypothesize that the loss of these DNA repair genes and the 5q11-35 region, contributes to genomic instability and mutability, ultimately causing high proliferation rates and aggressive behaviors. Our integrated studies of gene expression and genomic DNA copy number have identified important pathway-based determinants of Basal-like cancers and a possible therapeutic biomarker.

All relevant gene expression and copy number data new to this manuscript can be found in the GEO database under series GSE10893.

2.3 Materials and Methods

Breast Cancer Patient Datasets

Three patient datasets were used in this study, each of which contained gene expression and DNA copy number microarray data. We combined 2 sets into a single training set (n=180 with expression and copy number) so that we could have increased statistical power to detect subtype-specific CNA. The combined training set included one set of breast tumors from the United States ("UNC") (n=77) and another set of tumors from Norway ("NW") (n=103). The third data set ("Jonsson") was used as a validation/testing set (n=359)(Jonsson et al.). All samples were collected using IRB-approved protocols and all patients signed informed consent forms. Tumors in the training set were assayed for gene expression patterns using Agilent DNA microarrays. In total, the UNC gene expression cohort contained 337 human breast tumors (UNC337), taken from Prat et al. 2010 (Prat et al.). Log_2 ratio data was taken from the UNC Microarray Database (UMD), filtering for a lowess normalized intensity value of 10 or above for each channel, and 70% good data, and then used for further analyses. Data is available from Gene Expression Omnibus under series GSE10893 (http://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?token=tdodziqaacgembc&acc=GSE10893). Sample information including clinical data, subtype, source, GEO Sample ID and overlap with copy number information, can be found in the UMD (https://genome.unc.edu) under published data (Weigman et al.). A subset of this UNC expression data set (n=77) was assayed for copy number changes using

Infinium Human-1 109K BeadChip (Illumina, San Diego, CA, USA) and results are described here.

The Norway (NW) data set consists of 103 previously published breast tumors assayed on custom Agilent Microarrays for gene expression (GSE3985), and for copy number changes also using Infinium Human-1 109K BeadChip (Naume et al. 2007; Nordgard et al. 2008) (also listed under GEO GSE10893). The Jonsson dataset consists of 359 breast tumors assayed using custom produced oligonucleotide microarrays for gene expression, BAC microarrays for copy number, and which has been previously published (Jonsson et al.) in GEO series GSE22133. Clinical and gene expression data was also used from an additional dataset, the Netherlands Cancer Institute breast cancer dataset (NKI-295, n=295) (van de Vijver et al. 2002).

Classifying tumors for expression-based subtype classification

The Lowess normalized R/G Log_2 ratio data from the UNC337, NW103 and Jonsson-359, were used and median centered independently within each data set, prior to collapsing (via averaging) from probes to HGNC gene symbols. Datasets were then limited to the gene symbols shared across them. After column standardization of both datasets (samples at N(0,1)), Distance Weighted Discrimination (DWD) (Benito et al. 2004) was used to remove platform bias and combine the UNC and NW datasets prior to classification. The PAM50 gene set predictor(Parker et al. 2009) in conjunction with a newly identified Claudin-low predictor(Prat et al.) was used to assign subtypes to the tumors in these three datasets.

Assessment of tumor genomic DNA copy number changes

77 UNC and 103 NW samples had blood/normal and tumor DNA pairs each assayed using the Infinium Human-1 109K BeadChip (Illumina, San Diego, CA, USA), with an average resolution of 1 probe per 26kb (median of 1 probe per 13kb). Tumor content for each sample was assessed by histology to be greater than 50% tumor tissue. Additionally, tumor cellularity was also assessed genomically by the ASCAT (Van Loo et al. 2010) and genoCNA (Sun et al. 2009) algorithms. Each sample was labeled, hybridized, scanned and analyzed according to the manufacturer's protocol (Gunderson et al. 2005). To account for germline specific copy number polymorphisms, a pool of 118 blood-drawn DNA samples was compiled and used as Illumina's standard clustering reference through the BeadStudio software, which estimates allele-specific intensity boundaries. Genotype reports and LogR values were extracted with reference to dbSNP's (build 125) forward allele orientation using BeadStudio (v3.1, Illumina, San Diego, CA, USA). Sample information is in Additional Data File 1 and LogR (A+B signal) values can be found on GEO under series GSE10893, platform GPL8139.

Identification of CNA with SWITCHdna

To determine regions of copy number aberration (CNA), we developed a new analysis method that is a modification of the SupWald method (Andrews 1993; Bai 1998), where we created an R suite of functions called "SWITCHdna", to identify breakpoints in aCGH data. Statistical problems in aCGH analysis are the detection of transition points (positions where DNA copy numbers change) along with identifying most-likely false positive candidates. In SWITCHdna, the estimation of

number of transition points (K) is performed sequentially. First, we test for the presence of at least one transition point (versus none) across the entire chromosome. If the test shows no significant deviation from the null model (or no change points) we accept K=0. Otherwise, we find the change point that maximizes the F statistic and continue and test for the presence of the two change points (versus one) and so on. We accepted a change point when the minimum of two criteria were met: 1) size of region on either side of the breakpoint was larger than α (number of observations (in our case, array probes) which would cover user-defined range) and 2) the F statistic of each segment was the maximum value possible and above the user-specified threshold. Following detection of the transition points, a segment average value and corresponding z-score, are determined along with the number of observations used.

A significance filter was applied to the raw SWITCHdna-identified segments in order to reduce noise and increase the probability of identifying biologically relevant regions. To determine this significance cutoff, we created 100 permuted datasets from our LogR ratio values, which had values shuffled within the space of each chromosome. SWITCHdna was then run on each of these permuted datasets to gather information about random segment statistics. We found that a z-score of 3 or greater, and intensity cutoff of 0.1 was sufficient to select for values with a familywise error rate of 5% (data not shown). All subsequent plots and tables were produced after applying this significance filter to our data. Plotting functions (plot.freq.SW) display the prevalence of subtype-shared CNA in genome-space.

SWITCHdna is provided as a source script in R(R Development Core Team 2009) and available for download at: https://genome.unc.edu/pubsup/SWITCHdna/.

Determining subtype-specific CNA

Using the cnaGENE function of SWITCHdna, the segment output file was converted into an indicator matrix, where for each sample, each gene's copy state was represented as -1 = loss, 0 = no change, 1 = gain. This happens first on a segment basis, with individual gene data then being taken from the segment value; specifically SWITCHdna first identifies the gain/loss segments of a sample, the cnaGENE function filters segments by providing thresholds for Z score and Intensity levels (that can be determined through permuting probe loci and rerunning SWITCH). When a segment passes filtering, all genes that fall completely within that segment are assigned a gain/loss status as appropriate to the sign of the segment. For genes that may be split by a segment, we assigned a value of unchanged=0, because it was split and thus, giving 2 possible values; the split genes were a minority of the genes and this only affected <0.5% of genes across all samples. For each subtype, the counts of gains and losses were compared versus all other samples in order to identify subtype-specific CNA. A Fisher's Exact test was performed on the subtype vs. rest counts for each gene. The resulting p-values were adjusted by the Benjamini-Hochberg method (Benjamini and Hochberg 1995) to correct for multiple-hypothesis testing and then genes with p-values < 0.05 were then gathered for each subtype. Regions within the cytobands of localized CNA were determined by the significant genes found within each cytoband, allowing for more separation of CNA loci (Supplementary Table 2.1).

RNAi Cell Line experiments

Stable RNAi knockdown lines were generated from hTERT-immortalized, human mammary epithelial cell lines (HME-CC and ME16C) cultured as described in Troester et al (Troester et al. 2006). Genes were targeted with shRNA constructs for *RAD17* (Oligoengine) and *RAD50* (Origene) (**RAD17** – UGCCAUACCUUGCUCUACU, GUGGAAAGACAACGACCUU; **RAD50:** TI308519 – CTTCAGACAGGATTCTTGAACTGGACCAG, CTCAACTGTGGCATCATTGCCTTGGATGA) and control lines were concurrently made using a non effective construct (Origene – TR30003, Oligoengine – VEC-PRT-0006). 10μg of each vector was transfected into Phoenix 293T cells using Lipofectamine 2000 (Invitrogen) as per manufacturer's instructions to produce retrovirus. Viral supernatant was collected and applied to the cell line with 75ug of polybrene. Stable populations were selected by culturing in 400mg/mL geneticin for the RAD17 knockdown line and its control and 1mg/mL puromycin for the RAD50 knockdown line and control.

Western Blot Analysis

Cells were grown in 10-cm tissue culture-treated dishes to 80% confluence and then harvested, protein isolated, and quantified as described previously (Troester et al. 2004). Membranes were probed for RAD17 (AB3261; Millipore, Billerica, MA, USA), RAD50 (#3427; Cell Signaling Technology, Danvers, MA, USA), alpha-tubulin (sc-9104, Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA), or actin (GTX 21801, GeneTex, Irving, CA, USA) followed by anti-rabbit or anti-goat IgG horseradish peroxidase-linked whole antibody (sc-2020; Santa Cruz Biotechnology, Inc., Santa Cruz, CA, USA or Amersham Biosciences, Piscataway, NJ USA) and

detected using SuperSignal West Pico Chemiluminescent Substrate (Pierce, Rockford, IL, USA).

Cell Cytotoxicity Assays

Sensitivity to drugs was assessed by a modified mitochondrial dye-conversion assay (Cell Titer 96, Promega #G4100, Madison, WI, USA) as described (Troester et al. 2004; Hoadley et al. 2007). Carboplatin and paclitaxel were purchased from UNC Hospital Pharmacy (Chapel Hill, NC, USA). ABT-888 was generously provided by the Center for Integrative Chemical Biology and Drug Discovery (CICBDD) and the University Cancer Research Fund at the University of North Carolina. 72h inhibitory concentrations that caused a 50% reduction in MTT (3-[4,5-dimethylthiazol-2-yl]-2,5-diphenyl tetrazolium bromide) dye conversion (IC50) were determined using nonlinear regression (Van Ewijk and Hoekstra 1993).

BRCA1 Foci Formation Assay

HME-CC cells were seeded into 4-chamber well slides at an initial concentration of 50,000 cells per well, allowed to adhere, then irradiated with 2.5 Gy of ionizing radiation with 20 minute recovery, or treated with 200 μM ABT-888 for 24h, or left untreated. After treatment, cells were fixed with 4% paraformaldehyde, permeabilized with 0.2% Triton X-100 solution, then immunostained with primary antibody (BRCA1, Santa Cruz, sc-6954) and Alexa-488 (Invitrogen) conjugated secondary antibody and counterstained with DAPI. Cells were also stained with geminin antibody (GMNN, ProteinTech, 10802-1-AP) and Alexa-568 (Invitrogen) conjugated secondary antibody. Images were taken with Leica SP2 confocal microscope with a 63X objective lens. Foci counts were

performed using a MetaMorph 7.0 (Molecular Devices) analysis module with foci being scored as areas of 2-5 pixel width with 70 gray levels above local background intensity. At least 5 fields of view were scored, per sample, per treatment category.

Survival Analysis

A mean expression value for *RAD17*+*RAD50*, and *INPP4B*, for each patient tumor in the UNC337 and NKI295 (van de Vijver et al. 2002) datasets was determined and patients were rank-ordered and separated into three equal groups representing low, medium, and high average expression for each gene or combined genes. Survival analyses were performed using the Kaplan-Meier test in R.

2.4 Acknowledgements

This work was supported by funds from the NCI Breast SPORE program (P50-CA58223), RO1-CA138255, T32-GM008719, F30-ES018038, R03-CA132143, P50-CA125183, the Breast Cancer Research Foundation, the EIF-Lee Jeans Translational Research Fund, and the V Foundation for Cancer Research. We thank the UNC CICBDD for ABT-888, which is directed by Stephen Frye, and compound provision is managed by Jian Jin.

2.5 Figures and Legends

Figure 2.1: Copy number frequency plots from SWITCHdna show regions of aberrations shared by members of the same subtype.

Grey shading indicates regions of change with the y-axis representing frequency of aberration at each site within each subtype. Regions in black were statistically associated with a particular subtype and remained significant after Benjamini-Hochberg correction. Regions below the center (negative values) represent losses, and areas above the center (positive values) indicate gains. A) Basal-like, B) Claudin-low, C) HER2-enriched, D) Luminal A, E) Luminal B, and F) Normal-like. G) Expanded view of the Basal-like copy number landscape. *INPP4B*, *MAP3K8*, *FAM107B*, and *ZEB1*, each in Basal-like specific regions of CNA, are marked. *BRCA1*, *BRCA2*, *PTEN*, *RB1*, and *TP53*, are genes/regions that were frequently, but not specifically, lost in the Basal-like subtype, and *KRAS*, which is frequently but not specifically gained in the Basal-like subtype, are also noted. The dotted horizontal lines indicate 50% gain or loss. H) Enlarged view of the Basal-like chromosome 5q region showing the location of *RAD17*, *MSH3*, *RAD50*, and *RAP80*. Loss frequency is indicated on the y-axis and the level of 50% loss is highlighted by the dotted line.

Figure 2.1

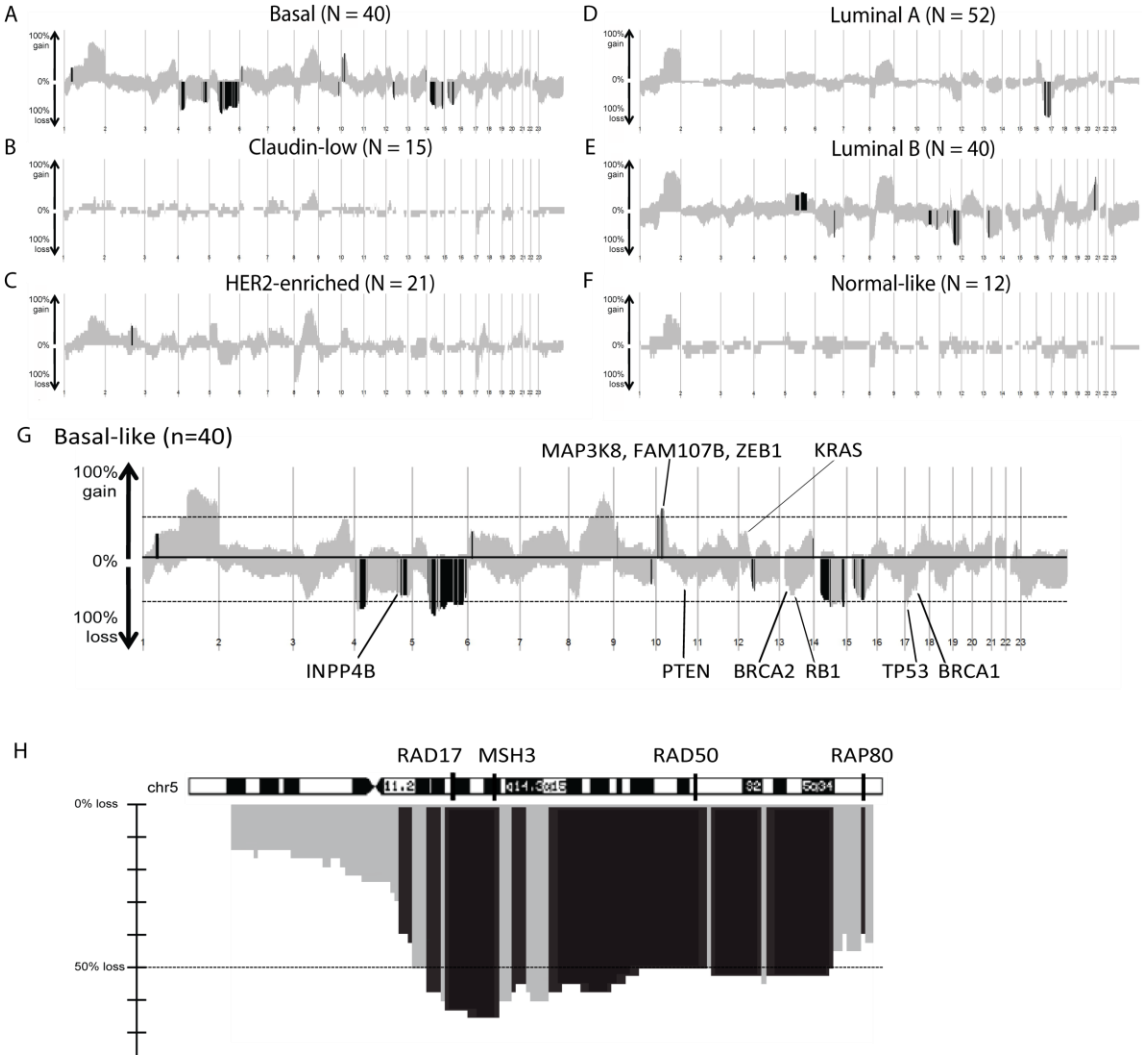

A
Basal (N = 40)
D
Luminal A (N = 52)
B
Claudin-low (N = 15)
E
Luminal B (N = 40)
C
HER2-enriched (N = 21)
F
Normal-like (N = 12)
100% gain
0%
100% loss
G
Basal-like (n=40)
MAP3K8, FAM107B, ZEB1
KRAS
INPP4B
PTEN
BRCA2
RB1
TP53
BRCA1
H
RAD17
MSH3
RAD50
RAP80
chr5
0% loss
50% loss

Figure 2.2: Numerous genes in the UNC-Norway dataset have low gene expression associated with DNA copy number loss.

Gene Expression values for *RAD17*, *RAD50*, *RAP80*, *MSH3*, *BRCA1*, *BRCA2*, *PTEN*, *RB1*, *TP53*, and *INPP4B* in the UNC-Norway dataset (n=180) separated by copy number status (DNA copy number loss vs. no loss). p-values determined by ANOVA test.

Figure 2.2

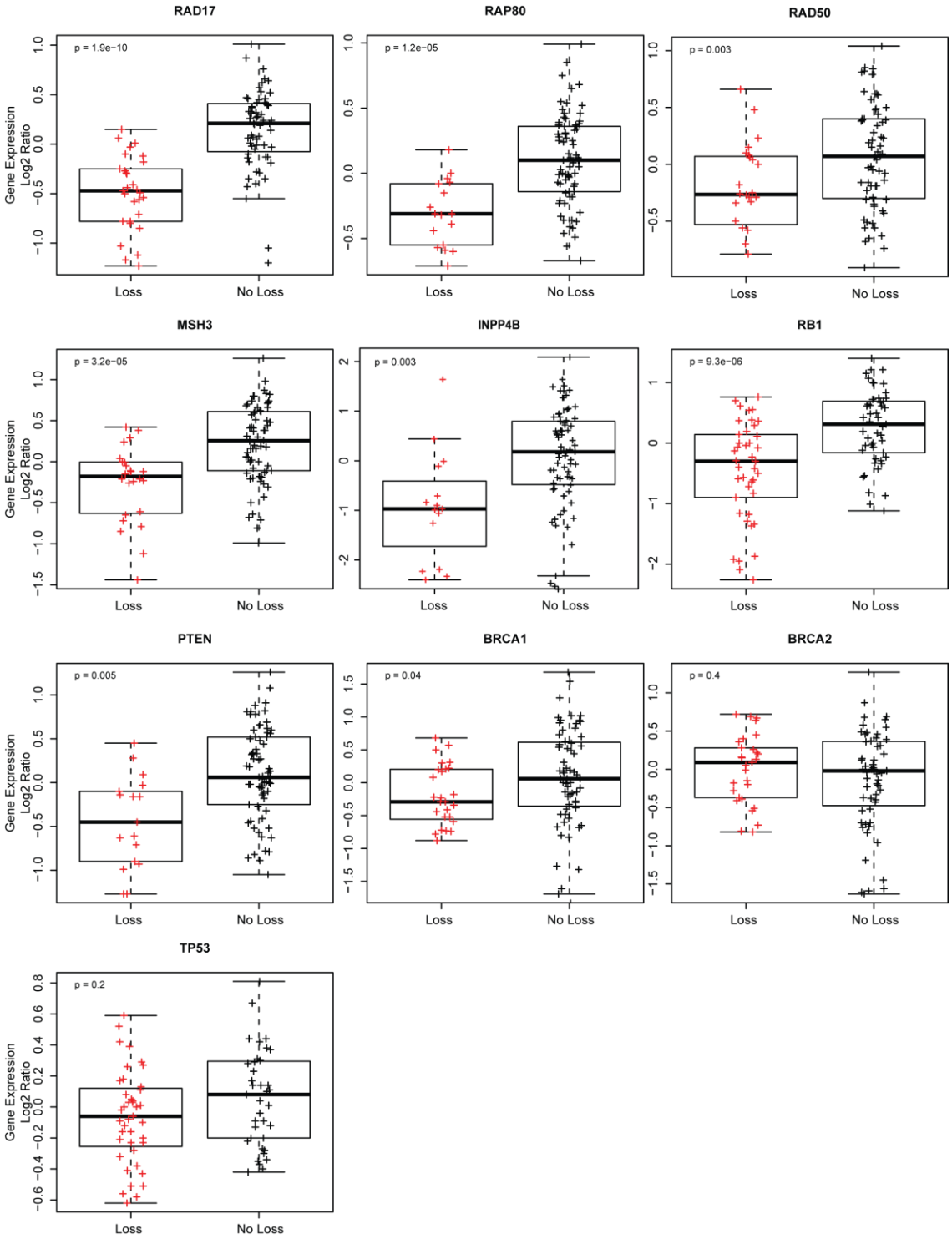

RAD17
p = 1.9e−10
RAP80
p = 1.2e−05
RAD50
p = 0.003
MSH3
p = 3.2e−05
INPP4B
p = 0.003
RB1
p = 9.3e−06
PTEN
p = 0.005
BRCA1
p = 0.04
BRCA2
p = 0.4
TP53
p = 0.2
Gene Expression
Log2 Ratio
Loss
No Loss

Figure 2.3: ANOVA boxplots for individual genes that are commonly lost in Basal-like cancers according to intrinsic subtype determined using the UNC337 sample set.

p-values were determined by 2-way ANOVA. A) *RAD17*, B) *RAD50*, C) *RAP80*, D) *PTEN*, E) *MSH3*, and F) *INPP4B*.

Figure 2.3

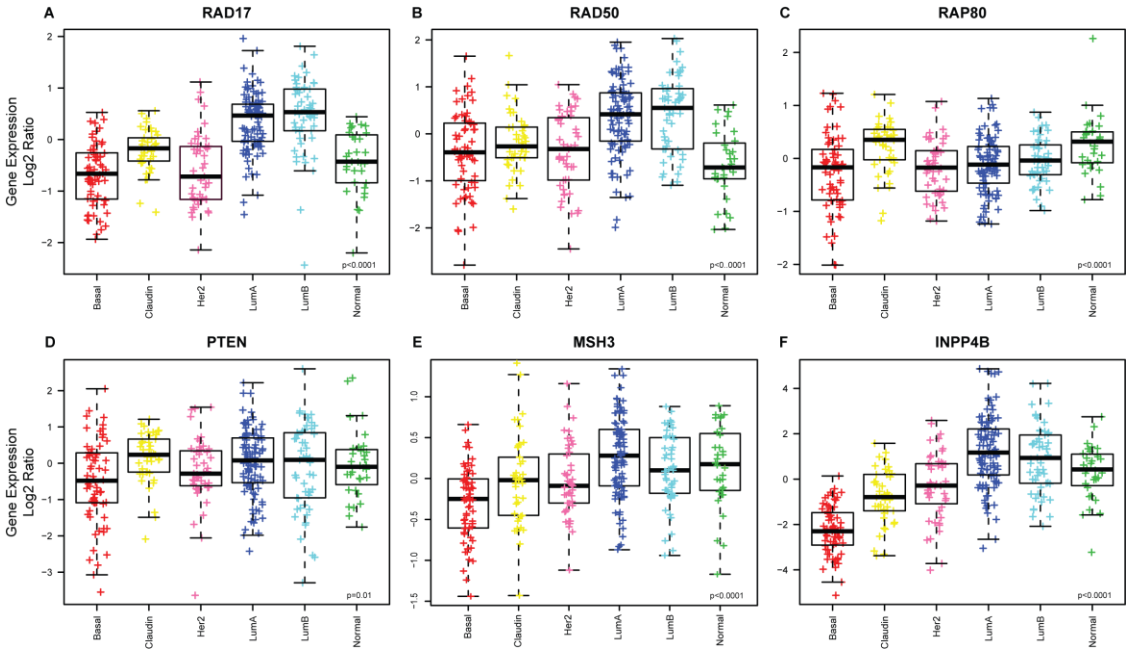

A
RAD17
B
RAD50
C
RAP80
D
PTEN
E
MSH3
F
INPP4B
Gene Expression
Log2 Ratio
Basal
Claudin
Her2
LumA
LumB
Normal
p<0.0001
p<0.0001
p<0.0001
p=0.01
p<0.0001
p<0.0001

Figure 2.4: Survival analysis according to expression of RAD17+RAD50 and INPP4B.

Patients in the UNC337 and NKI295 datasets were rank-ordered organized by average gene expression values of A) *RAD17+RAD50* combined, or B) *INPP4B*. The patients were split into 3rds based upon rank-order expression values and Kaplan Meier analysis was done on the three groups to examine trends in relapse-free survival and overall survival. p-values determined by log-rank test.

Figure 2.4

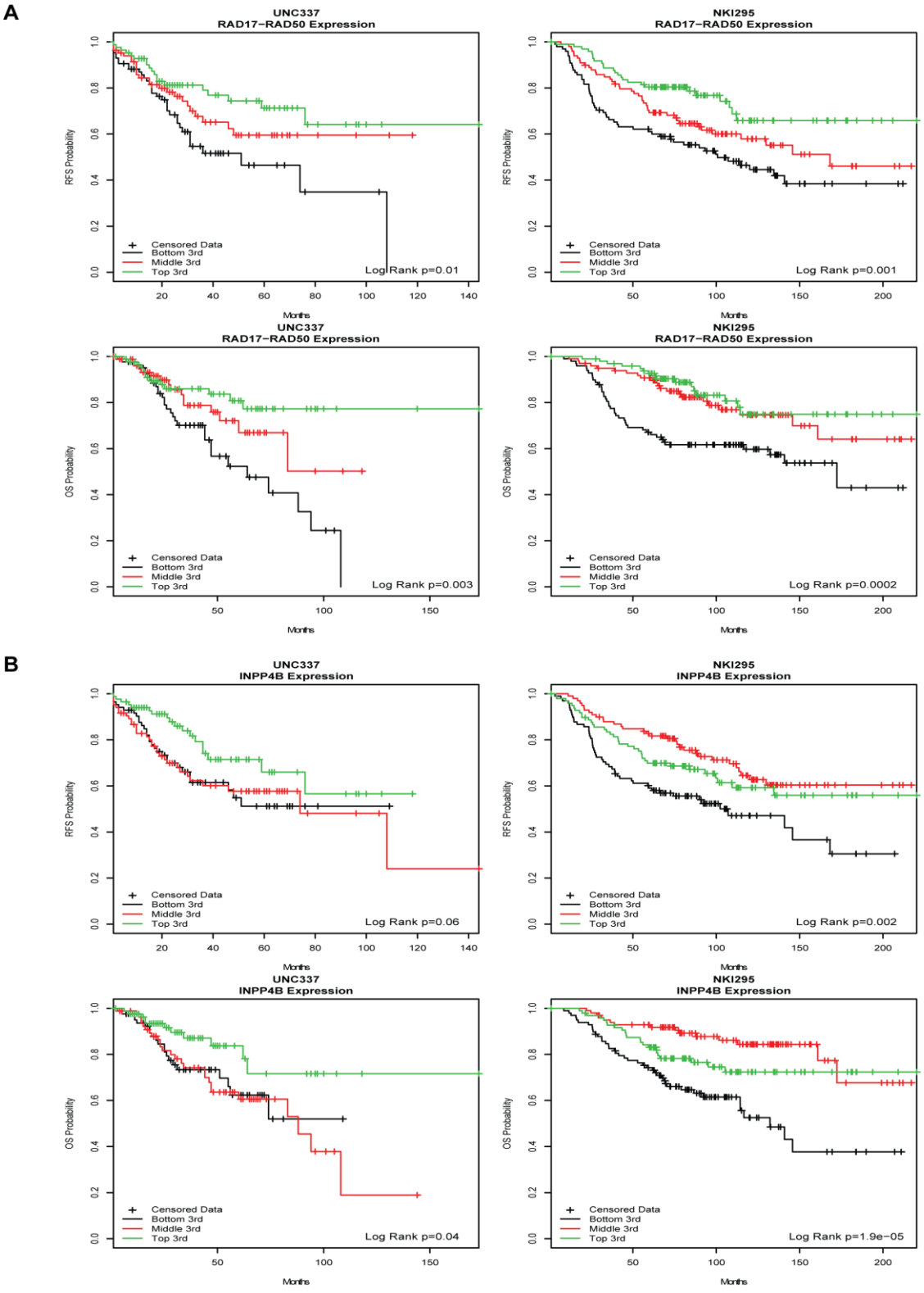
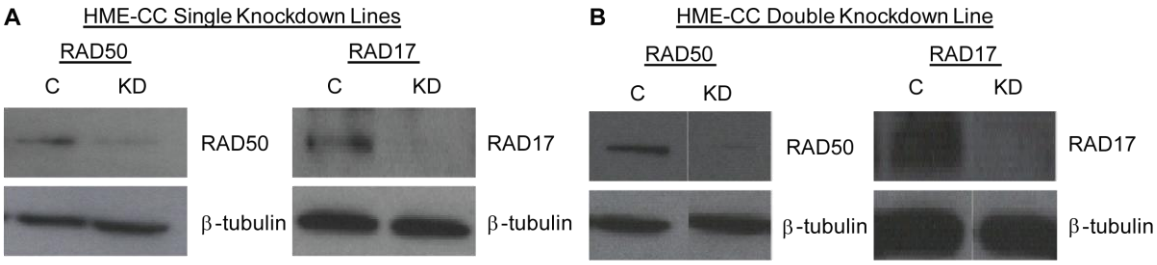

A
UNC337
RAD17−RAD50 Expression
RFS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=0.01
Months
NKI295
RAD17−RAD50 Expression
RFS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=0.001
Months
UNC337
RAD17−RAD50 Expression
OS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=0.003
Months
NKI295
RAD17−RAD50 Expression
OS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=0.0002
Months
B
UNC337
INPP4B Expression
RFS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=0.06
Months
NKI295
INPP4B Expression
RFS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=0.002
Months
UNC337
INPP4B Expression
OS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=0.04
Months
NKI295
INPP4B Expression
OS Probability
Censored Data
Bottom 3rd
Middle 3rd
Top 3rd
Log Rank p=1.9e−05
Months

Figure 2.5: RNAi knockdown experiments in an immortalized HMEC (BABE cell line).

Western blot analysis showing reduction of RAD17 and RAD50 protein expression in HME-CC A) single, or B) double RNAi knockdown lines. (KD, knockdown line; C, vector control line). Tubulin staining was performed as a loading control. C) Estimated IC50 with 95% CI for ABT-888, Carboplatin, and Paclitaxel based on mitochondrial dye-conversion assay. Results are based on the average of two experiments per condition, each done in triplicate, with knockdown-control pairs with significant differences in IC50 are designated with a *.

Figure 2.5



C

	ABT-888 IC50 (95% CI)	Carboplatin IC50 (95% CI)	Paclitaxel IC50 (95% CI)
RAD17 KD (n=6)	262.2 μM (248.3 - 276.0)*	29.6 μM (27.1 - 32.2)*	2.9 nM (2.6 - 3.2)
RAD17 Control	296.3 μM (283.5 - 309.1)	43.3 μM (38.5 - 48.2)	3.0 nM (2.8 - 3.3)
RAD50 KD (n=6)	205.7 μM (196.9 - 214.5)	25.9 μM (22.5 - 29.3)*	5.5 nM (3.8 - 7.2)
RAD50 Control	191.7 μM (176.7 - 206.7)	14.0 μM (11.9 - 16.1)	5.3 nM (3.4 - 7.2)
RAD17-RAD50 Double KD (n=6)	142.4 μM (128.3 - 156.4)*	17.3 μM (15.2 - 19.4)*	3.8 nM (3.5 - 4.1)
RAD17-RAD50 Double Control	244.2 μM (231.4 - 256.9)	31.7 μM (28.5 - 34.9)	4.0 nM (3.6 - 4.3)

* $p<0.001$

Figure 2.6: BRCA1-mediated DNA repair foci formation assay.

A) Representative images of BRCA1 foci formation in RAD17-RAD50 Double Knockdown cells and control cells after treatment with 2.5 Gy of ionizing irradiation and 20 minute recovery (Ionizing Radiation), or no treatment (Untreated). B) Representative images of BRCA1 foci formation in RAD17-RAD50 Double Knockdown cells and control cells with 200 μM ABT-888 (ABT-888), or no treatment (Untreated). Green channel - BRCA1, Red channel – Geminin, Blue Channel - DAPI images. All images were taken with a 63x objective and post processed to 300% of its original size. Automated BRCA1 foci counting results from each cell line for C) ionizing radiation and D) ABT-888 treatment. Error bars represent 95% confidence intervals (* $p<.05$ of knockdown relative to control). p-values were calculated from t-tests comparing foci counts in treated Double Knockdown cells versus treated Control cells or untreated Double Knockdown cells versus untreated Control cells.

Figure 2.6

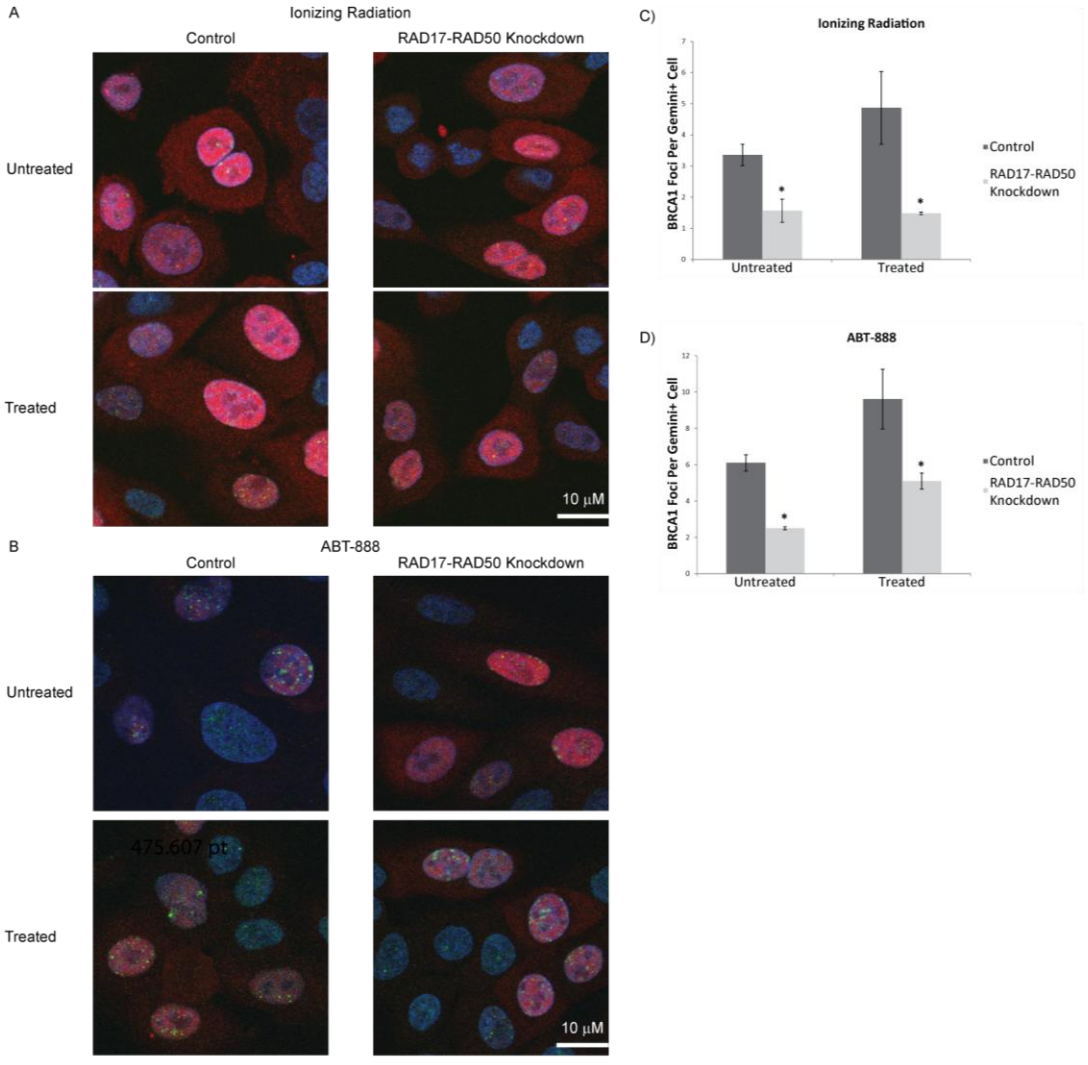

A
Ionizing Radiation
Control
RAD17-RAD50 Knockdown
Untreated
Treated
10 μM
B
ABT-888
Control
RAD17-RAD50 Knockdown
Untreated
Treated
475.607 pt
10 μM
C)
Ionizing Radiation
BRCA1 Foci Per Gemini+ Cell
Untreated
Treated
Control
RAD17-RAD50 Knockdown
D)
ABT-888
BRCA1 Foci Per Gemini+ Cell
Untreated
Treated
Control
RAD17-RAD50 Knockdown

Table 2.1: Frequency of copy number alterations data for the UNC-Norway combined dataset.

Data is shown for selected A) deletions, B) amplifications, C) average number of changes, D) % Tumor Cellularity, and E) co-occurrences. Values are presented in 'Count (%)' format. Specific counts are given for individual deletions or co-deletions, with each sample only classified into one category. C) Counts for average gains/losses for each subtype. Total number of aberrations is the sum of all individual gene gains and losses. Average segment number and length were calculated from the SWITCHdna generated segments for each sample within each subtype. D) % Tumor Cellularity generated by ASCAT algorithm or genoCNA algorithm. E) Rates of co-occurrence of 5q cluster loss with other gene alterations are shown (N refers to the number of total samples with 5q loss). Fisher's Exact tests or Chi-square approximations were done to determine if the rates of occurrence, or co-occurrence, were at statistically significant levels.

Table 2.1

	UNC-NW													All ($n = 180$)	
	Basal ($n = 40$)		Claudin ($n = 15$)		Her2 ($n = 21$)		LumA ($n = 52$)		LumB ($n = 40$)		Normal-like ($n = 12$)		*P*-Value[a]	No.	%
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%			
(a) Deletions															
No 5q Genes Lost	13	32.5	11	73.3	11	52.4	42	80.8	32	80.0	10	83.3	<0.001	119	66.1
5q13.2 (RAD17)	5	12.5	1	6.7	1	4.8	4	7.7	2	5.0	1	8.3	<0.001	14	7.8
5q31.1 (RAD50)	1	2.5	0	0.0	1	4.8	0	0.0	2	5.0	0	0.0	<0.001	4	2.2
5q35.2 (RAP80)	1	2.5	1	6.7	0	0.0	2	3.8	2	5.0	0	0.0	0.001	6	3.3
RAD17/RAD50	5	12.5	2	13.3	3	14.3	2	3.8	0	0.0	0	0.0	<0.001	12	6.7
RAD17/RAP80	1	2.5	0	0.0	0	0.0	0	0.0	1	2.5	1	8.3	<0.001	3	1.7
RAD50/RAP80	0	0.0	0	0.0	0	0.0	1	1.9	1	2.5	0	0.0	<0.001	2	1.1
RAD17/RAD50/RAP80	14	35.0	0	0.0	5	23.8	1	1.9	0	0.0	0	0.0	<0.001	20	11.1
10q23.31 (PTEN)	9	22.5	1	6.7	5	23.8	12	23.1	12	30.0	2	16.7	0.6	41	22.8
13q14.2 (RB1)	18	45.0	3	20.0	7	33.3	11	21.1	33	82.5	4	33.3	<0.001	76	42.2
17p13.1 (TP53)	20	50.0	6	40.0	12	57.1	18	34.6	20	50.0	4	33.3	0.4	80	44.4
17q21 (BRCA1)	17	42.5	2	12.5	11	52.4	11	21.2	7	17.5	3	25.0	0.01	51	28.3
13q12.1 (BRCA2)	12	30.0	3	18.8	5	23.8	8	15.4	21	52.5	3	25.0	0.01	52	28.9
4q31.21 (INPP4B)	16	40.0	1	11.1	1	4.8	6	11.5	4	10.0	1	8.3	0.002	29	16.1
(b) Amplifications															
17q12 (ERBB2)	9	22.5	2	13.3	12	57.1	9	17.3	15	37.5	2	16.7	0.007	49	27.2
12p12.1 (KRAS)	14	35.0	4	26.7	5	23.8	9	17.3	9	22.5	2	16.7	0.5	43	23.9
12q15 (MDM2)	6	15.0	1	6.7	4	19.0	12	23.1	18	45.0	1	8.3	0.01	42	23.3
8q24.21 (MYC)	26	65.0	7	46.7	10	47.6	18	34.6	29	72.5	3	25.0	0.002	93	51.7
10p11.23 (MAP3K8)	16	40.0	1	6.7	6	28.6	3	5.8	3	7.5	2	16.7	<0.001	31	17.2
10p11.22 (ZEB1)	16	40.0	1	6.7	5	23.8	3	5.8	3	7.5	2	16.7	<0.001	30	16.7
10p13 (FAM107B)	20	50.0	1	6.7	5	23.8	4	7.7	6	15.0	1	8.3	<0.001	37	20.6
(c)															
Average # of Gains	3943		1543		2970		2847		3885		2326			3192	
Average # of Losses	4854		1906		3347		2560		4634		2891			3590	
Total # of Aberrations	8797		3450		6317		5408		8519		5218			6782	
Average # of Segments	194		143		223		150		222		139			183	
Segment Length (kb	14988		20421		13068		19424		13113		21062			15923	
(d)															
% Tumor Cellularity (ASCAT)	52.5		38.4		38.2		53.0		50.2		38.1			48.4	
% Tumor Purity (genoCNA)	75.0		78.0		79.5		68.0		68.0		70.0			71.0	

(e) Co-occurrence of 5q[b] loss with additional gene loss

	N	No.	%	*P*-Value[c]
10q23.31 (PTEN) Loss	32	14	43.8	<0.001
13q14.2 (RB1) Loss	32	26	81.3	<0.001
17q21 (BRCA1) Loss	32	17	53.1	<0.001
10p Amplicon	32	15	46.9	<0.001
17p13.1 (TP53) Loss	32	19	59.4	0.08
4q31.21 (INPP4B) Loss	32	16	50.0	<0.001

Values are presented in 'Count (%)' format. Specific counts are given for individual deletions or co-deletions, with each sample only classified into one category. c Counts for average gains/losses for each subtype. Total number of aberrations is the sum of all individual gene gains and losses. Average segment number and length were calculated from the SWITCHdna generated segments for each sample within each subtype. d % Tumor Cellularity generated by ASCAT algorithm or genoCNA algorithm. e Rates of co-occurrence of 5q cluster loss with other gene alterations are shown (*N* refers to the number of total samples with 5q loss). Fisher's exact tests or Chi-square approximations were done to determine if the rates of occurrence, or co-occurrence, were at statistically significant levels

[a] Chi-square approximation

[b] RAD17+RAD50 loss OR RAD17+RAD50+RAP80 loss

[c] Fisher's exact test

Table 2.2: Frequency of copy number alterations data for the Jonsson dataset.

Data is shown for selected A) deletions, B) amplifications, C) average number of changes, and D) co-occurrences. Values are presented in 'Count (%)' format. Specific counts are given for individual deletions or co-deletions, with each sample only classified into one category. C) Counts for average gains/losses for each subtype. Total number of aberrations is the sum of all individual gene gains and losses. Average segment number and length were calculated from the SWITCHdna generated segments for each sample within each subtype. D) Rates of co-occurrence of 5q cluster loss with other gene alterations are shown (N refers to the number of total samples with 5q loss). Fisher's Exact tests or Chi-square approximations were done to determine if the rates of occurrence, or co-occurrence, were at statistically significant levels.

Table 2.2

	Jonsson													All ($n = 356$)	
	Basal ($n = 61$)		Claudin ($n = 43$)		Her2 ($n = 46$)		LumA ($n = 117$)		LumB ($n = 55$)		Normal-like ($n = 34$)		*P*-Value[a]	No.	%
	No.	%	No.	%	No.	%	No.	%	No.	%	No.	%			
(a) Deletions															
No 5q Genes Lost	15	24.6	29	67.4	28	58.3	107	90.7	44	80.0	26	76.5	<0.001	249	69.9
5q13.2 (RAD17)	9	14.8	4	9.3	3	6.3	3	2.5	5	9.1	3	8.8	<0.001	27	7.6
5q31.1 (RAD50)	4	6.6	0	0.0	2	4.2	3	2.5	3	5.5	1	2.9	<0.001	13	3.7
5q35.2 (RAP80)	0	0.0	0	0.0	0	0.0	2	1.7	0	0.0	1	2.9	<0.001	3	0.8
RAD17/RAD50	16	26.2	2	4.7	12	25.0	2	1.7	2	3.6	2	5.9	<0.001	36	10.1
RAD17/RAP80	3	4.9	1	2.3	0	0.0	0	0.0	0	0.0	0	0.0	<0.001	4	1.1
RAD50/RAP80	2	3.3	0	0.0	1	2.1	0	0.0	1	1.8	0	0.0	<0.001	4	1.1
RAD17/RAD50/RAP80	12	19.7	7	16.3	2	4.2	1	0.8	0	0.0	1	2.9	<0.001	23	6.5
10q23.31 (PTEN)	21	34.4	10	23.3	9	18.8	14	11.9	17	30.9	2	5.9	<0.001	73	20.5
13q14.2 (RB1)	33	54.1	16	37.2	13	27.1	32	27.1	31	56.4	8	23.5	<0.001	133	37.4
17p13.1 (TP53)	21	34.4	6	14.0	17	35.4	33	28,0	19	34.5	11	32.4	0.2	107	30.1
17q21 (BRCA1)	19	31.1	6	14.0	7	15.2	12	10.3	5	9.1	5	14.7	<0.001	54	15.2
13q12.1 (BRCA2)	22	36.1	13	30.2	12	26.1	26	22.2	30	54.5	7	20.6	<0.001	110	30.9
4q31.21 (INPP4B)	29	47.5	14	32.6	16	34.8	13	11.1	11	20.0	2	5.9	<0.001	85	23.9
(b) Amplifications															
17q12 (ERBB2)	9	14.8	5	11.6	31	64.6	23	19.5	17	30.9	6	17.6	<0.001	91	25.6
12p12.1 (KRAS)	12	19.7	7	16.3	3	6.5	3	2.5	10	18.2	0	0.0	<0.001	35	9.8
12q15 (MDM2)	2	3.3	4	9.3	7	15.2	13	11.1	17	30.9	0	0.0	<0.001	43	12.1
8q24.21 (MYC)	42	68.9	19	44.2	22	47.8	48	41.0	44	80.0	14	41.2	<0.001	189	53.1
10p11.23 (MAP3K8)	16	26.2	8	18.6	7	14.6	6	5.1	8	14.5	1	2.9	0.001	46	12.9
10p11.22 (ZEB1)	16	26.2	8	18.6	8	16.7	6	5.1	8	14.5	1	2.9	<0.001	47	13.2
10p13 (FAM107B)	29	47.5	10	23.3	6	12.5	6	5.1	11	20.0	3	8.8	<0.001	65	18.3
c)															
Average # of Gains	2853		2169		2481		1963		3281		1469			4675	
Average # of Losses	5089		3171		3523		2430		3872		2395			6597	
Total # of Aberrations	7942		5341		6003		4393		7153		3864			11272	
Average # of Segments	167		130		134		97		129		93			122	
Segment Length (kb)	16522		21289		20573		28542		21462		29841			22610	
(d) Co-occurrence of 5q[b] loss with additional gene loss															
	N		No.	%	*P*-Value[c]										
10q23.31 (PTEN) Loss	67		28	41.8	<0.001										
13q14.2 (RB1) Loss	67		39	58.2	<0.001										
17q21 (BRCA1) Loss	67		24	35.8	<0.001										
10p Amplicon	67		15	22.4	0.02										
17p13.1 (TP53) Loss	67		29	43.3	0.01										
4q31.21 (INPP4B) Loss	67		31	46.3	<0.001										

Values are presented in Count (%) format. Specific counts are given for individual deletions or co-deletions, with each sample only classified into one category. c Counts for average gains/losses for each subtype. Total number of aberrations is the sum of all individual gene gains and losses. Average segment number and length were calculated from the SWITCHdna generated segments for each sample within each subtype. d Rates of co-occurrence of 5q cluster loss with other gene alterations are shown (*N* refers to the number of total samples with 5q loss). Fisher's exact tests or Chi-square approximations were done to determine if the rates of occurrence, or co-occurrence, were at statistically significant levels

[a] Chi-square approximation

[b] RAD17+RAD50 loss OR RAD17+RAD50+RAP80 Loss

[c] Fisher's exact test

Table 2.3: Comparison of Jonsson et al. copy number based classifications versus intrinsic subtypes.

Subtype classifications using both the original labels in the Jonsson dataset (Jonsson et al.) using copy number defined subtypes, versus PAM50 plus Claudin-low gene expression subtypes is shown. p-value determined by Chi-square approximation.

Table 2.3

		Jonsson subtypes					
		Amplifier	Luminal-complex	Mixed	17q12	Luminal-simple	Basal-complex
PAM50+Claudin low subtypes	Basal	7	4	4	3	1	42
	LumA	18	42	15	10	31	2
	Claudin	11	4	8	4	3	13
	LumB	11	36	3	2	1	2
	Her2	1	9	3	30	0	5
	Normal-like	4	10	5	2	10	3

Subtype classifications using both the original labels in the Jonsson dataset [14] using copy number defined subtypes, versus PAM50 plus Claudin-low gene expression subtypes is shown. *P* value determined by Chi-square approximation

P-value < 0.001

Table 2.4: Examination of possible correlations between the specific CNA and overall genomic instability.

The average numbers of CNA for gains, losses, or both, are shown for the entire dataset and within sets of tumors with a given copy number alteration (5q, *PTEN*/10q23.31, *RB1*/13q14.2, *TP53*/17p13.1 and 10p). A Wilcoxon-rank sum test was performed to see if the rate of copy number aberration between each group (Pairwise: Aberration vs. Other, or No Aberration) was significantly different (*).

Table 2.4

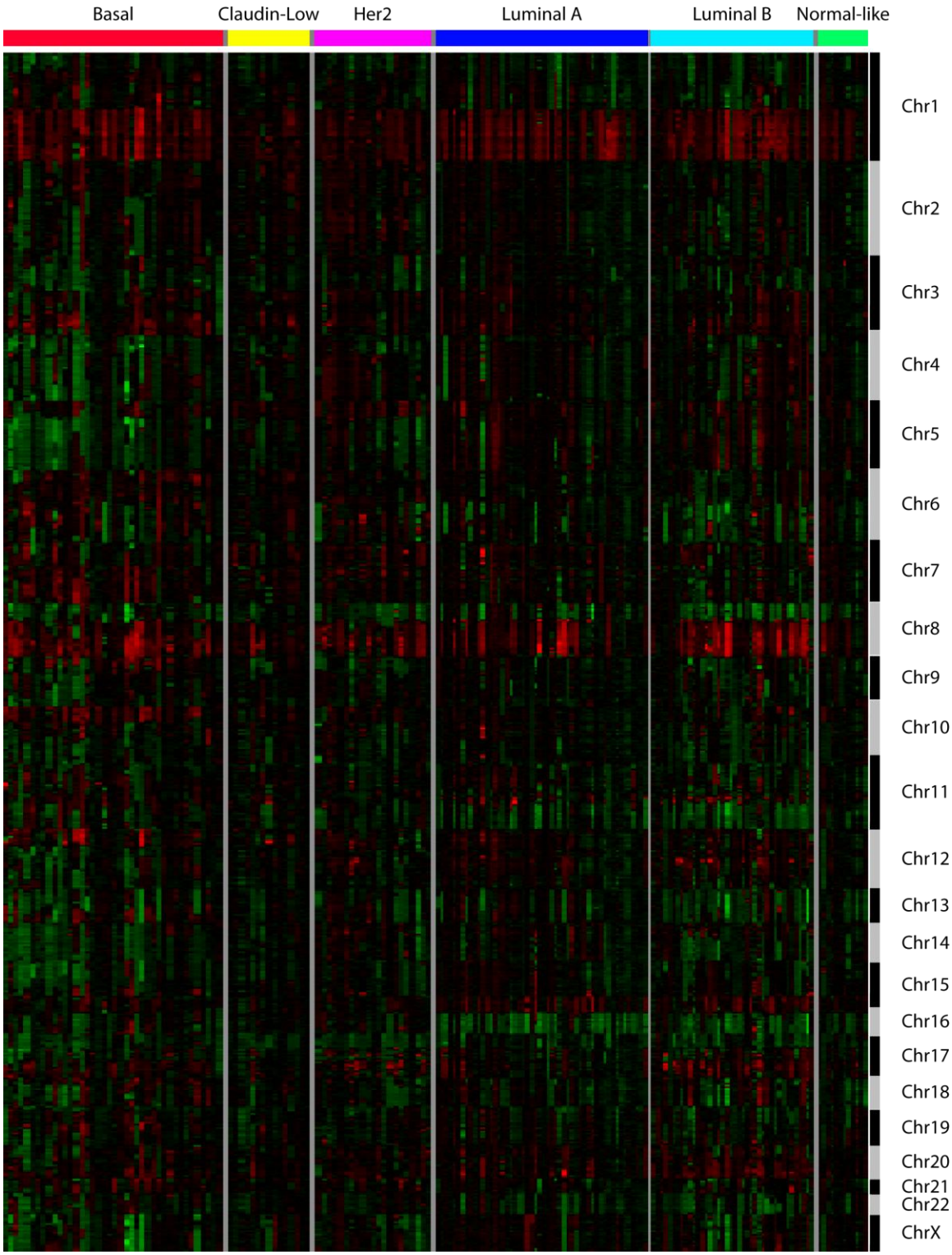
Average amount of CNA by event and class in combined UNC/Norway dataset	Average gain	Average loss	Average total
All samples mean ($n = 180$)	3,192	3,590	6,783
All samples median ($n = 180$)	2,790	3,230	6,560
RAD17 + RAD50 ($n = 12$)	3,948 ($P = 0.1$)	5,614 ($P = 0.0004$*)	9,562 ($P = 0.003$*)
RAD17 + RAD50 +/− RAP80 ($n = 32$)	4,451 ($P = 0.0002$*)	6,681 ($P < 0.00001$*)	11,131 ($P < 0.00001$*)
RAD17 + RAD50 + RAP80 ($n = 20$)	4,752 ($P = 0.0002$*)	7,320 ($P < 0.00001$*)	12,073 ($P < 0.00001$*)
Other ($n = 148$)	2,920	2,922	5,842
10q23.31 (PTEN) loss ($n = 31$)	4,202 ($P = 0.009$*)	6,169 ($P < 0.00001$*)	10,371 ($P < 0.00001$*)
No 10q23.31 (PTEN) loss ($n = 149$)	2,982	3,054	6,036
13q14.2 (RB1) loss ($n = 66$)	4,127 ($P = 0.00002$*)	5,638 ($P < 0.00001$*)	9,765 ($P < 0.00001$*)
No 13q14.2 (RB1) loss ($n = 114$)	2,652	2,404	5,056
17p13.1 (TP53) loss ($n = 80$)	3,900 ($P = 0.00008$*)	4,8578 ($P < .00001$*)	8,757 ($P < 0.00001$*)
No 17p13.1 (TP53) loss ($n = 100$)	2,627	2,576	5,203
10p Amplicon ($n = 34$)	5,016 ($P < 0.00001$*)	5,232 ($P = 0.0002$*)	10,248 ($P < 0.00001$*)
No 10p Amplicon ($n = 146$)	2,768	3,208	5,975

The average numbers of CNAs for gains, losses, or both, are shown for the entire dataset and within sets of tumors with a given copy number alteration (5q, *PTEN*/10q23.31, *RB1*/13q14.2, *TP53*/17p13.1, and 10p). A Wilcoxon-rank sum test was performed to see if the rate of copy number aberration between each group (Pairwise: Aberration vs. Other, or No Aberration) was significantly different (*)

Supplementary Figure 2.1: Heatmap of copy number aberration landscape of the UNC-NW tumor training sample set.

Samples were ordered by sample name within subtype and the data is presented in chromosomal order. Red indicates increased copy number, black unchanged relative to normal, and green indicates decreased copy number.

Supplementary Figure 2.1

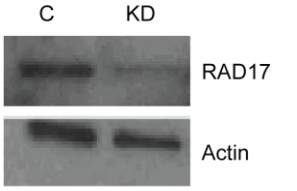


Supplementary Figure 2.2: RNAi knockdown experiments in a second immortalized HMEC (ME16C).

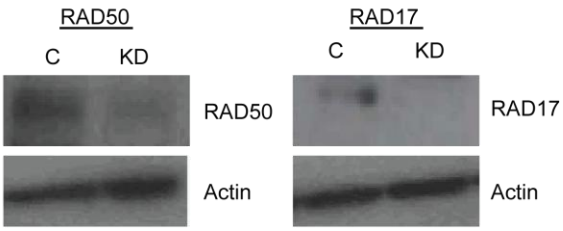
Western blot analysis showing reduction of RAD17 and RAD50 protein expression in A) single or B) double RNAi knockdown lines. (KD, knockdown line; C, vector control line). Actin staining was performed as a loading control. C) Estimated IC50 with 95% CI for ABT-888, Carboplatin, and Paclitaxel based on mitochondrial dye-conversion assay. Results for each condition are based on the average of an experiment done in triplicate, and knockdown-control pairs with significant differences in IC50 are designated with a *.

Supplementary Figure 2.2

A) ME16C Single Knockdown Line



B) ME16C Double Knockdown Line



C)

	ABT-888 IC50 (95% CI)	Carboplatin IC50 (95% CI)	Paclitaxel IC50 (95% CI)
RAD17 KD (n=3)	233.8 μM (218.4 - 249.2)	57.6 μM (56.0 - 59.2)*	.04 nM (.04 - .05)
RAD17 Control	213.1 μM (195.1 - 231.1)	102.1μM (97.2 - 107.0)	.07 nM (.05 - .08)
RAD17-RAD50 Double KD (n=3)	277.7 μM (271.0 - 283.4)*	52.0 μM (49.9 - 54.1)*	3.6 nM (3.0 - 4.2)
RAD17-RAD50 Double Control	330.2 μM (324.6 - 335.8)	136.0μM (126.4-145.6)	3.0 nM (2.6 - 3.4)

* p<0.001

Supplementary Table 2.1: List of the subtype-specific regions of aberration after multiple hypothesis correction, organized by subtype.

Supplementary Table 2.1

Chromosome	Start Position	Stop Position	Cytoband	Gain/Loss
Basal				
1	41717036	43943775	1p34.2	GAIN
1	43945805	44960161	1p34.1	GAIN
1	46541957	46555032	1p33	GAIN
4	1234177	1683843	4p16.3	LOSS
4	3220565	5553626	4p16.2	LOSS
4	5577784	10295484	4p16.1	LOSS
4	10979549	15266111	4p15.33	LOSS
4	15313739	17632477	4p15.32	LOSS
4	19864333	22430289	4p15.31	LOSS
4	23402742	26636101	4p15.2	LOSS
4	30331135	30753569	4p15.1	LOSS
4	37121738	38804810	4p14	LOSS
4	41057559	42353879	4p13	LOSS
4	69085500	69570979	4q13.2	LOSS
4	71492582	75938920	4q13.3	LOSS
4	76623381	79684447	4q21.1	LOSS
4	79691766	80079606	4q21.21	LOSS
4	114219937	114355024	4q26	LOSS
4	143168636	143603331	4q31.21	LOSS
4	151218876	154929678	4q31.3	LOSS
4	155375138	160500747	4q32.1	LOSS
4	162524501	164492534	4q32.2	LOSS
4	164612456	170167997	4q32.3	LOSS
4	170251982	170915637	4q33	LOSS
4	174326479	176135905	4q34.1	LOSS
4	176791083	177486490	4q34.2	LOSS
5	49730236	50173926	5q11.1	LOSS
5	52119531	58183162	5q11.2	LOSS
5	58300629	62108926	5q12.1	LOSS
5	63292034	63701397	5q12.2	LOSS
5	63838208	66501179	5q12.3	LOSS
5	66513872	67633403	5q13.1	LOSS
5	68425638	72913540	5q13.2	LOSS
5	73958991	76396789	5q13.3	LOSS
5	76408288	81082828	5q14.1	LOSS
5	81303630	82685333	5q14.2	LOSS
5	82803339	90714877	5q14.3	LOSS

Chromosome	Start Position	Stop Position	Cytoband	Gain/Loss
5	92250850	96544700	5q15	LOSS
5	98132900	102642260	5q21.1	LOSS
5	102912456	102926389	5q21.2	LOSS
5	106744250	109231328	5q21.3	LOSS
5	109877429	111137815	5q22.1	LOSS
5	111507433	112958878	5q22.2	LOSS
5	113725565	115180304	5q22.3	LOSS
5	115193714	121517156	5q23.1	LOSS
5	121675719	127022221	5q23.2	LOSS
5	127447382	129549383	5q23.3	LOSS
5	130522776	135427403	5q31.1	LOSS
5	135493104	138988200	5q31.2	LOSS
5	139008130	142795270	5q31.3	LOSS
5	143171919	147191453	5q32	LOSS
5	147238467	151765017	5q33.1	LOSS
5	152850499	154326721	5q33.2	LOSS
5	155686345	159845035	5q33.3	LOSS
5	159922707	167623739	5q34	LOSS
5	167651670	171366482	5q35.1	LOSS
5	172591744	174888201	5q35.2	LOSS
5	176660805	180009206	5q35.3	LOSS
6	13729714	14088212	6p23	GAIN
6	42772314	43039595	6p21.1	GAIN
9	831690	1047552	9p24.3	GAIN
9	12683435	13935586	9p23	GAIN
9	14071847	14900353	9p22.3	GAIN
9	122980237	124410900	9q33.2	LOSS
10	170643	1769670	10p15.3	GAIN
10	3099740	3205003	10p15.2	GAIN
10	3811260	6662244	10p15.1	GAIN
10	7244255	12332593	10p14	GAIN
10	12431738	17283687	10p13	GAIN
10	17311250	19896508	10p12.33	GAIN
10	20145174	20609292	10p12.32	GAIN
10	21110098	22746531	10p12.31	GAIN
10	22863772	24876778	10p12.2	GAIN
10	24604620	28074753	10p12.1	GAIN
10	28141103	31360860	10p11.23	GAIN
10	31648148	32903498	10p11.22	GAIN
12	42034279	44588086	12q12	LOSS

Chromosome	Start Position	Stop Position	Cytoband	Gain/Loss
12	44599181	47397048	12q13.11	LOSS
12	47446248	48387464	12q13.12	LOSS
12	48421630	53099317	12q13.13	LOSS
12	53136012	55168448	12q13.2	LOSS
12	55201993	55269875	12q13.3	LOSS
12	56621627	57600529	12q14.1	LOSS
13	104916365	104941384	13q33.2	GAIN
13	106620319	107317084	13q33.3	GAIN
13	109973420	114110898	13q34	GAIN
14	21961377	23608756	14q11.2	LOSS
14	23609961	31698685	14q12	LOSS
14	31868274	34078694	14q13.1	LOSS
14	34100051	35410919	14q13.2	LOSS
14	35837522	37090214	14q13.3	LOSS
14	37128947	38971371	14q21.1	LOSS
14	49113809	52311409	14q22.1	LOSS
14	52311662	54325595	14q22.2	LOSS
14	54378476	55837783	14q22.3	LOSS
14	57736586	58085299	14q23.1	LOSS
14	69312342	69568836	14q24.2	LOSS
14	92049878	92765306	14q32.12	LOSS
14	92773649	95250286	14q32.13	LOSS
14	95740950	100105884	14q32.2	LOSS
15	28193434	29022600	15q13.2	LOSS
15	29020705	31945591	15q13.3	LOSS
15	31945721	37862331	15q14	LOSS
15	37880223	38115089	15q15.1	LOSS
15	49136093	51017946	15q21.2	LOSS
15	51593230	55368004	15q21.3	LOSS
15	56490060	57176541	15q22.1	LOSS
15	57184612	61461128	15q22.2	LOSS
15	61583863	61913200	15q22.31	LOSS
15	65899096	66370691	15q23	LOSS
15	75011134	75564998	15q24.3	LOSS
19	39664720	39962225	19q13.11	GAIN
22	37740208	38258806	22q13.1	GAIN
Her2				
2	159800558	163403274	2q24.2	GAIN
2	164158152	168812365	2q24.3	GAIN

Chromosome	Start Position	Stop Position	Cytoband	Gain/Loss
Luminal A				
16	45393460	45522699	16q11.2	LOSS
16	45546775	51138307	16q12.1	LOSS
16	51646446	54466783	16q12.2	LOSS
16	54782803	56638303	16q13	LOSS
16	56705000	65205296	16q21	LOSS
16	65206155	69392572	16q22.1	LOSS
16	69398793	73198518	16q22.3	LOSS
16	73739922	78192112	16q23.1	LOSS
16	79132355	80549399	16q23.2	LOSS
16	80588753	82633263	16q23.3	LOSS
16	82644872	85172804	16q24.1	LOSS
16	85920445	87225756	16q24.2	LOSS
16	87232502	87749670	16q24.3	LOSS
Luminal B				
5	58300629	59320301	5q12.1	GAIN
5	68425638	71536822	5q13.2	GAIN
5	101597589	102642260	5q21.1	GAIN
5	102912456	102926389	5q21.2	GAIN
5	125787000	126200608	5q23.2	GAIN
5	128458341	128477620	5q23.3	GAIN
6	116369389	117096650	6q22.1	LOSS
6	117109043	117359993	6q22.2	LOSS
10	118946990	119027085	10q25.3	LOSS
10	119032598	121691235	10q26.11	LOSS
10	122206456	122883255	10q26.12	LOSS
10	123223889	124739898	10q26.13	LOSS
11	49383103	49393881	11p11.12	LOSS
11	76710709	85146692	11q14.1	LOSS
11	85633463	87710586	11q14.2	LOSS
11	87881006	92269284	11q14.3	LOSS
11	92342437	95765250	11q21	LOSS
11	104262627	109547776	11q22.3	LOSS
11	110631916	112776199	11q23.1	LOSS
11	112785528	114880322	11q23.2	LOSS
11	116803699	119605859	11q23.3	LOSS
13	27611014	27767472	13q12.2	LOSS
13	27774389	30804409	13q12.3	LOSS
13	31211674	31275009	13q13.1	LOSS

Chromosome	Start Position	Stop Position	Cytoband	Gain/Loss
20	40134806	41603948	20q13.11	GAIN
20	41621022	45719023	20q13.12	GAIN
20	45719063	49181611	20q13.13	GAIN
20	49438571	52269898	20q13.2	GAIN
20	61689399	61755226	20q13.33	GAIN

CHAPTER THREE

MICRO-SCALE GENOMIC COPY NUMBER ABERRATIONS AS ANOTHER MEANS OF MUTAGENESIS IN BREAST CANCER

3.1 Introduction

In breast cancers, the Basal-like subtype has high levels of genomic instability relative to other breast cancer subtypes with many basal-like-specific regions of aberration. There is evidence that this genomic instability extends to smaller scale genomic aberrations as well, as shown by a previously described micro-event in the *PTEN* gene in the Basal-like SUM149 breast cancer cell line. We sought to identify if small regions of genomic change exist, by using a high density, gene centric comparative genomic hybridization (CGH) array on both cell lines and primary tumors. A custom Agilent tiling array for CGH (244,000 probes, 200bp tiling resolution) was created to identify small regions of genomic change and was focused on previously identified basal-like specific, and general, cancer genes. Tumor genomic DNA from 94 patients and 2 breast cancer cell lines was labeled and hybridized to these arrays. Aberrations were called using SWITCHdna and the smallest 25% of SWITCHdna-defined genomic segments being called micro-aberrations (<64 contiguous probes, ~ <15kb). Our data showed that primary tumor breast cancer genomes frequently contained areas of small-scale copy number gains and losses, termed micro-aberrations, which are undetectable using lower-density genome-wide platforms. The basal-like subtype exhibited the highest

incidence of these events. These micro-aberrations sometimes altered expression of the involved gene as suggested by data from microarray and mRNA-seq studies. We confirmed the presence of the *PTEN* micro-amplification in SUM149 and by mRNA-seq showed that this resulted in loss of expression of all exons downstream of this event. Micro-aberrations disproportionately affected the 5' regions of the affected genes, including the promoter region, and a high frequency of micro-aberrations was associated with poor survival outcomes. Using a high probe density, gene-centric aCGH microarray, we present evidence of small-scale genomic aberrations that contribute to gene inactivation, and thus, genomic instability and tumor formation through a mechanism not detected using conventional copy number analyses.

A hallmark of many human cancers is genomic instability, and cancer itself can be thought of as the result of altered ploidy (Shen). The importance of studying the cancer genome cannot be understated as genomic alterations have been linked to cancer causation both broadly and in specific subgroups of patients (Rowley 1973; Seshadri et al. 1989; Seshadri et al. 1993; Sakakura et al. 1999; Pinkel and Albertson 2005). Alterations on a genomic level are likely to cause associated changes in gene expression (Tran et al. ; Jarvinen et al. 2008). Previous global gene expression profiling studies of breast carcinoma samples have identified at least five distinct subtypes of breast cancer (Perou et al. 1999; Perou et al. 2000; Sorlie et al. 2001; Sorlie et al. 2003; Hu et al. 2006) and specific patterns of copy number aberrations (CNA) can also define genetic events associated with different subtypes (Weigman et al. ; Bergamaschi et al. 2006). Continued investigation of copy number

abnormalities in breast cancer is likely to yield insights into the pathogenesis of the disease.

With the continued advancement in technology for copy number analysis and high-throughput microarray studies and next-generation sequencing, previously difficult to detect varieties of genetic abnormalities are beginning to be unraveled. Through the use of a high-density array comparative genomic hybridization (HD-aCGH) platform, it is possible to detect both gross and fine-scale aberrations in genes (Xing et al. ; Saal et al. 2008). Small-scale CNA, here termed micro-aberrations, represent a previously under-investigated source of copy number variation that may shed light on breast subtype characteristics and tumorigenesis. Enhanced detection, cataloguing, and validation of these events could be an avenue through which we can gain a greater understanding of breast cancer genomes through the improved ability to detect genetic events affecting gene expression and function.

3.2 Results and Discussion

Copy Number Micro-aberrations are Present in Breast Tumor Subtypes

In order to test the hypothesis that primary breast cancer genomes contain areas of small-scale copy number gains and losses, termed micro-aberrations, in addition to large-scale amplifications and deletions, we designed a custom, high-resolution, high-density, comparative genomic hybridization tiling array (HD-aCGH) with an average probe resolution of 200 base pairs. We assembled a dataset of 94 tumors and 2 cell lines and tested them on this HD-aCGH array. Each tumor was classified into one of five previously defined expression subtypes using the published PAM50 identifier (Parker et al. 2009) and genomic aberrations were identified using the SWITCHdna algorithm (Weigman et al.).

We were able to identify both previously observed large scale amplifications and deletions, and novel small-scale copy number aberrations, which we have highlighted a few selected examples here. The Sum149 cell line has a previously identified micro-amplification in exon 2 of the *PTEN* gene (Saal et al. 2008), which we also clearly observed using our HD-aCGH platform (Figure 3.1a). A number of other intra-genic aberrations were also detected, including a focal deletion in *PTEN* in a basal-like tumor (Figure 3.1b), an intra-genic deletion in *RB1* in a basal-like tumor (Figure 3.1c), and a small *RB1* amplification in a basal-like tumor (Figure 3.1d).

To quantify micro-aberrations, we established a size definition of the smallest 25% of SWITCHdna identified aberrations in this dataset (<64 contiguous probes, <~15 kb). The genomic landscape of the basal-like subtype exhibited many

of these micro-aberrations, as basal-like tumors have the highest incidence of these events (Table 3.1). Basal-like tumors were seen to have the highest average number of micro-aberrations per sample, the highest median number of micro-aberrations per sample, and the greatest percentage of samples exhibiting micro-aberrations (Table 3.1).

The potential power of the high-density tiling array platform is shown in the example of the *RB1* data (Figure 3.2 and 3.3). Each identified micro- and macro-aberration segment for this gene is plotted with the RB1 exons identified by each gray stripe, along with the location of each RB1 probe on the tiling array (green segments) and an earlier 109,000 feature single nucleotide polymorphism (SNP) platform (5 probes: orange dots). The micro-aberration segments overlap with at most one genome-wide probe, and thus would have never been called a loss given how most aCGH programs call "changed segments", which was the case for SWITCHdna. In addition, several macro-aberrations identified from the tiling array platform have minimal overlap with the 109K genome-wide probes. 49/94 samples had previously been assayed on the 109k SNP platform (Weigman et al.), and the results of this overlap set for *RB1* were compared directly in terms of CNA assignment agreement by SWITCHdna. Again, focusing on the *RB1* gene as our example case, we observed six copy number aberrations in these 49 samples by the tiling array for *RB1* (Table 3.2). Only two of these six CNA were detectable by the 109k SNP array and these were the aberrations that spanned the whole gene. The remaining four, all of which were intra-genic, were completely missed by the copy number segments generated from the whole-genome array (Table 3.2). This is

illustrated in two example *RB1* gene plots directly comparing the probes and segments from the high-density tiling array and the 109k SNP array (Figure 3.3).

Copy Number Status Correlated with Expression

Beyond simply determining the frequency of micro-aberrations present within each gene or tumor, we wanted to assess whether the presence of micro-aberrations would result in functional consequences. We first assigned a copy number status for each gene and each sample (gross copy number gain, gross copy number loss, micro-amplification, or micro-deletion) and then within each gene, examined whether the corresponding gene expression was concordant with the type of genomic aberration observed, i.e. do micro-amplifications result in increased expression and/or do micro-deletions result in decreased expression of the involved gene. We found similar rates of concordant expression between micro-aberrations and gross aberrations, with 30-40% of the genes showing 100% agreement between aberration type and gene expression, meaning that for these genes, every sample that displayed a micro-amplification in the gene also had greater than median expression of the gene and likewise every sample that had a micro-deletion in the gene had less than median expression of the gene. Another 50-75% of the tested genes showed at least 50% concordance between aberration type and gene expression meaning at least half of the samples that displayed a copy number aberration had altered expression of the affected gene in the same direction as the CNA (Table 3.3). These findings suggest that the micro-aberrations have functional effects upon gene expression similar to what is seen with larger scale CNA. A number of genes displaying micro-aberrations also showed differential expression

of the involved gene when comparing the aberrant vs. non-aberrant groups by ANOVA (Figure 3.4). ANOVA box plots are shown for the genes *NUF2* (Figure 3.4a) and *UBE2T* (Figure 3.4b), where samples with micro-amplifications have significantly higher expression of the gene than those without micro-amplifications. Also shown are *ZNF217* (Figure 3.4c), where samples with micro-deletions have lower expression of the gene than those without micro-deletions and *SLC7A6* (Figure 3.4d), where the samples with micro-deletions have higher expression than the samples that do not; we do note that the sample size is small in some cases, but the trends are present.

Genomic Micro-amplification Causes Exon Skipping

We performed a closer examination of the micro-amplification of the *PTEN* gene in the SUM149 cell line, as it is a validated aberration that has now been shown by multiple groups. Using mRNA-seq data, we assessed the expression of the *PTEN* gene on an exon level to determine the functional consequence of the DNA micro-amplification. For comparison, we also examined the data for the SUM102 cell line, which has no genomic alterations in *PTEN*. The distribution of aligned reads for each exon is shown for each cell line (Figure 3.5). In SUM149, there is a lack of *PTEN* gene expression starting from the middle of exon 2, which coincides with the location of the genomic DNA micro-amplification. In comparison, the SUM102 cell line has aligned reads throughout the entirety of the *PTEN* gene, thus the micro-amplification causes a loss of expression of all downstream exons.

The 5' and Promoter Regions of Genes are Most Commonly Affected by Micro-aberrations

In order to assess whether certain regions of genes were more commonly affected by micro-aberrations than others, we portioned each gene on the array into four quadrants: 5' End, 5' Middle, 3' Middle, and 3' End based upon a proportional splitting of each gene into 4 equal segments. For every micro-aberration instance (n=330), we noted the quadrants that it occupied, and then for each quadrant, determined what proportion of the total possible quadrants were affected by a micro-aberration event (Figure 3.6a). We found that the 5' end of the gene was disproportionately affected by micro-aberrations (~88% vs. <40%). Additional refinement of the affected region was also performed and a large percentage of micro-aberrations also affected the promoter (defined as area upstream of the coding region) and 5' UTR regions as well (80.6% and 79.7% respectively). Of the aberrations whose area of effect was limited only to the 5' End, the promoter region was affected at a higher rate than the 5' UTR region (Figure 3.6b, 92.0% vs. 68.1%).

Micro-aberration Frequency Associated with Poorer Survival

The survival outcomes of patients with varying levels of copy number aberrations were also assessed to determine if there was an association with outcomes. We identified the number of copy number micro-aberrations per sample using our SWITCHdna criteria and rank-ordered the patients in terms of micro-aberration frequency. Patients were assigned to one of two groups depending on whether they were in the top 67% of microCNA or the bottom 33%. Kaplan-Meier analysis was performed examining overall (Figure 3.7a) and relapse-free survival

(Figure 3.7b). We saw that patients with the least genomic instability as assessed by SWITCHdna-called micro-aberrations had significantly better outcomes in terms of both overall and relapse-free survival.

Cell Cycle Genes are Frequently Micro-aberrant

Lastly, we examined the frequency of micro-aberrations for each gene. Here, we list the top 17 most micro-aberrant genes among those that were tested on this tiling array (Table 3.4a). Gene Set Enrichment Analysis shows that genes exhibiting multiple micro-aberrations in our study were likely to be found in cell cycle-related pathways (Table 3.4b).

Discussion

The previous discovery of focal aberrations within genes (Saal et al. 2008) using a high-density aCGH array, and the lack of description of such features in many whole-genome aCGH-based breast cancer studies (Weigman et al. ; Bergamaschi et al. 2006; Chin et al. 2006; Wood et al. 2007) suggests that these micro-aberrations may occur regularly in breast cancer genomes and that they simply have not yet been detected in previous studies. To address this hypothesis, we assembled a dataset of 94 breast tumors and two breast cancer cell lines and tested them on a custom-designed aCGH tiling array; this array was targeted to 128 gene panel focused on important cancer relevant genes (Supplemental Table 3.1) and previously identified basal-like cancer specific regions and genes (Fedele et al. ; Weigman et al. ; Bergamaschi et al. 2006; Johannsdottir et al. 2006).

By utilizing a previously tested segmentation and aberration calling algorithm called SWITCHdna (Weigman et al.), we analyzed the tiling array data and

proceeded to generate a numerical definition of a micro-aberration (Supplementary Table 3.2). Essentially all of these micro-aberrations would be mostly undetectable using lower-resolution genome-wide platforms, as these segments would be covered by at most one probe on such arrays (Figure 3.2 and 3.3). An analysis of the frequency of micro-aberrations within our dataset samples showed that the basal-like subtype had the most frequent occurrences of these events, mirroring their high overall genomic instability (Ding et al. ; Russnes et al. ; Bergamaschi et al. 2006; Chin et al. 2006; Van Loo et al. 2010); thus the presence of micro-aberrations did correlate with the presence of large aberrations. Our data also shows that at least in some cases, these events have functional downstream consequences (Figure 3.4 and 3.5, Table 3.3). The finding of exon skipping at the point of the focal amplification in the PTEN gene in the SUM149 cell line is particularly interesting given the otherwise normal copy number. The exact mechanism that induces this exon skipping is yet to be determined, but one can imagine that some aspect of the amplified DNA sequence results in an alteration to the pre-processed transcript that could cause early truncation or some sort of structural interference (Monaco et al. 1988; Liu et al. 2001).

We did find that the initial 5' end of these genes was the most heavily affected by micro-aberration events, specifically the promoter region of the gene (Figure 3.6). It is unclear what leads to this predilection, but it does suggest that this specific portion of the gene may be more prone to genomic stress. The involvement of the promoter region does suggest that this site of active transcriptional processing may lead to a structural genomic weakness that causes a predisposition

towards micro-aberrations. This finding may also aid in explaining either the factors involved in the formation of micro-aberrations versus macro-aberrations or the possible downstream consequences of such events.

A limitation of our study is that we are currently unable to determine if any of these micro-aberrations are subtype specific. If they are, this would 1) mirror whole genome study findings, and 2) potentially showcase an alternative means of gene disruption that unravels previously unexplained expression data. As one example, RB1 dysfunction has been shown to be associated with the basal-like subtype (Jiang et al. 2010). We have found in our own studies that RB1-LOH was highly correlated with gene expression patterns and patient outcomes while RB1 protein expression on the same samples was not (Herschkowitz et al. 2008). An intra-genic micro-aberration could potentially explain such cases, as the genetic abnormality may only affect a portion of the gene such that a protein is still produced and the majority of it intact, but it does not function properly. Expanded studies with high-resolution platforms like whole genome sequencing will allow us to answer this question.

From a broader viewpoint, it stands to reason that other tumor types may also exhibit these types of events, but as yet they have not been widely described (2008). However, similar intragenic deletions in *RB1* and *PTEN* were recently described in melanoma cell lines (SKMEL-207, A2058, SKMEL-178 (Xing et al.)), suggesting that these types of micro-genomic events are present in other cancers and will reveal themselves using the newer methodologies. It may also be true that

only certain cancer types display these aberrations though, which would suggest that there are mechanisms of genomic instability unique to these cancers.

In examining the specific genes on the tiling array that displayed micro-aberrations, we noted that genes that were found in cell-cycle related pathways were particularly prone to these small-scale events. Given that our panel of targeted genes was focused on cancer relevant genes, there was some inherent enrichment for this class of genes, but even within the background of genes on the array itself, there was a predilection for cell cycle genes. Coupled with our finding of micro-aberrations being localized to promoter regions, it may be that cell cycle genes are more prone to these events because of active transcriptional activity.

We were also able to make the observation that higher genomic instability in the form of micro-aberrations in our dataset was associated with poorer survival outcomes (Figure 3.7). An overall high level of genomic instability has been found to associate with worse survival (Trope et al. 2000; Kristensen et al. 2003), and here we see that finding extended to micro-instability in our dataset. There was however, a high concordance of overlap between patients that had many large scale changes and many micro-aberrations, and overall, patients with higher total numbers of CNA were associated with poorer outcomes (data not shown). Nonetheless, the same rank-ordering split that was performed on the micro-aberrations did not result in identical findings for overall aberrations, suggesting that while there could be confounding of the micro-aberration survival findings by overall genomic instability, there may also be characteristics unique to the micro-aberrations themselves. Furthermore, the concordance between gross aberrations

and micro-aberrations suggests that there may be common mechanisms of genomic instability at play, which may yield insights into how micro-aberrations arise.

Conclusions

In addition to exhibiting gross copy number changes, breast tumor genomes contain focal micro-aberrations as well when examined using high-resolution platforms. These micro-aberrations occur within the background of global genomic instability and can have disruptive effects upon gene expression. These micro-events represent a potential means of mutagenesis in genes that have been otherwise determined to be normal in terms of gross copy number or SNV-based somatic mutations. Continued investigation into these events with improved tools will allow their increased detection and use as identifying characteristics of specific tumor.

3.3 Materials and Methods

Breast Cancer Patient Dataset

The dataset chosen for this study contained both gene expression and high-density array comparative genomic hybridization (HD-aCGH) copy number data from a set of breast tumors from UNC "HD-UNC94" (n=94). Additionally, the SUM102 and SUM149 breast cancer cell lines were also tested using these high-density tiling arrays. All samples were collected using IRB-approved protocols and all patients signed informed consent forms. Tumors in the dataset were assayed for gene expression patterns using Agilent DNA microarrays. Log_2 ratio data was taken from the UNC Microarray Database (UMD), filtering for a lowess normalized intensity value of 10 or above for each channel, and 70% good data, and then used for further analyses.

Classifying tumors for expression-based subtype classification

The Lowess normalized R/G Log_2 ratio data from the HD-UNC94 was used from the two different microarray platforms employed in this study (GPL5325 Agilent UNC Perou Lab Homo sapiens 1X44K Custom Array, Agilent UNC Perou Lab Homo sapiens 1X244k Custom Array). The datasets were then limited to the probes shared across both platforms. After column standardization of both platforms (samples at N(0,1)), Distance Weighted Discrimination (DWD) (Benito et al. 2004) was used to remove platform bias prior to classification. After correction, the R/G Log_2 ratio data was collapsed (via averaging) from probes to HGNC gene symbols. The PAM50 gene set predictor(Parker et al. 2009) was used to assign subtypes to the tumors.

Tiling array design

The custom HD-aCGH tiling platform was designed using Agilent's E-array v5.0 online (https://earray.chem.agilent.com/earray/) software and built on the Human 244k Custom Oligo platform. 230,606 probes cover a total region of 45 Mb, which includes the full genomic sequence of the 128 genes of interest as well as the region 150 kb upstream and downstream of each gene (Supplementary Table 3.1) for an average resolution of 200bp. Labeling and hybridization were performed according to manufacturer's instructions using the Agilent Genomic DNA Labeling Kit PLUS. A Human Genomic DNA Pool (Promega) was used as reference DNA. Microarrays were scanned on an Agilent DNA Microarray scanner (G2565CA) and the data uploaded to the University of North Carolina Microarray Database (UMD, www.genome.unc.edu).

Identification of CNA and microCNA with SWITCHdna

To determine regions of copy number aberrations (CNA), we utilized the SWITCHdna algorithm (Weigman et al.), focusing on individual genes. For the purposes of this study, the analysis window was limited to the genomic region + 5kb upstream/downstream (Supplementary Table 3.1). In order to further filter the identified segments, we set the cutoff for the absolute value of the $\log_2$ ratio to be greater than 0.30 in order to reduce false positives. After identifying all genomic segments of alteration with SWITCHdna, we analyzed the distribution of sizes of aberrant segments and established a cutoff of <64 contiguous SWITCH probes, or ~≤15kb as the definition of a micro-aberration, which equates to the smallest 25% of CNA in this dataset (Supplementary Table 3.2).

In order to identify the regions of genes, most commonly affected by micro-aberrations, each gene was divided into four quadrants based on proportional splitting of each gene into four equal segments: 5' End (promoter region, 5'UTR, beginning regions of gene), 5' Middle (first ½ of gene), 3' Middle (second ½ of gene), 3' End (end regions of gene, 3' UTR, downstream region). For every micro-aberration instance, the affected quadrants were tallied and for each quadrant, the proportion of affected quadrants out of all possible quadrants was calculated. Similarly, each micro-aberration instance was assessed in terms of whether it encompassed the promoter or 5' untranslated region (UTR) of each gene, with the promoter region defined as genomic space upstream of the transcription start site.

mRNA-seq

mRNA-seq was performed on RNA isolated from cell lines and tumors using the Qiagen RNeasy Mini Kit (Cat. No. 74104). Library preparation was performed using the TruSeq RNA Sample Kit from Illumina (Cat. No. RS-930-2001) following the low throughput protocol detailed in the manufacturer's guidelines. 1x76bp nucleotide reads were generated using an Illumina GAII sequencer.

Survival Analysis

The patients in the dataset were rank ordered by total number of SWITCHdna-defined aberrations and micro-aberrations and separated in the top 67% and bottom 33%. Survival analyses were performed using the Kaplan-Meier test in R (R Development Core Team 2009).

3.4 Acknowledgements

This work was supported by funds from the NCI Breast SPORE program (P50-CA58223-09A1), by RO1-CA138255, RO1-CA148761, T32-GM008719, F30-ES018038, and by the Breast Cancer Research Foundation.

3.5 Figures and Legends

Figure 3.1: Selected examples of intra-genic micro-aberrations.

A) The previously identified *PTEN* exon 2 micro-amplification in SUM149 cell line DNA is observed, as is B) an intra-genic deletion also in *PTEN* in a basal-like tumor. Both C) focal deletion and D) amplification are seen in the *RB1* gene, both in basal-like tumors.

Figure 3.1

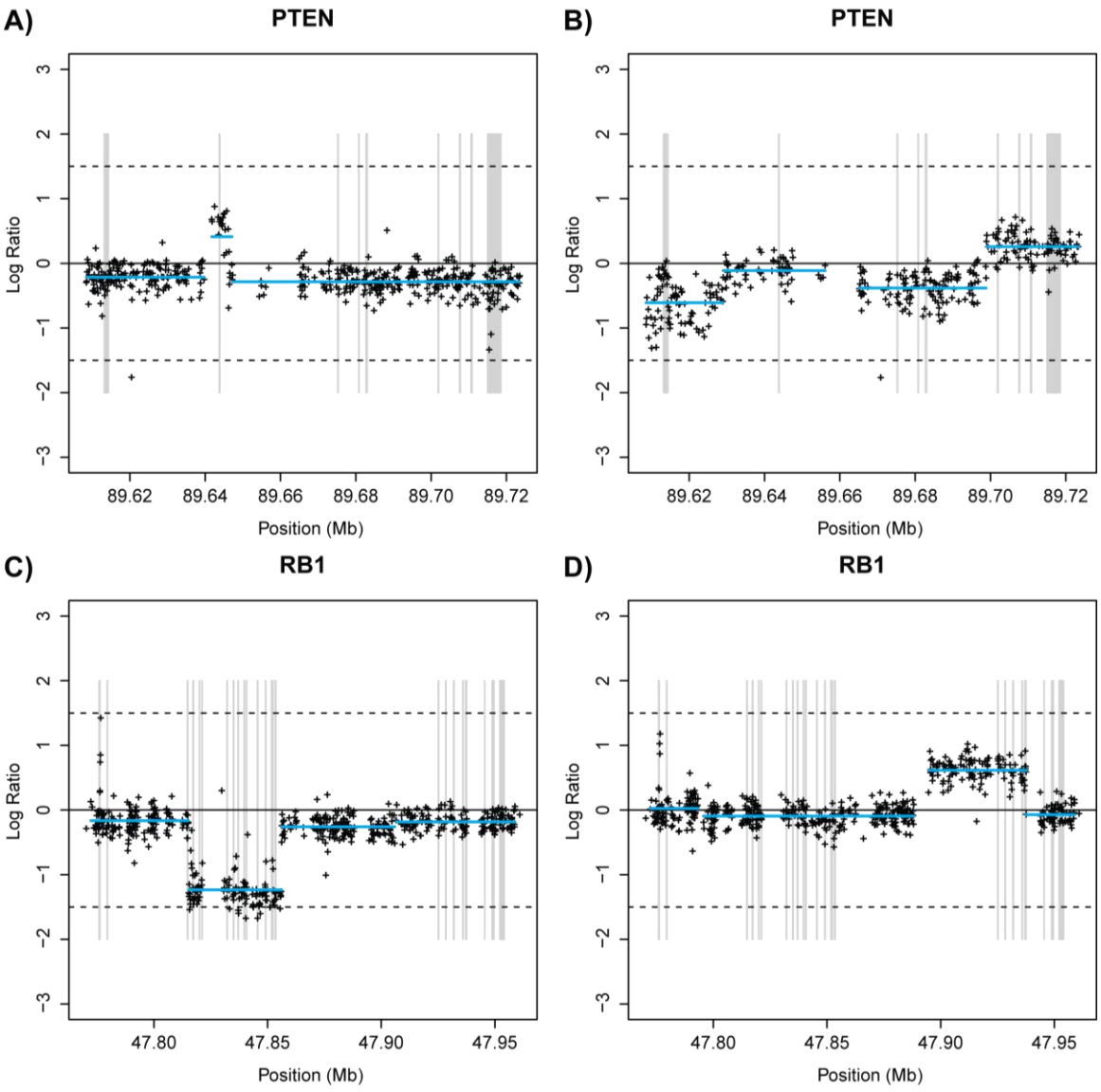

A)
PTEN
B)
PTEN
C)
RB1
D)
RB1
Log Ratio
Position (Mb)
89.62
89.64
89.66
89.68
89.70
89.72
47.80
47.85
47.90
47.95
3
2
1
0
−1
−2
−3

Figure 3.2: Aberrations and probe locations for the *RB1* gene.

Each copy number aberrant segment called by SWITCHdna is plotted by its location within the *RB1* gene. Red segments indicate micro-aberrations and blue segments indicate macro-aberrations. Exons are highlighted in grey bars. The locations of the tiling array probes are indicated with green circles and the corresponding locations of probes from a previously used 109k genome-wide SNP platform are indicated with orange circles.

Figure 3.2

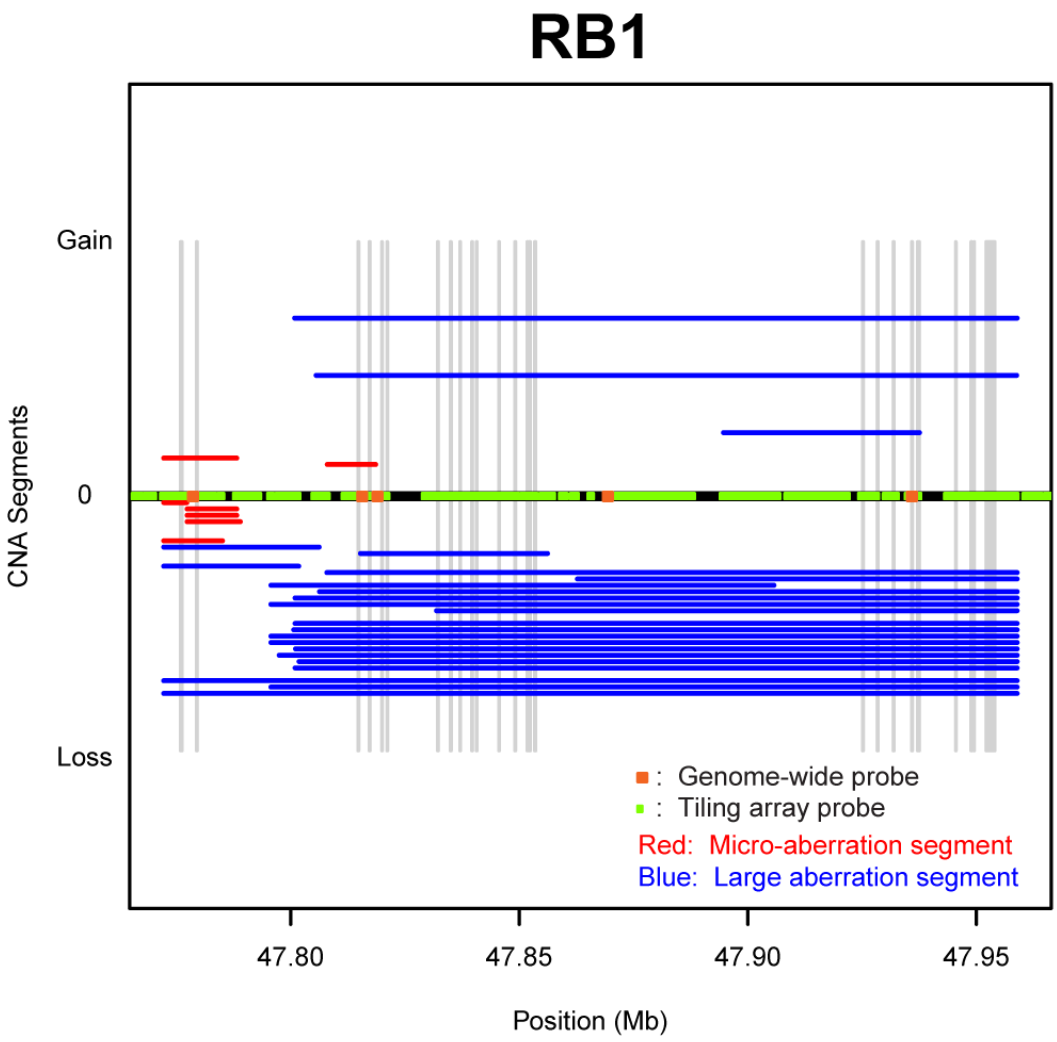

RB1
Gain
0
Loss
CNA Segments
: Genome-wide probe
: Tiling array probe
Red: Micro-aberration segment
Blue: Large aberration segment
47.80
47.85
47.90
47.95
Position (Mb)

Figure 3.3: Comparison of tiling array and 109k SNP platform generated SWITCHdna copy number segments for *RB1*.

Two representative examples are shown illustrating the different detection thresholds achieved by the tiling array platform and the 109k SNP platform for CNA in *RB1*. Exons are highlighted in grey bars. The locations of the tiling array probes are indicated with crosses and the corresponding locations of the 109k genome-wide SNP probes are indicated with orange circles. SWITCHdna called segments for each platform are also shown with tiling array segments in blue and 109k platform segments in orange.

Figure 3.3

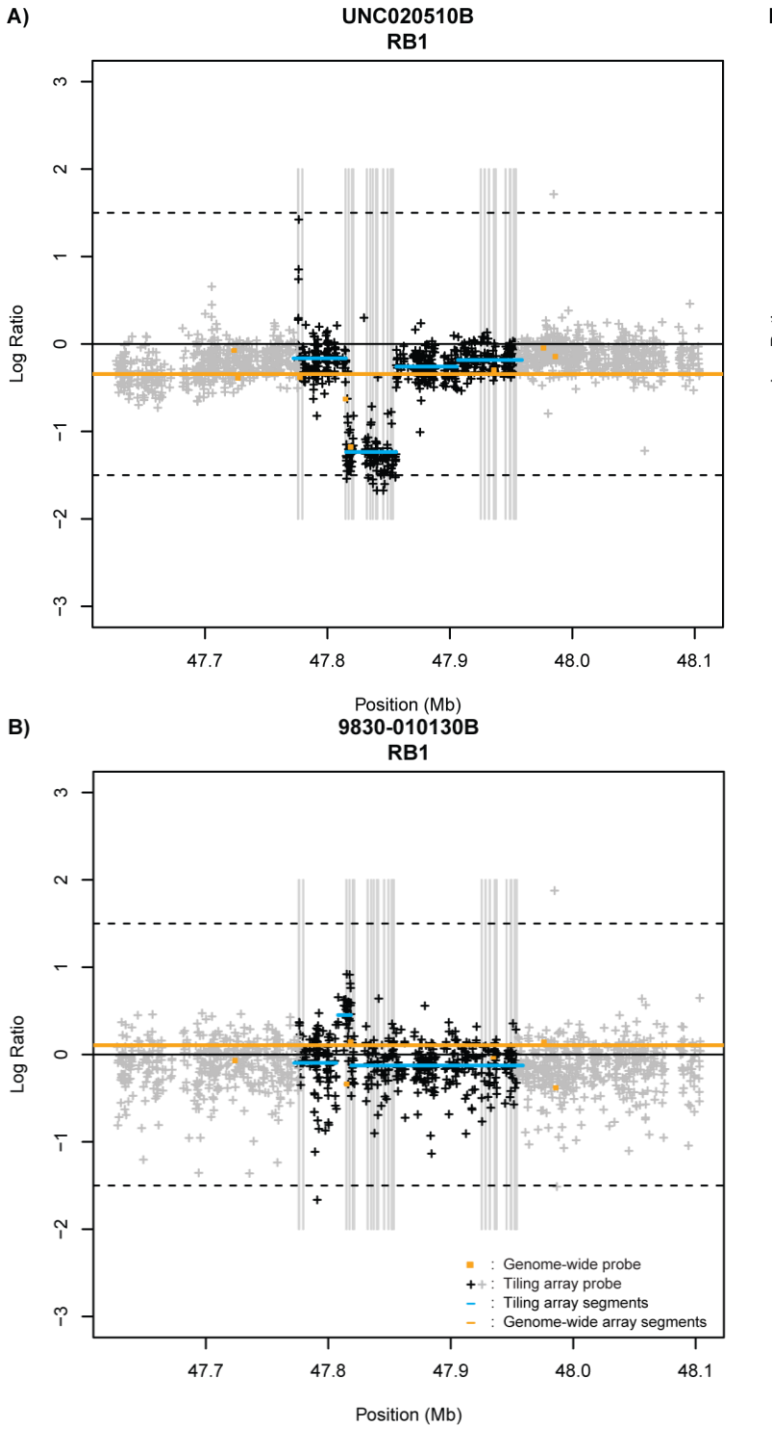

A)
UNC020510B
RB1
Log Ratio
3
2
1
0
−1
−2
−3
47.7
47.8
47.9
48.0
48.1
Position (Mb)
B)
9830-010130B
RB1
Log Ratio
3
2
1
0
−1
−2
−3
: Genome-wide probe
: Tiling array probe
: Tiling array segments
: Genome-wide array segments
47.7
47.8
47.9
48.0
48.1
Position (Mb)

Figure 3.4: The presence of micro-aberrations can result in differential expression by copy number status.

Samples with micro-amplifications in A) *NUF2* and B) *UBE2T* are associated with significantly higher expression of the gene than samples without these aberrations. Samples with micro-deletion in C) *ZNF217* are associated with significantly lower expression of the gene than samples without these aberrations. Samples with micro-deletion in D) *SLC7A6* are associated with significantly higher expression of the gene than samples without these aberrations.

Figure 3.4

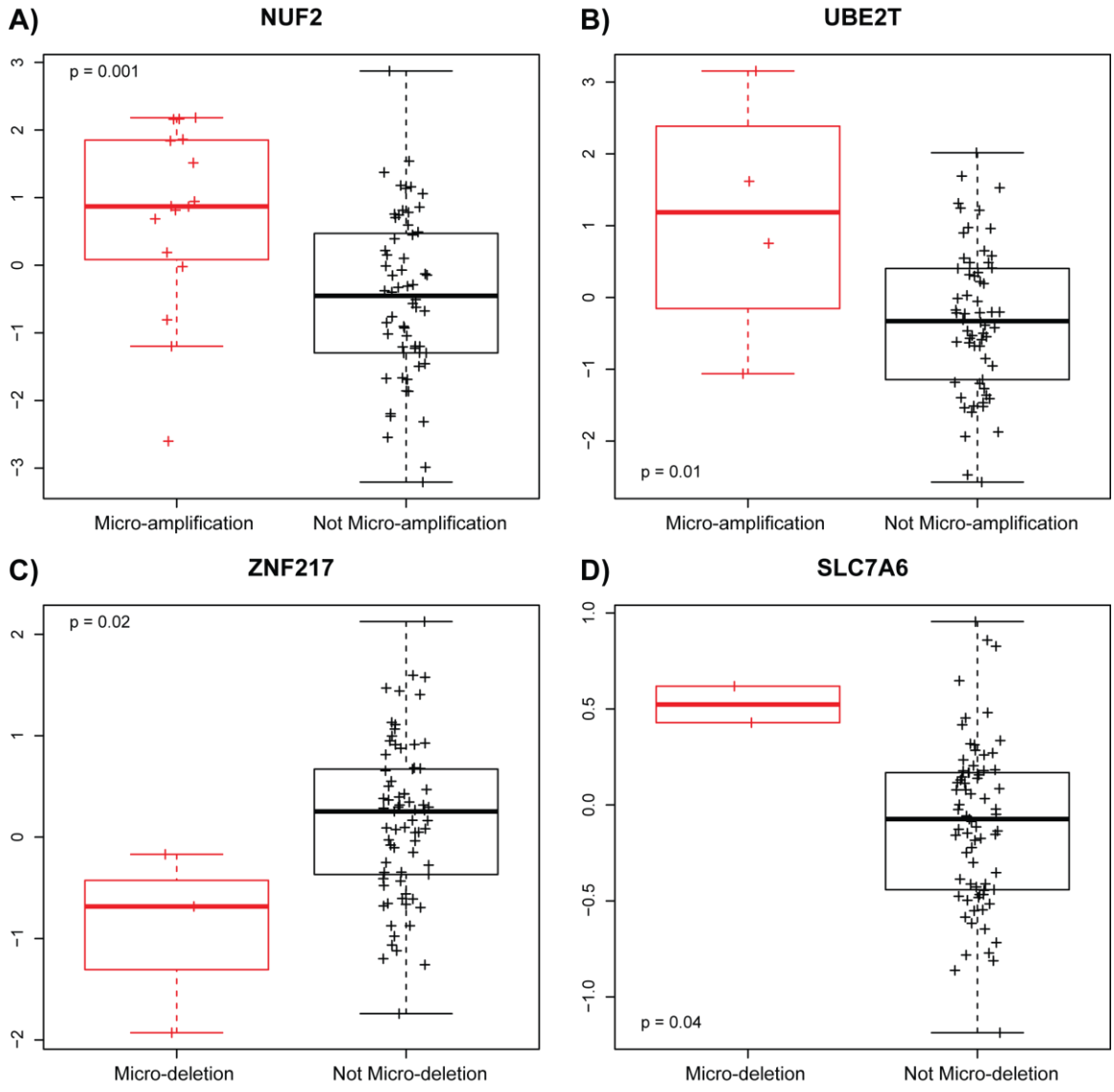

A)
NUF2
p = 0.001
Micro-amplification
Not Micro-amplification
B)
UBE2T
p = 0.01
Micro-amplification
Not Micro-amplification
C)
ZNF217
p = 0.02
Micro-deletion
Not Micro-deletion
D)
SLC7A6
p = 0.04
Micro-deletion
Not Micro-deletion

Figure 3.5: A micro-amplification in *PTEN* in the SUM149 cell line results in lack of exon expression from the site of the aberration onwards.

A) The distribution of mRNA-seq reads by exon is shown for the SUM149 and SUM102 cell lines. B) For reference, the genomic space of the *PTEN* gene is shown, along with the copy number status across the *PTEN* gene in SUM149.

Figure 3.5

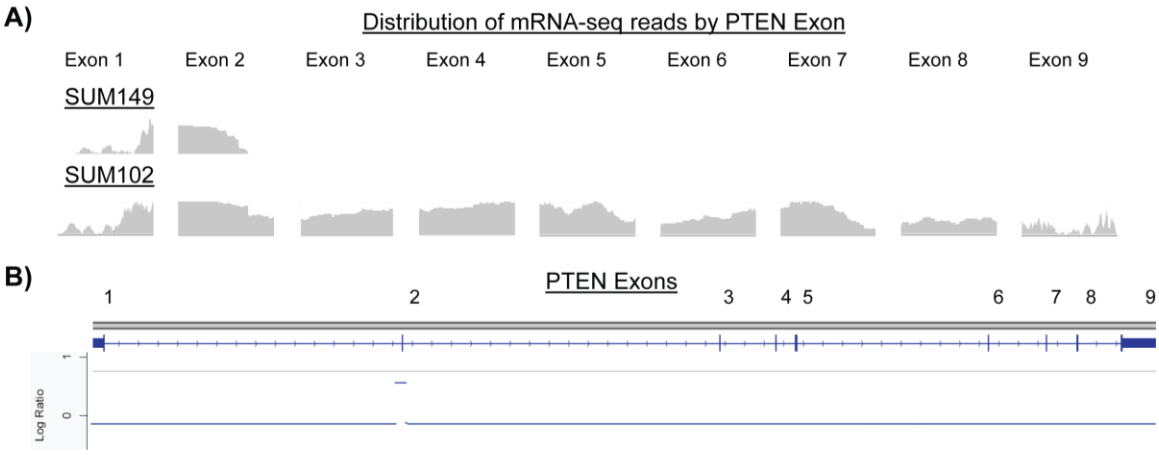
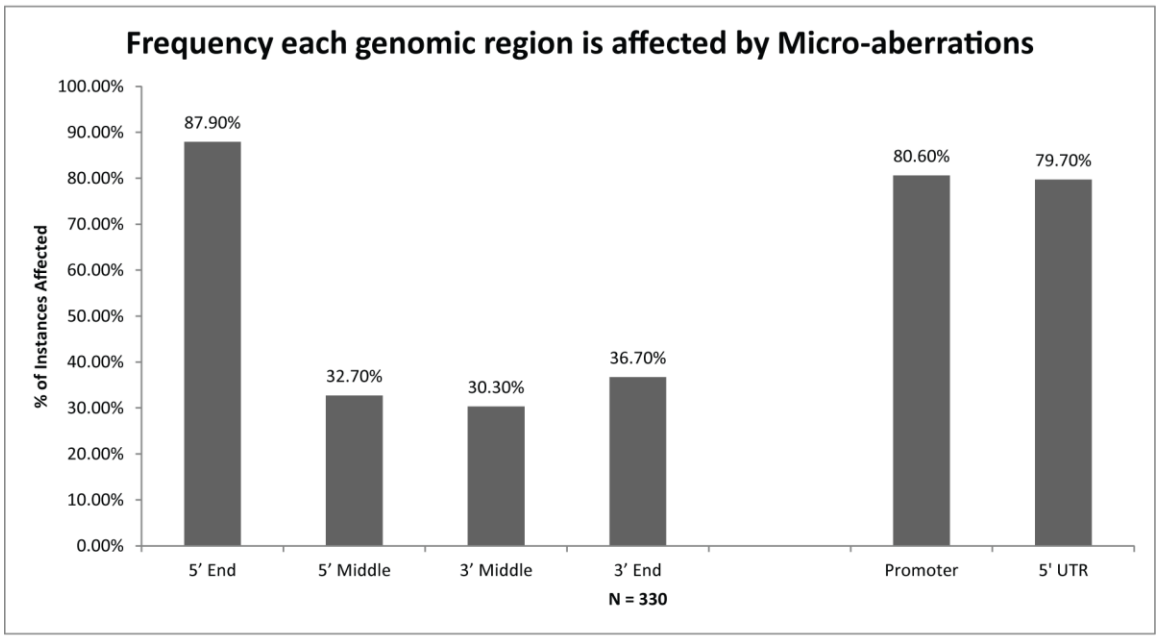

A)
Distribution of mRNA-seq reads by PTEN Exon
Exon 1
Exon 2
Exon 3
Exon 4
Exon 5
Exon 6
Exon 7
Exon 8
Exon 9
SUM149
SUM102
B)
PTEN Exons
1
2
3
4 5
6
7
8
9
Log Ratio
1
0

Figure 3.6: Frequency that genomic quadrants are affected by micro-aberrations.

A) The percentage of total possible instances that a given genomic region is overlapped by a micro-aberration segment is displayed for each genomic quadrant (5' End, 5' Middle, 3' Middle, 3'End) and the 5' UTR and promoter regions. B) For the micro-aberrations that only affected the 5' End region, the % of instances where it affected the promoter or 5' UTR is listed.

Figure 3.6

A)


Frequency each genomic region is affected by Micro-aberrations
% of Instances Affected
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
87.90%
32.70%
30.30%
36.70%
80.60%
79.70%
5′ End
5′ Middle
3′ Middle
3′ End
Promoter
5' UTR
N = 330

B)

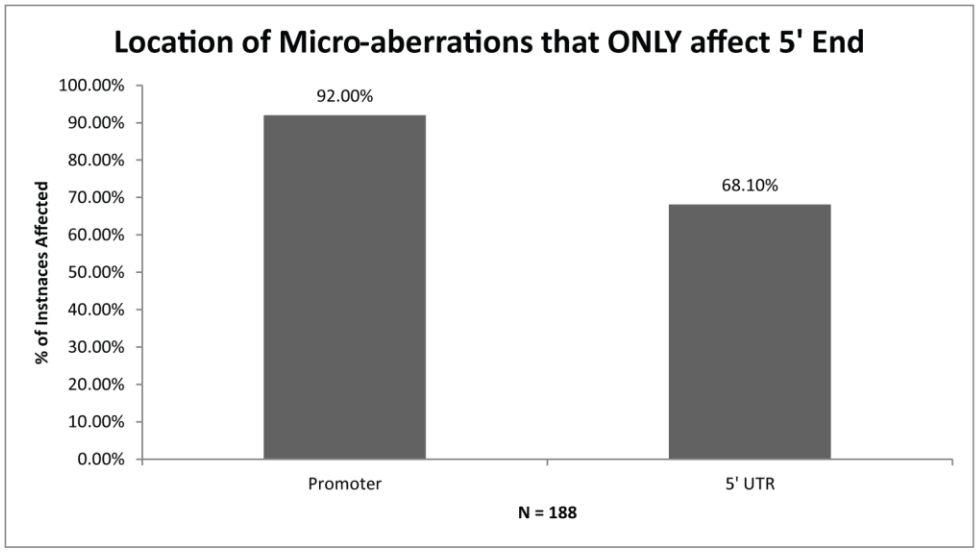

Location of Micro-aberrations that ONLY affect 5' End
% of Instnaces Affected
100.00%
90.00%
80.00%
70.00%
60.00%
50.00%
40.00%
30.00%
20.00%
10.00%
0.00%
92.00%
68.10%
Promoter
5' UTR
N = 188

Figure 3.7: Higher levels of micro-aberrations are associated with worse survival outcomes.

A) Kaplan-Meier plots for overall survival and B) relapse-free survival are shown for the patients in the tiling array datasets. The patients were split into two groups, the top 67% in terms of total micro-aberrations versus the bottom 33%.

Figure 3.7

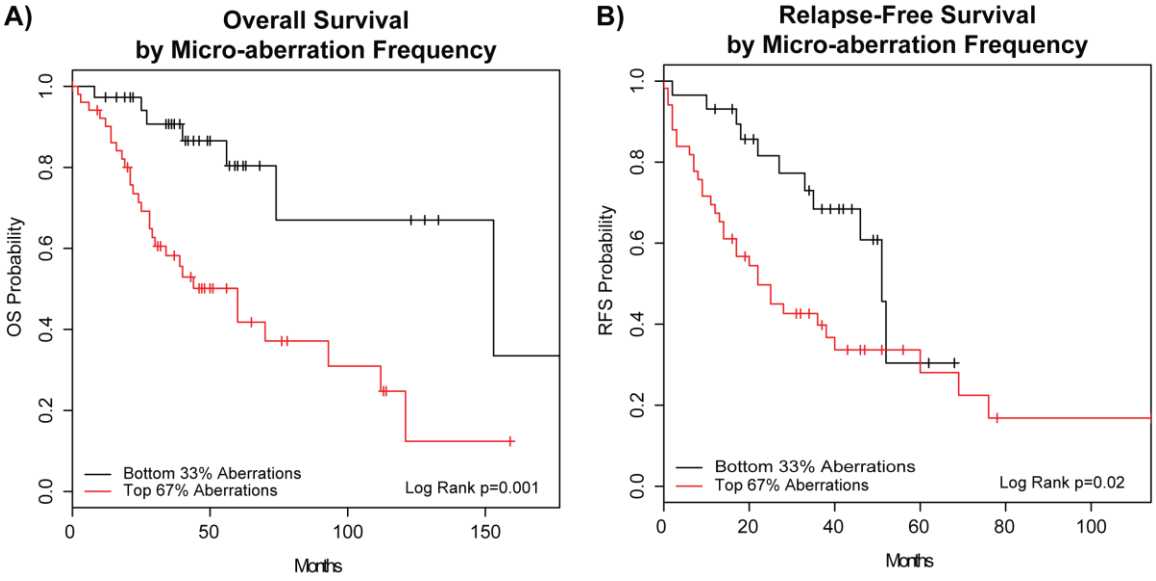

A)
Overall Survival
by Micro-aberration Frequency
OS Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
50
100
150
Months
Bottom 33% Aberrations
Top 67% Aberrations
Log Rank p=0.001
B)
Relapse-Free Survival
by Micro-aberration Frequency
RFS Probability
0.0
0.2
0.4
0.6
0.8
1.0
0
20
40
60
80
100
Months
Bottom 33% Aberrations
Top 67% Aberrations
Log Rank p=0.02

Table 3.1: Copy number micro-aberrations by subtype.

The mean and median numbers of micro-aberrations for samples within each subtype are shown, as is the percentage of samples within each subtype that exhibited any micro-aberrations.

Table 3.1

	Average Micro-aberrations/Sample	Median Micro-aberrations/Sample	% Samples with Micro-aberration
Subtype			
Basal (n=31)	4.29	4	83.87
LumA (n=27)	3.07	1	74.07
LumB (n=21)	3.81	2	80.95
Her2 (n=10)	2.90	2	80.00
Normal (n=5)	2.00	0	40.00

Table 3.2: *RB1* Copy Number Status on 49 samples from the tiling array and 109k platforms.

p-value determined by Fisher's Exact Test.

Table 3.2

	109k SNP array Copy Number Aberration	109k SNP array Copy Number Normal	Total
HD-aCGH array Copy Number Aberration	2	4	6
HD-aCGH array Copy Number Normal	0	43	43
Total	2	47	49
			p-value = 0.01

Table 3.3: Percentage of genes exhibiting each class of micro-aberration with concordant gene expression.

Results are shown for gross copy number gains, gross copy number losses, micro-amplifications, and micro-deletions and frequency that genes are 100% concordant by gene expression or >=50% concordant by gene expression. A concordant sample is one where a copy number gain or micro-amplification is accompanied by high expression of the affected gene or a copy number loss or micro-deletion is accompanied by low expression of the affected gene.

Table 3.3

		Genes with CNA with Concordant Gene Expression			
		Gross Gains	Gross Losses	Micro-amplifications	Micro-deletions
# 100% Concordant / # with Aberration	(%)	32/105 (30.5%)	25/83 (30.1%)	24/64 (37.5%)	22/53 (41.5%)
# >=50% Concordant / # with Aberration	(%)	79/105 (75.2%)	54/83 (65.1%)	34/64 (53.1%)	36/53 (67.9%)

Table 3.4: Listing of frequently micro-aberrant genes and pathway enrichment.

A) The top 17 genes that displayed the most micro-aberrations are shown, along with the number of micro-aberrations seen within each gene. B) Results of a Gene Set Enrichment Analysis of the genes exhibiting more than one micro-aberration are displayed, showing an enrichment of cell-cycle related pathways.

Table 3.4

A)

Top 17 Genes with the Most Micro-aberrations		
	Gene Name	**# of Micro-aberration Occurrences**
1.	NUF2	23
2.	NAT1	21
3.	FZD7	11
4.	MYC	11
5.	ELOVL5	11
6.	PIK3CA	11
7.	CENPF	10
8.	UBE2T	9
9.	S100A11	8
10.	ZNF217	8
11.	RB1	7
12.	CCNB1	7
13.	CELSR1	6
14.	MIA	6
15.	TP53BP2	5
16.	MDM2	5
17.	PTEN	5

B)

Gene Set Name	**Description**	**# Genes in Overlap**	**p-value**
BIOCARTA_P53_PATHWAY	p53 Signaling Pathway	7	$2.12\ e^{-12}$
BIOCARTA_CELLCYCLE_PATHWAY	Cyclins and Cell Cycle Regulation	6	$3.34\ e^{-9}$
BIOCARTA_RACCYCD_PATHWAY	Influence of Ras and Rho proteins on G1 to S Transition	5	$3.88\ e^{-7}$

Supplementary Table 3.1: Gene list and tiling array coverage and analysis windows.

List of the 128 genes on the tiling array with EntrezGene ID along with their genomic coordinates by NCBI build 36.1 (hg18) of the reference human genome. Coordinates are also provided showing the total coverage on the tiling array for each gene as well as the +/- 5kb analysis window that was used.

Supplementary Table 3.1

Gene Name	Entrez Gene ID	Genomic Coordinates of Gene	Gene Coverage Start Position	Gene Coverage End Position	Gene Analysis Start Position	Gene Analysis End Position
ABCC3	8714	chr17:46067227-46124062	45917227	46274062	46062227	46129062
AGR2	10551	chr7:16811133-16797960	16961133	16647960	16816133	16792960
AGR3	155465	chr7:16888138-16865555	17038138	16715555	16893138	16860555
AKT3	10000	chr1:242073176-241718158	242223176	241568158	242078176	241713158
AR	367	chrX:66680599-66860844	66530599	67010844	66675599	66865844
AURKA	6790	chr20:54400758-54377852	54550758	54227852	54405758	54372852
AVEN	57099	chr15:32118595-31945720	32268595	31795720	32123595	31940720
BAG1	573	chr9:33254761-33242469	33404761	33092469	33259761	33237469
BCL11A	53335	chr2:60634137-60531806	60784137	60381806	60639137	60526806
BCL2	596	chr18:59137593-58941559	59287593	58791559	59142593	58936559
BCL2A1	597	chr15:78050698-78040290	78200698	77890290	78055698	78035290
BIRC5	332	chr17:73721872-73733311	73571872	73883311	73716872	73738311
BRCA1	672	chr17:38530994-38449840	38680994	38299840	38535994	38444840
BRCA2	675	chr13:31787617-31871809	31637617	32021809	31782617	31876809
BTG3	10950	chr21:17907033-17887842	18057033	17737842	17912033	17882842
BUB1	699	chr2:111152135-111111881	111302135	110961881	111157135	111106881
CA12	771	chr15:61402783-61461128	61252783	61611128	61397783	61466128
CCNB1	891	chr5:68498669-68509828	68348669	68659828	68493669	68514828
CCND1	595	chr11:69165054-69178423	69015054	69328423	69160054	69183423
CCNE1	898	chr19:34994741-35007059	34844741	35157059	34989741	35012059
CDC20	991	chr1:43597213-43601461	43447213	43751461	43592213	43606461
CDC6	990	chr17:35697672-35712939	35547672	35862939	35692672	35717939
CDCA8	55143	chr1:37930746-37947978	37780746	38097978	37925746	37952978
CDH1	999	chr16:67328696-67676586	67178696	67826586	67323696	67681586
CDH3	1001	chr16:67236277-67290443	67086277	67440443	67231277	67295443
CDK4	1019	chr12:56432431-56428270	56582431	56278270	56437431	56423270
CDKN1B	1027	chr12:12761576-12766570	12611576	12916570	12756576	12771570
CDKN2A	1029	chr9:21984490-21957751	22134490	21807751	21989490	21952751
CDKN2B	1030	chr9:21999312-21992902	22149312	21842902	22004312	21987902
CELSR1	9620	chr22:45311731-45135395	45461731	44985395	45316731	45130395
CENPA	1058	chr2:26862386-26870961	26712386	27020961	26857386	26875961
CENPF	1063	chr1:212843155-212904537	212693155	213054537	212838155	212909537
CEP55	55165	chr10:95246399-95278839	95096399	95428839	95241399	95283839
CFLAR	8837	chr2:201689135-201737260	201539135	201887260	201684135	201742260
CKS2	1164	chr9:91115933-91121438	90965933	91271438	91110933	91126438
CLDN4	1364	chr7:72883129-72884951	72733129	73034951	72878129	72889951

CLDN7	1366	chr17:7106519-7103946	7256519	6953946	7111519	7098946
COX6C	1345	chr8:100975071-100959548	101125071	100809548	100980071	100954548
CTPS	1503	chr1:41217594-41250822	41067594	41400822	41212594	41255822
E2F1	1869	chr20:31737854-31727150	31887854	31577150	31742854	31722150
EGFR	1956	chr7:55054219-55242525	54904219	55392525	55049219	55247525
ELOVL5	60481	chr6:53321901-53240155	53471901	53090155	53326901	53235155
EPCAM	4072	chr2:47449971-47467661	47299971	47617661	47444971	47472661
ERBB2	2064	chr17:35097919-35138441	34947919	35288441	35092919	35143441
ESR1	2099	chr6:152170379-152466099	152020379	152616099	152165379	152471099
EXO1	9156	chr1:240078158-240119671	239928158	240269671	240073158	240124671
FBP1	2203	chr9:96441624-96405236	96591624	96255236	96446624	96400236
FGFR4	2264	chr5:176446527-176457733	176296527	176607733	176441527	176462733
FIGF	2277	chrX:15312498-15273639	15462498	15123639	15317498	15268639
FOXA1	3169	chr14:37134240-37128940	37284240	36978940	37139240	37123940
FOXC1	2296	chr6:1555680-1559131	1405680	1709131	1550680	1564131
FZD7	8324	chr2:202607555-202611405	202457555	202761405	202602555	202616405
GALNT7	51809	chr4:174326479-174481693	174176479	174631693	174321479	174486693
GATA3	2625	chr10:8136673-8157170	7986673	8307170	8131673	8162170
GPR160	26996	chr3:171239397-171285878	171089397	171435878	171234397	171290878
GRB7	2886	chr17:35147713-35157064	34997713	35307064	35142713	35162064
GSTM1	2944	chr1:110031965-110037890	109881965	110187890	110026965	110042890
GSTM3	2947	chr1:110084563-110078077	110234563	109928077	110089563	110073077
ID4	3400	chr6:19945596-19948894	19795596	20098894	19940596	19953894
IGF1	3479	chr12:101398454-101313806	101548454	101163806	101403454	101308806
IGFBP2	3485	chr2:217206372-217237404	217056372	217387404	217201372	217242404
KIF13B	23303	chr8:29176529-28980714	29326529	28830714	29181529	28975714
KIF2C	11004	chr1:44978138-45006003	44828138	45156003	44973138	45011003
KPNA1	3836	chr3:123716474-123623438	123866474	123473438	123721474	123618438
KRAS	3845	chr12:25295121-25249447	25445121	25099447	25300121	25244447
KRT14	3861	chr17:36996673-36992059	37146673	36842059	37001673	36987059
KRT17	3872	chr17:37034335-37029220	37184335	36879220	37039335	37024220
KRT18	3875	chr12:51628922-51632952	51478922	51782952	51623922	51637952
KRT19	3880	chr17:36938167-36933395	37088167	36783395	36943167	36928395
KRT5	3852	chr12:51200510-51194626	51350510	51044626	51205510	51189626
KRT8	3856	chr12:51585127-51577238	51735127	51427238	51590127	51572238
MAPT	4137	chr17:41327624-41461547	41177624	41611547	41322624	41466547
MDM2	4193	chr12:67488247-67520481	67338247	67670481	67483247	67525481
MIA	8190	chr19:45973140-45975235	45823140	46125235	45968140	45980235
MKI67	4288	chr10:129814645-129784913	129964645	129634913	129819645	129779913
MMP11	4320	chr22:22445036-22456503	22295036	22606503	22440036	22461503
MYB	4602	chr6:135544146-135582003	135394146	135732003	135539146	135587003

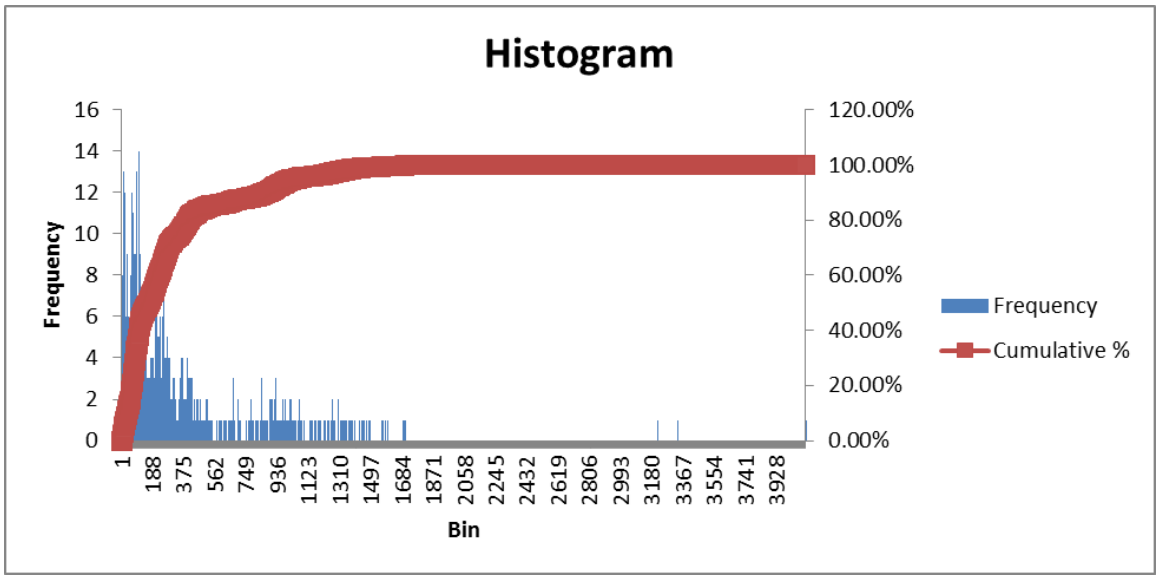
MYBL2	4605	chr20:41729123-41778536	41579123	41928536	41724123	41783536
MYC	4609	chr8:128817498-128554220	128967498	128404220	128822498	128549220
NAT1	9	chr8:18111895-18125100	17961895	18275100	18106895	18130100
NDC80	10403	chr18:2561605-2606629	2411605	2756629	2556605	2611629
NUF2	83540	chr1:161558347-161592177	161408347	161742177	161553347	161597177
ORC6L	23594	chr16:45281059-45289807	45131059	45439807	45276059	45294807
P4HTM	54681	chr3:49002345-49019586	48852345	49169586	48997345	49024586
PCNA	5111	chr20:5055268-5043599	5205268	4893599	5060268	5038599
PGR	5241	chr11:100506465-100414313	100656465	100264313	100511465	100409313
PHGDH	26227	chr1:120056033-120088361	119906033	120238361	120051033	120093361
PIK3CA	5290	chr3:180349005-180435194	180199005	180585194	180344005	180440194
PRC1	9055	chr15:89338808-89310272	89488808	89160272	89343808	89305272
PTEN	5728	chr10:89613175-89718512	89463175	89868512	89608175	89723512
PTTG1	9232	chr5:159781443-159788324	159631443	159938324	159776443	159793324
RACGAP1	29127	chr12:48705488-48669212	48855488	48519212	48710488	48664212
RAD17	5884	chr5:68700880-68746384	68550880	68896384	68695880	68751384
RAD50	10111	chr5:131920529-132007498	131770529	132157498	131915529	132012498
RAD51C	5889	chr17:54124962-54166691	53974962	54316691	54119962	54171691
RAD51L1	5890	chr14:67356262-68132367	67206262	68282367	67351262	68137367
RARA	5914	chr17:35718972-35767420	35568972	35917420	35713972	35772420
RB1	5925	chr13:47775884-47954027	47625884	48104027	47770884	47959027
REEP6	92840	chr19:1442165-1448924	1292165	1598924	1437165	1453924
RRM2	6241	chr2:10180184-10188074	10030184	10338074	10175184	10193074
S100A11	6282	chr1:150276135-150271606	150426135	150121606	150281135	150266606
S100A8	6279	chr1:151630173-151629132	151780173	151479132	151635173	151624132
S100A9	6280	chr1:151596954-151600127	151446954	151750127	151591954	151605127
SCUBE2	57758	chr11:9069731-8998511	9219731	8848511	9074731	8993511
SFRP1	6422	chr8:41286137-41238635	41436137	41088635	41291137	41233635
SLC39A6	25800	chr18:31963355-31942492	32113355	31792492	31968355	31937492
SLC40A1	30061	chr2:190153858-190133561	190303858	189983561	190158858	190128561
SLC5A6	8884	chr2:27288575-27275962	27438575	27125962	27293575	27270962
SLC7A6	9057	chr16:66855924-66893223	66705924	67043223	66850924	66898223
SLC9A3	6550	chr5:577447-526425	727447	376425	582447	521425
SPDEF	25803	chr6:34632069-34613558	34782069	34463558	34637069	34608558
STK11	6794	chr19:1156798-1179434	1006798	1329434	1151798	1184434
TAP1	6890	chr6:32929726-32920964	33079726	32770964	32934726	32915964
TCF7L1	83439	chr2:85214245-85391016	85064245	85541016	85209245	85396016
TK1	7083	chr17:73694880-73681755	73844880	73531755	73699880	73676755
TOP2A	7153	chr17:35827695-35798321	35977695	35648321	35832695	35793321
TP53	7157	chr17:7531642-7512445	7681642	7362445	7536642	7507445
TP53BP2	7159	chr1:222100297-222034218	222250297	221884218	222105297	222029218

TSPAN13	27075	chr7:16759876-16790686	16609876	16940686	16754876	16795686
TYMS	7298	chr18:647651-663492	497651	813492	642651	668492
UBE2C	11065	chr20:43874662-43879003	43724662	44029003	43869662	43884003
UBE2T	29089	chr1:200577707-200567408	200727707	200417408	200582707	200562408
UIMC1	51720	chr5:176366049-176264612	176516049	176114612	176371049	176259612
VEGFA	7422	chr6:43845931-43862202	43695931	44012202	43840931	43867202
WIPF2	147179	chr17:35629100-35691965	35479100	35841965	35624100	35696965
XBP1	7494	chr22:27526560-27520548	27676560	27370548	27531560	27515548
YBX1	4904	chr1:42920653-42940607	42770653	43090607	42915653	42945607
ZNF217	7764	chr20:51633043-51617017	51783043	51467017	51638043	51612017

Supplementary Table 3.2: SWITCHdna segment size distribution.

Histogram showing the distribution of SWITCHdna segments by size.

Supplementary Table 3.2



CHAPTER FOUR

INPP4B IS A PUTATIVE IDENTIFYING AND PROGNOSTIC MARKER OF BASAL-LIKE BREAST TUMORS

4.1 Introduction

Basal-like breast tumors have been shown to have frequent and specific copy number loss of *INPP4B* in multiple studies and low mRNA expression of this gene is seen in BBT. INPP4B has been shown to be a key component of the PI3K signaling pathway and may act as a tumor suppressor that inhibits Akt activation, proliferation, and promotes apoptosis. The association between INPP4B DNA, RNA, and protein expression, and breast cancer subtype is further examined in this study using a tumor slide dataset with associated gene expression and copy number data (UNC, n=125) and two different tissue microarrays, one with outcome data (SPECS-TMA, n=106; UBC-TMA, n=318). IHC staining was performed on the slide dataset and tissue microarrays to determine protein expression level. It was shown that low INPP4B IHC staining levels were highly associated with BBT over other tumor subtypes and also that *INPP4B* copy number states, RNA gene expression, and protein expression were all strongly associated with one another. Negative staining for cytoplasmic INPP4B was associated with significantly shorter mean relapse-free survival time on the UBC-TMA dataset. Additionally, tumors which were able to achieve pathologic complete response had significantly lower RNA expression of *INPP4B* compared to those that had residual disease. These findings indicate that

low INPP4B levels are specific to the basal-like subtype and track the known biology of BBT. INPP4B may have utility as a clinical marker for BBT, and/or as a predictor for treatment response.

The inherent heterogeneity of breast tumors shows that there are a host of oncogenes and tumor suppressors, which when dysregulated, lead to cancer-promoting processes. Previously, it has been shown that basal-like breast tumors (BBT) have a strong association with dysfunctional DNA repair pathways, illustrating one example of a defective cell pathway leading to tumor formation. However, one pathway alone is certainly insufficient to explain the development of all tumors; even within one specific subtype of breast cancer, it is not the case that all tumors of that subtype are only disrupted in a single manner.

One pathway of interest is the phosphoinositide 3-kinase (PI3K) signaling pathway, which promotes cell proliferation, survival, migration, angiogenesis, and genomic instability (Puc et al. 2005; Saal et al. 2005; Janzen and Scadden 2006; Baker 2007) (Figure 4.1). This pathway has been shown to be dysregulated in a wide variety of cancers and mutations in key members of this pathway have been observed to induce neoplastic transformations in various model systems (Vivanco and Sawyers 2002; Engelman et al. 2006).

Under normal cellular function, PI3K pathway activation occurs in response to extracellular stimuli, and increased production of the phosphoinositides, PtdIns(3,4,5)P_3 and PtdIns(3,4)P_2, occurs at the plasma membrane and leads to promotion of cell growth and cell survival pathways (Plas and Thompson 2005; Manning and Cantley 2007). These phosphoinositides bind and activate several

downstream effectors, most notably the proto-oncogene Akt, and are necessary for full Akt activation *in vivo* (Franke et al. 1997; Scheid et al. 2002; Ma et al. 2008).

Downstream signaling of the PI3K pathway is normally countered by phosphatase degradation of PtdIns(3,4,5)P_3. One such phosphatase is phosphatase and tensin homolog (PTEN), a 3-phosphotase and tumor suppressor that dephosphorylates PtdIns(3,4,5)P_3 to form PtdIns(4,5)P_2 (Ma et al. 2008). Loss of PTEN expression or function is often seen in human cancers (Vazquez and Sellers 2000) and murine models of *PTEN* deletion result in many tumor types (Di Cristofano et al. 1998; Kishimoto et al. 2003). Furthermore, reduced PTEN is also seen to be associated with BBT (Weigman et al. ; Saal et al. 2008).

Further processing of PI3K pathway phosphoinositides occur as PtdIns(4,5)P_2 is hydrolyzed by inositol polyphosphate 4-phosphatases (Norris and Majerus 1994). INPP4B is one such phosphatase and this gene has recently generated interest as it has been shown by us and others to be frequently and specifically deleted in BBT (Fedele et al. ; Weigman et al. ; Naylor et al. 2005; Bergamaschi et al. 2006; Chin et al. 2007). Loss of *INPP4B* was also frequently associated with loss of *PTEN* in tumors (Fedele et al.). Other observations also suggest an important role for *INPP4B*, such as the finding that shRNA targeting of the gene induces anchorage-independent growth in human mammary epithelial cells (HMEC) (Westbrook et al. 2005). INPP4B knockdown is also shown to promote Akt activation, increased cell proliferation and tumor growth and low INPP4B protein expression is associated with poor outcomes in breast and ovarian cancers (Gewinner et al. 2009).

Though a strong association has been shown between *INPP4B* loss and BBT, extensive analysis of *INPP4B* function within this subtype has not yet been performed. Here, we have investigated *INPP4B* expression and function within breast tumor subtypes and determined the utility of differential *INPP4B* expression as a subtype and treatment response predictor.

.

4.2 Results

In Chapter 2, we showed that in addition to being frequently and specifically lost on a copy number level in BBT, *INPP4B* also has significantly lower RNA expression when copy number loss is present (Figure 2.2) and of all the breast cancer subtypes, has the lowest expression in BBT (Figure 2.3F). Here, we analyze the protein expression of INPP4B in our breast tumors using immunohistochemistry (IHC) and proceed to draw correlations with different aspects of breast cancer biology. IHC staining was performed on a set of 125 breast tumor slides and scored according to level of INPP4B staining (0-3 scale) (Figure 4.2). The distribution of IHC scores was assessed across subtypes and we found that weak or absent INPP4B was much more likely to be found in ER-negative subtypes, and most frequently in basal-like samples, while high-intensity INPP4B staining was more prevalent in the two luminal subtypes (Table 1, p-value <.0001). Furthermore, there was excellent correlation between the level of INPP4B protein expression, represented by IHC score, and the level of gene expression. Each IHC staining level increase was associated with a commensurate increase in *INPP4B* gene expression (Figure 4.3).

Finally, the association between *INPP4B* copy number status and INPP4B protein level was examined. 60 of the 125 IHC stained breast tumors had DNA copy number array data available and the copy number state of *INPP4B* was determined for this subset, with each sample being assigned a condition of gain, loss, or unchanged. Within each of these copy number conditions, the tumors that were unchanged or gained for *INPP4B* on a DNA level were far more likely to have intermediate to strong staining for INPP4B by IHC. Conversely, the breast tumors

that were determined to have copy number loss were significantly more likely to show weak or absent INPP4B protein levels (Table 4.2).

Through collaboration with the Nielsen group at the University of British Columbia, we continued to determine the strength of the association between INPP4B levels and clinicopathologic characteristics by leveraging the use of tissue microarray (TMA) sets available at that institution. Preliminary results from analysis of one of these TMAs (SPECS) (n=106), reveals a similar strong association with BBT as was seen in the set of 125 UNC slides. Each tumor from the SPECS-TMA was classified as basal-like or non-basal and put into one of two categories based on INPP4B staining, <=5% of cells stained (low) or >5% of cells stained (high) (Table 4.3). Nearly all the non-basal samples exhibited high INPP4B staining (69/70, 98.6%), while all but one of the samples that had low INPP4B staining was a basal-like tumor (22/23, 95.7%).

The ability of *INPP4B* expression to predict survival and therapeutic response was also assessed within available datasets. A separate TMA with survival data was available to our collaborators from UBC (UBC) (n=318) and was utilized for this purpose. Here, the samples were broken out by level of cytoplasmic staining intensity and median relapse-free survival times within each group were determined. Although the magnitude of the difference is small, for those patients with negative staining for INPP4B, there was significantly shorter mean relapse-free survival time (Table 4.4).

In order to test the capability of *INPP4B* expression levels to predict pathologic complete response to neoadjuvant chemotherapy, the Hess et al. and

Hatzis et al. (Hatzis et al. ; Hess et al. 2006) combined dataset was employed. Here, the samples were split into two groups, those that achieved pathologic complete response (pCR) and those that had residual disease (RD). Within each group, the gene expression of *INPP4B* was determined for each sample, and the two groups were compared by ANOVA. The ANOVA result showed that tumors that achieved pCR had significantly lower expression of *INPP4B* than those that did not (Figure 4.4)

Lastly, one other means by which we examined the relationship between *INPP4B* and breast cancer subtypes was to examine the level of *INPP4B* expression across a panel of breast cancer mouse models. Similar to our findings on human breast tumors, we found that the basal-like tumor mouse models had the lowest mean expression for this gene, while models of other breast cancer subtypes had comparatively higher expression (Figure 4.5).

4.3 Discussion and Ongoing Studies

It has been seen that the different gene expression subtypes of breast cancer are often associated with distinct genetic alterations. By understanding which subtype-specific genetic events are occurring in what types of pathways, we can potentially identify therapeutic targets to develop and pursue based on existing understanding of pathway function. We expanded upon a previous study identifying basal-like specific regions of deletion and followed up on the finding that BBT have common loss of chromosome 4q, which contains the gene *INPP4B* (Fedele et al. ; Weigman et al. ; Gewinner et al. 2009). We previously showed that in addition to being frequently and specifically copy-number lost for *INPP4B*, BBT also had the lowest average expression of the gene. It was also seen that across all breast tumors, tumors with copy number loss of *INPP4B* had significantly lower expression, suggesting that the copy number deletions were functional (Weigman et al.). In this study, we continue the examination between subtype and *INPP4B* expression, as well as more closely examining INPP4B expression on a protein level by IHC.

Utilizing INPP4B IHC staining and scoring by the level of staining, we found that BBT was also the subtype most strongly associated with absent or low INPP4B protein levels (Table 4.1). These findings were observed not only on our UNC slide dataset, but also replicated on a separate TMA from UBC, showing that this is a robust finding (Table 4.3). This same staining analysis is currently being performed on two, different larger TMAs (n=348, n=4000) and we anticipate that similar results will be observed.

Examination of INPP4B on a DNA, RNA, and protein level showed that there was great concordance in expression between these three levels as copy number lost samples were much more likely to have low INPP4B IHC staining and the level of IHC staining followed very closely in line with the level of RNA expression (Figure 4.3, Table 4.2). The combination of these factors points to the fact that low *INPP4B* expression appears to be an excellent marker for the basal-like subtype regardless of what methodology is used to determine expression level. The strength of the association between IHC staining and BBT is particularly notable in terms of clinical utility as nucleic-acid analysis based laboratory testing is still in the minority compared to IHC staining (Netto et al. 2003).

INPP4B expression levels also appear to have some functionality in terms of predicting survival times and therapeutic response. We observed that those tumors with the lowest expression levels of this gene had the shortest relapse-free survival times (Table 4.4) and that patients who were able to achieve pCR had significantly lower expression of this gene compared to those who had residual disease (Figure 4.4). It is possible and likely that much of this is due simply to the association with the basal-like subtype, which also possesses these characteristics (Perou et al. 2000; Carey et al. 2007), but this does speak to the strength of its connection with this subtype.

When we examined *INPP4B* expression in a panel of mouse models of breast cancer, we saw that the basal-like models were the ones with the lowest expression of *INPP4B*, recapitulating the condition seen in humans (Figure 4.5). This fact is encouraging as it appears these models may potentially share the same disruptions

in the PI3K pathway and can likely be used as a surrogate for human BBT in future studies, increasing our ability to unravel the mechanisms at play.

Regarding the biology of *INPP4B* and the role it might play in basal-like formation, one potential model may tie together the defect in PI3K pathway signaling shown here along with previously discussed BBT associations with *TP53* and defective double-strand break repair. A model previously proposed is that an initiating event occurs through *TP53* disruption, through mutation or copy number loss, both of which are commonly seen in basal-like cancers. This then allows the cell to lose *BRCA1*, which would otherwise be a lethal event (Evers and Jonkers 2006), or, as we have proposed, lose other components of the BRCA1 repair pathway, eventually resulting in a BRCA1-dependent DSB repair defect. Increased genomic instability occurs as a result of improper DNA repair function, which triggers disruption of other genes, such as components of the PI3K pathway. *PTEN* is one such candidate for disruption, having been seen to be associated with BBT (Weigman et al. ; Saal et al. 2008); another potential candidate is *INPP4B*, as we have discussed here. It has been proposed that BBTs may be addicted to aberrant PI3K pathway signaling (Weinstein 2002) and this cumulative model ties together many characteristics of BBT that explain how this could come about (Weinstein 2002).

Further studies are required to more fully understand the role of *INPP4B* in BBT. As mentioned, IHC staining is currently underway on a large TMA from our UBC collaborators (n=4000) that will give an even more complete view of this protein in breast cancer. This set will also come with more robust clinical outcomes data and more associations can be drawn from this analysis, potentially including a

new prognostic signature derived from *INPP4B*. Functional studies are also underway to reconstitute INPP4B expression in basal-like tumor cell lines of both mouse and human origin. The aim of these studies is to more robustly determine the physiologic role of these genes in a cancer background and identify a transcriptional program that can be mined for a genetic signature and potential therapeutic targets.

4.4 Materials and Methods

Breast Cancer Patient Datasets

Three patient datasets were used in this study, one UNC tumor slide set (n=125) and two different TMAs from the Nielsen group, SPECS (n=106) and UBC (n=318). The UNC set contained gene expression and DNA copy number microarray data, while the UBC set had clinical outcomes data. For pCR vs. RD analyses, the previously published Hess et al. (Hess et al. 2006) and Hatzis et al. (Hatzis et al.) datasets were combined and utilized (n=588).

Survival and Response Analysis

Each patient tumor in the UBC-TMA was classified into one of four groups based on level of IHC staining (negative, low, moderate, strong). Survival analyses were performed using the Kaplan-Meier test. For association with pCR, the patients were split into those that achieved pCR and those that had residual disease and the expression of INPP4B was compared between those two groups by ANOVA test.

Tumor subtype classification

For all tumors, the Lowess normalized R/G Log_2 ratio data were median centered prior to collapsing (via averaging) from probes to HGNC gene symbols. The PAM50 gene set predictor (Parker et al. 2009) was then used to assign subtypes to the tumors in the UNC slide dataset and the UBC TMAs. For the UNC slide set, the Claudin-low predictor(Prat et al.) was also applied.

Copy Number Status Assignment

To determine the INPP4B copy number state of tumors on the UNC slide dataset, we used aCGH data from previously performed studies and applied the SWITCHdna algorithm (Weigman et al.).

Immunohistochemistry

Formalin-fixed, paraffin-embedded tissue sections (~5μm) were processed using standard immunostaining methods. Following deparaffinization in xylenes, slides were rehydrated through a graded series of alcohol and rinsed in phosphate buffered saline. Endogenous peroxidase activity was blocked with 3% hydrogen peroxidase. Samples were steamed for antigen retrieval with 10 mM citrate buffer (pH 6.0) for 30 minutes. Slides were then incubated for 20 minutes with diluted normal blocking serum. The sections were incubated for 60 minutes at room temperature with primary antibody to INPP4B (Epitomics 2512-1 1:5000). The slides were incubated for 45 minutes with diluted biotinylated secondary antibody (1:250 dilution) and 30 minutes with Vectastain Elite 102 ABC reagent (Vector Laboratories). Sections were incubated in peroxidase substrate solution for visualization. Slides were counterstained with hematoxylin and examined by light microscopy. Tumor immunoreactivity was scored 0=negative, 1=weak positive, 2=intermediate positive, and 3=strong positive.

4.5 Figures and Legends

Figure 4.1: PI3K Pathway.

The PI3K signaling pathway is illustrated. Under normal circumstances, receptor stimulation leads to the formation of $PI(3,4,5)P_3$ and $PI(3,4)P_2$ from PI3K activity. These phospholipids induce Akt activation and increased cell proliferation, survival, and migration. PI3K signaling is terminated through the activity of PTEN, which hydrolyzes $PI(3,4,5)P_3$ to $PI(4,5)P_2$, or INPP4B, which converts $PI(3,4)P_2$ to PI(3)P.

Figure 4.1

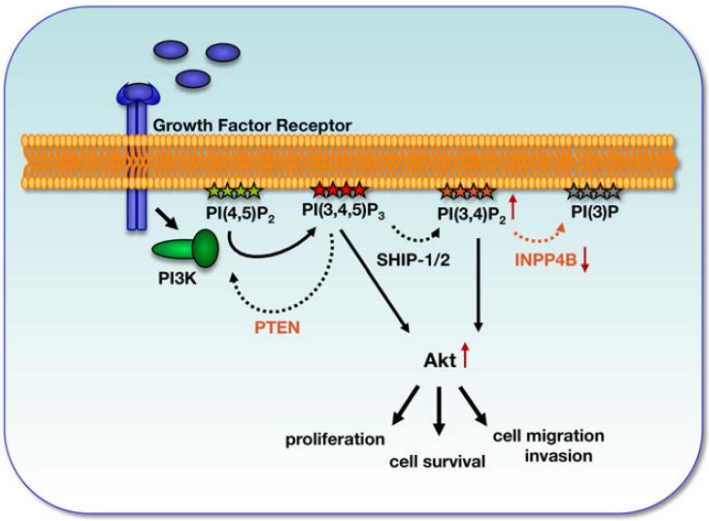


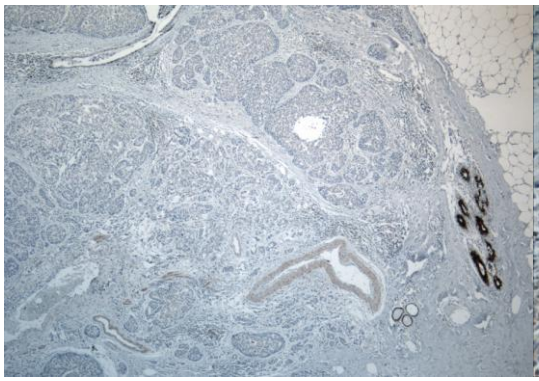
Figure adapted from Gewinner et al., *Cancer Cell* (2009)

Figure 4.2: INPP4B IHC Staining Scoring Levels.

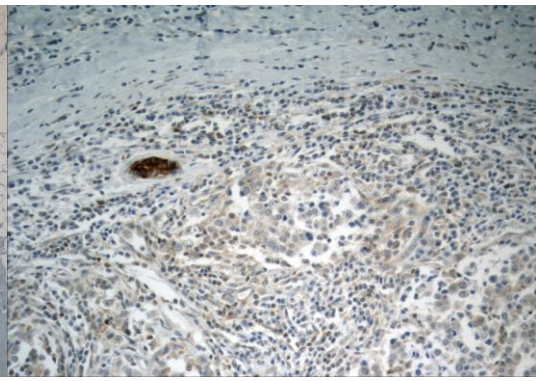
Representative images of each level of INPP4B IHC scoring is shown (IHC Score 0-3). IHC Score 0: no staining or less than 5% cells stained; IHC Score 1: weak staining; IHC Score 2: intermediate staining; IHC Score 3: strong staining.

Figure 4.2

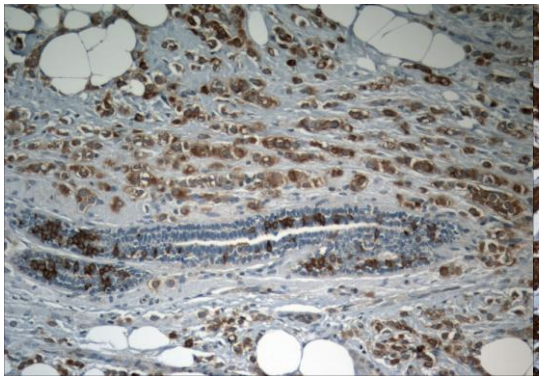
IHC Score 0



IHC Score 1



IHC Score 2



IHC Score 3

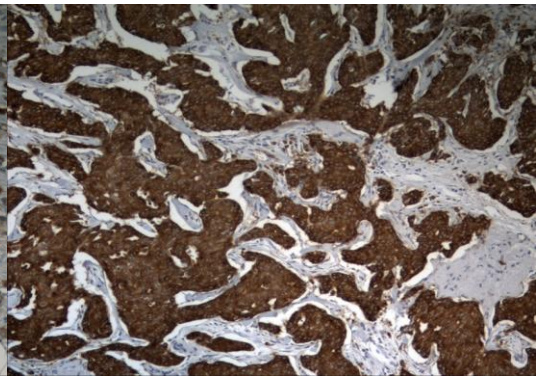


Figure 4.3: *INPP4B* Expression by INPP4B IHC Score.

Tumors from the UNC slide dataset (n=125) are grouped by their INPP4B IHC score and *INPP4B* expression is determined for each sample. p-value determined by ANOVA test.

Figure 4.3

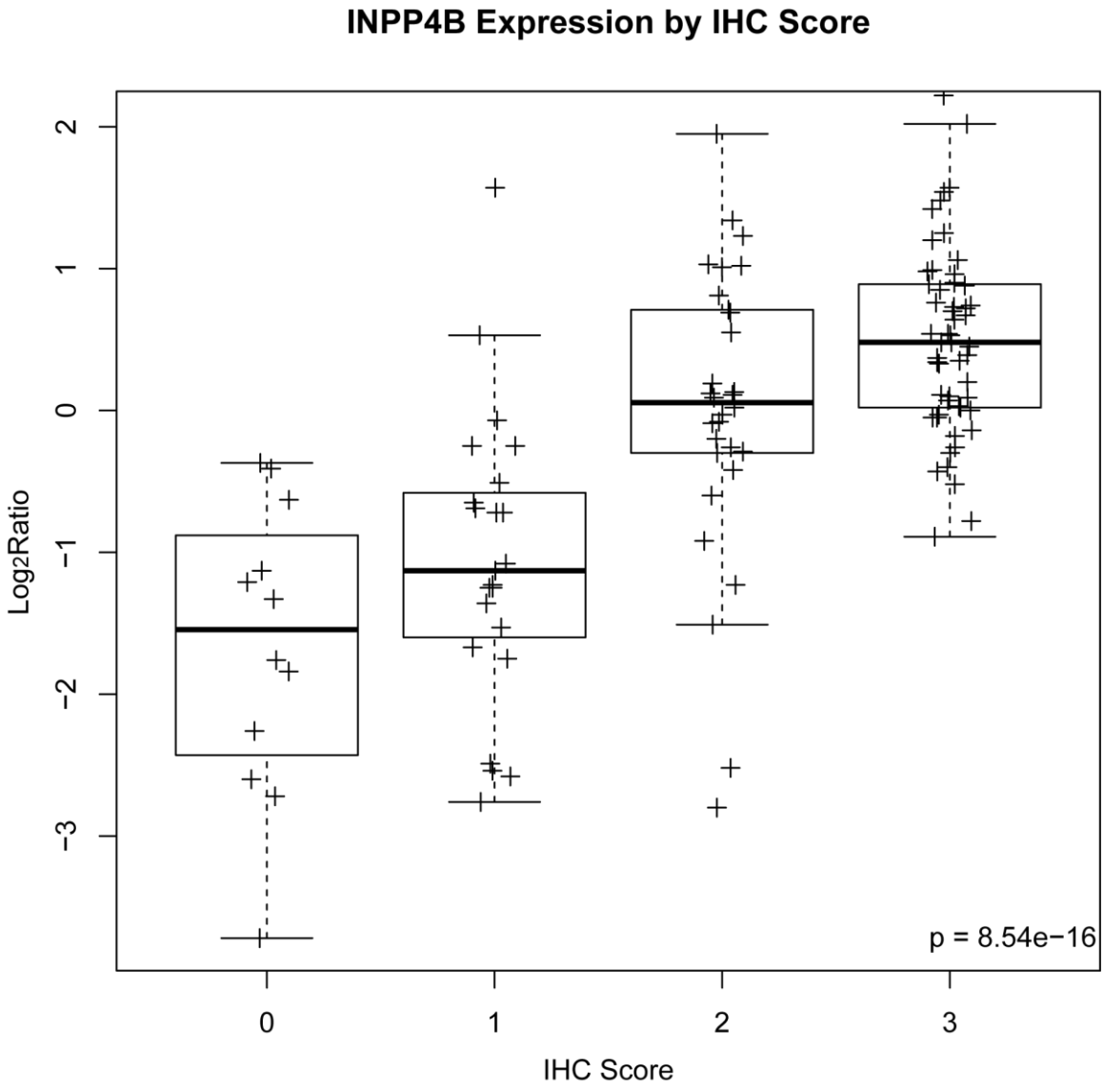

INPP4B Expression by IHC Score
Log2Ratio
2
1
0
−1
−2
−3
p = 8.54e−16
0
1
2
3
IHC Score

Figure 4.4: *INPP4B* Expression by pCR Status.

Tumor samples from the Hess et al. (Hess et al. 2006) and Hatzis et al. (Hatzis et al.) combined dataset are separated into groups by those with a pathologic complete response (pCR) or residual disease (RD) and *INPP4B* expression is evaluated for each sample within each group. p-value determined by ANOVA test.

Figure 4.4

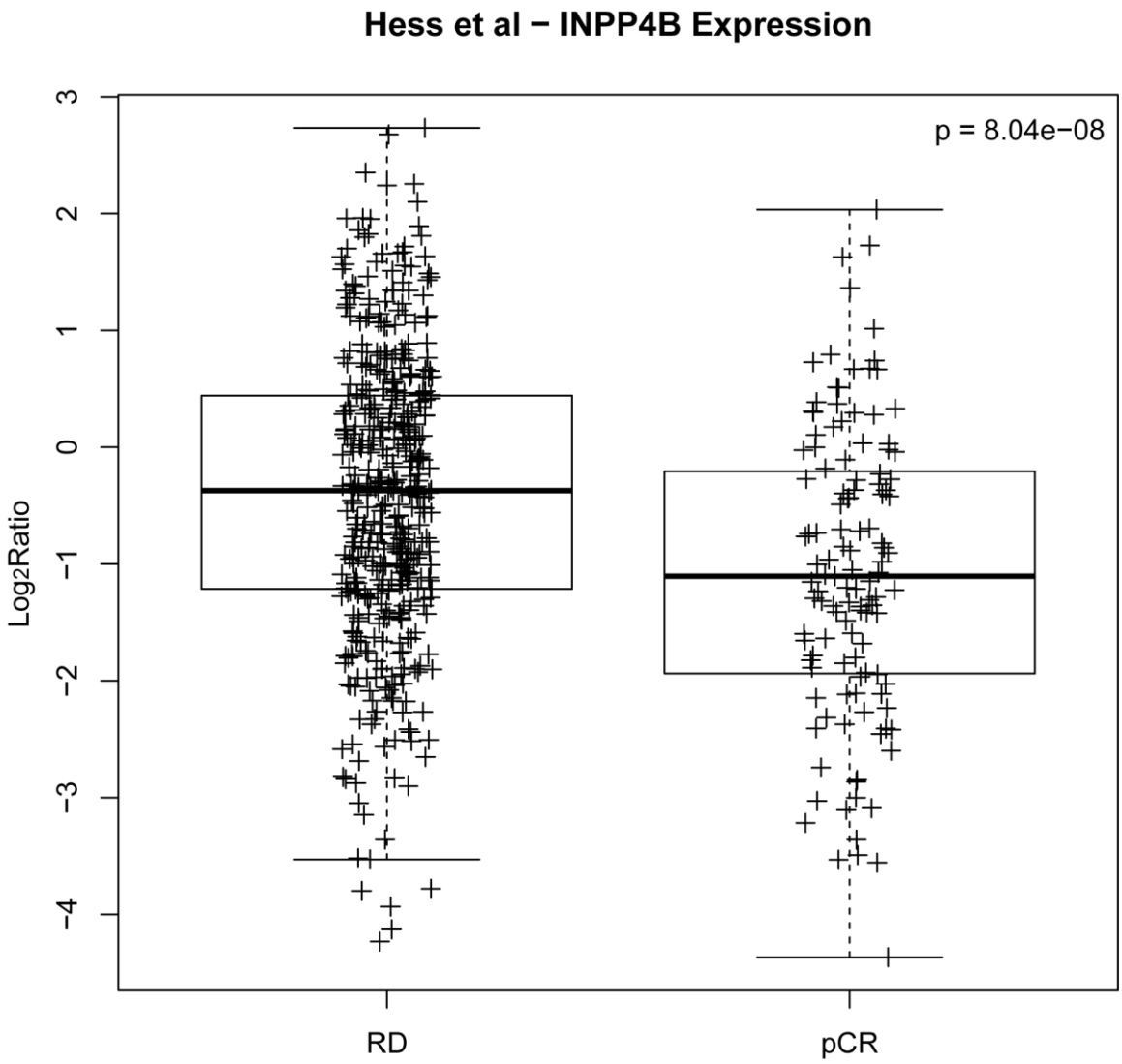

Hess et al − INPP4B Expression
p = 8.04e−08
Log2Ratio
3
2
1
0
−1
−2
−3
−4
RD
pCR

Figure 4.5: *INPP4B* Expression across a panel of murine breast cancer models.

INPP4B expression level was assessed across a panel of mouse breast cancer models. Each mouse tumor was grouped by model and *INPP4B* expression levels were assessed within each group. p-value determined by ANOVA test.

Figure 4.5

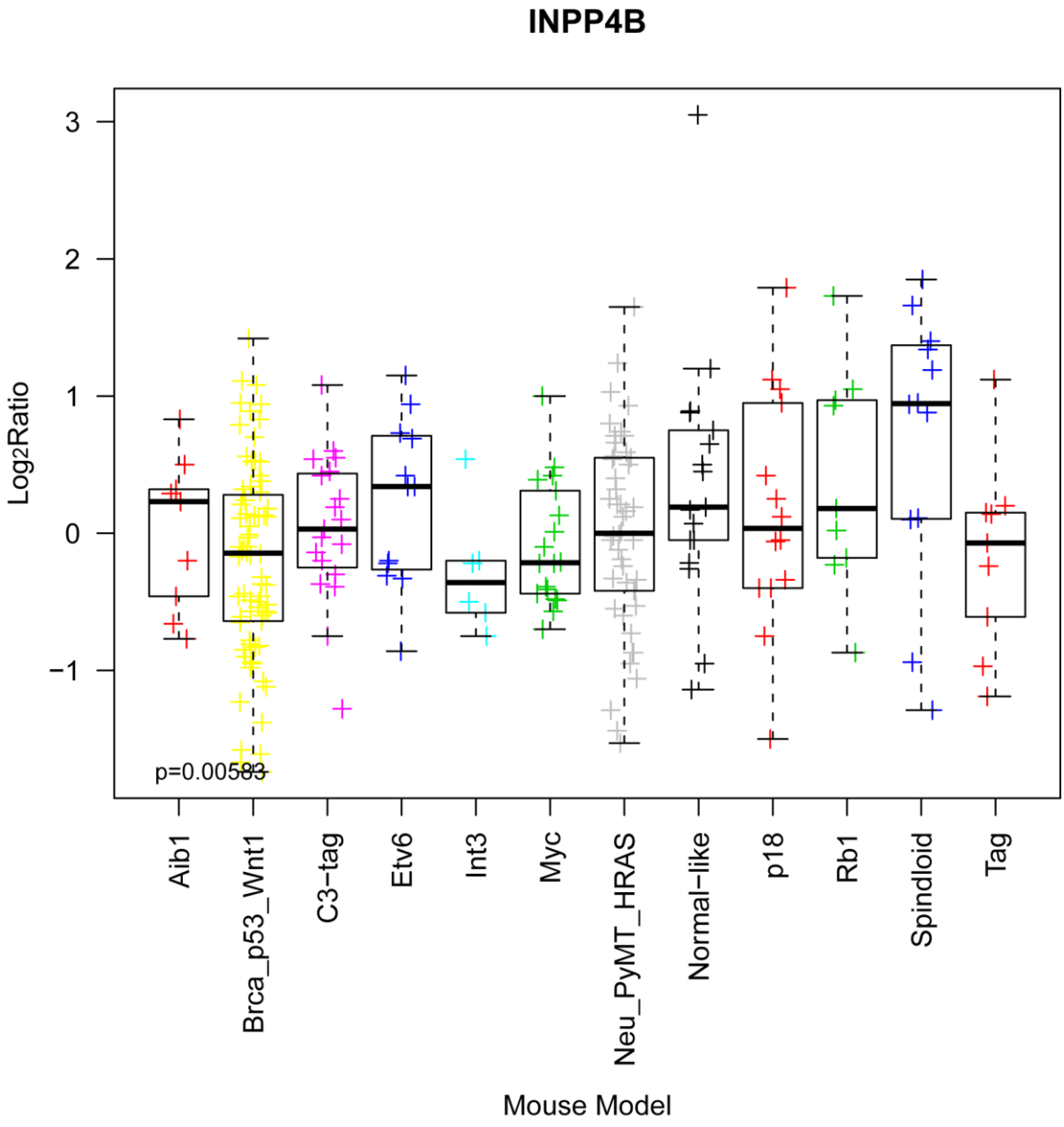

INPP4B
Log2Ratio
3
2
1
0
−1
p=0.00583
Aib1
Brca_p53_Wnt1
C3−tag
Etv6
Int3
Myc
Neu_PyMT_HRAS
Normal−like
p18
Rb1
Spindloid
Tag
Mouse Model

Table 4.1: INPP4B IHC Score by Subtype.

Each sample from the UNC slide dataset is displayed based on its classification by PAM50+Claudin subtype and IHC score. p-value determined by Fisher's Exact test.

Table 4.1

Subtype	IHC Score (0-3)				
	0	**1**	**2**	**3**	**Total**
Basal	7	12	2	4	25
Claudin	3	2	7	5	17
Her2	1	5	5	2	13
LumA	1	0	7	23	31
LumB	0	4	7	16	27
Normal	0	1	1	3	5
Unassigned	0	3	2	2	7
Total	12	27	31	55	125

p-value < .0001

Table 4.2: *INPP4B* Copy Number State by IHC Score.

For the 60 tumors on the UNC slide dataset with available copy number data, each was grouped based on its copy number status (gain, lost, unchanged) and IHC scoring level (low: 0-1, or high: 2-3). The percentage of tumors within each copy number state that had low or high IHC staining is displayed in parentheses. p-value determined by Fisher's Exact Test.

Table 4.2

	N=60					
		INPP4B 109k Status				
		Loss	No Change	Gain	Total	
INPP4B IHC Score	0-1	5 (62.5)	8 (19.5)	4 (36.4)	17	
	2+	3 (37.5)	33 (80.5)	7 (63.6)	43	
	Total	8	41	11		p-value = 0.03

Table 4.3: SPECS TMA INPP4B IHC Staining by Subtype.

The tumors on the SPECS-TMA were classified by subtype in terms of basal or not-basal and grouped according to subtype and IHC staining level (low: <=5% of tumor cells, or high: >5% of tumor cells). The percentage of tumors within each subtype with low or high IHC staining is displayed.

Table 4.3

INPP4B Basal vs. Non-Basal Crosstabulation

			Basal vs. Non-Basal		
			Non-Basal	Basal	Total
INPP4B	<=5% Tumor Cells	Count	1	22	23
		% Within Basal vs. Non-Basal	1.40%	61.10%	21.70%
	>5% Tumor Cells	Count	69	14	83
		% Within Basal vs. Non-Basal	98.60%	38.90%	78.30%
Total		Count	70	36	106
		% Within Basal vs. Non-Basal	100%	100%	100%

Table 4.4: Mean Relapse-Free Survival Time by level of cytoplasmic INPP4B intensity.

Tumors on the UBC-TMA were split into four groups based on INPP4B cytoplasmic staining intensity level (negative, low, moderate, strong) and mean+/-95% confidence interval relapse-free survival times are displayed for each group.

Table 4.4

Mean Relapse-Free Survival Time

INPP4B (cytoplasmic; intensity)		Mean			
				95% Confidence Interval	
		Estimate	Std. Error	Lower Bound	Upper Bound
	negative	7.651	.522	6.627	8.674
	low	10.292	.610	9.095	11.488
	moderate	12.624	.610	11.428	13.820
	strong	11.241	.672	9.925	12.558
	Overall	12.091	.524	11.064	13.119

CHAPTER FIVE

DISCUSSION

5.1 Summary of Findings

Since the discovery of the breast tumor intrinsic subtypes, our understanding of the molecular underpinnings of breast cancer has steadily increased and this has allowed more informed identification and treatment of the disease. To date, there have been at least six different breast cancer subtypes identified, luminal A, luminal B, normal breast-like, HER-enriched, claudin-low, and basal-like (Perou et al. 2000; Sorlie et al. 2001; Sorlie et al. 2003; Hu et al. 2006; Herschkowitz et al. 2007; Parker et al. 2009), each with biological and clinical relevance, as each subtype has unique clinical outcomes and response to treatment (Carey et al. 2006; Fan et al. 2006; Carey et al. 2007; Hoadley et al. 2007). These subtypes are defined both by the genes they express, such as ESR1 (luminal) and HER2 (HER2-enriched), and the genes they do not express, such as claudins 3, 4, and 7 (claudin-low). As technology continues to advance and new high-throughput technologies are developed and employed, the ability to study these cancers steadily improves. Already, it has been shown that breast cancers have distinct characteristics beyond expression profiles, including that they are distinct on a DNA copy number level as well (Russnes et al. ; Weigman et al. ; Jonsson et al. 2005; Bergamaschi et al. 2006; Haverty et al. 2008). There is a persistent need to further elucidate the biological basis for breast tumor

development as this disease continues to be a leading killer of women. The basal-like breast tumors, the focus of this work, are of particular importance, as they would be considered the 4th leading cancer killer of women if viewed as an individual disease.

In this work, I investigate the genomic landscape of BBT and examine basal-like specific areas of copy number aberration in order to define unique basal-like characteristics, and identify factors that contribute to its genomic instability phenotype. The main findings are summarized below:

Basal-like Breast Cancer DNA Copy Number Losses Identify Genes Involved in Genomic Instability, Response to Therapy and Patient Survival

1. Basal-like breast tumors exhibit frequent aberrations in copy number landscape, with the loss of regions that contain BRCA1-dependent DNA repair pathway genes.
2. Many DNA repair pathway genes that show copy number loss also exhibit low gene expression levels.
3. BRCA1 pathway genes located on chromosome 5q (*RAD17, RAD50, RAP80*) are frequently lost in unison and rarely individually.
4. Low expression of DNA repair genes on chromosome 5q (*RAD17, RAD50, RAP80*) is associated with poor survival.
5. Gene knockdown of RAD17 and/or RAD50 reduces the ability of the cell to form BRCA1 foci in response to DNA damage.

6. Disruption of the BRCA1 pathway by gene knockdown increased sensitivity to DNA damaging agents.

Micro-scale Genomic Copy Number Aberrations as Another Means of Mutagenesis in Breast Cancer

1. Breast tumors have a significant number of previously undetectable copy number micro-aberration events that are detectable by a high-resolution tiling copy number array.
2. Micro-aberrations are most commonly observed in basal-like breast tumors.
3. Copy number micro-aberrations can have functional consequences on the exon-level RNA expression of the aberrant gene.
4. Exon skipping is one observed functional consequence of micro-aberrations on gene expression.
5. Copy number micro-aberrations are disproportionately located in the promoter regions of genes.
6. Copy number micro-aberrations are enriched in cell-cycle related genes.
7. A high frequency of copy number micro-aberrations is associated with poor overall survival.

INPP4B is a Putative Identifying and Prognostic Marker of Basal-like Breast Tumors

1. Basal-like breast tumors have frequent copy number loss of *INPP4B*, low RNA expression of *INPP4B*, and low protein expression of INPP4B.
2. INPP4B levels are highly concordant on the DNA, RNA, and protein levels.

3. Low expression of *INPP4B* is associated with the ability to achieve a pathologic complete response, but also shorter relapse-free survival time.
4. Mouse models of basal-like breast cancer have the lowest average expression of *INPP4B* across a panel of breast cancer mouse models.
5. Low expression of INPP4B may be a biomarker of responsiveness to PIK3CA inhibitors.

5.2 Clinical Relevance

Breast Cancer Burden in the United States and North Carolina

At present, ~12% of women will develop invasive breast cancer during the course of her lifetime, with an estimated 230,480 new cases in 2011 in the United States, and another 57,650 new non-invasive cases. This represents the most commonly diagnosed form of cancer in women, excluding skin cancer, accounting for ~30% of all cancers in women. 39,520 women were expected to die from their disease in the same year. Though death rates have been decreasing since 1990, notably in women under age 50, breast cancer death rates are still higher than any other type of cancer for women, aside from lung cancer. Additionally, there are more than 2.6 million breast cancer survivors in the US (Statistics courtesty breastcancer.org).

Breast cancer in the State of North Carolina poses significant challenges to the health care system as there were 8,553 new cases in 2009 and another 8,597 new cases projected for 2011. In addition, breast cancer as a whole is a leading cancer killer of NC women and if BBT were to be treated as a unique disease separate from other breast cancers, it would represent the fourth leading cause of cancer deaths among women.

The research described here give us a greater understanding of how BBT arise. More importantly, these experiments help to highlight potential therapeutic targets, something that has been critically lacking for this subgroup so far. Currently, treatment regimens for basal-like breast tumors are limited to cytotoxic

chemotherapies, which have severe side effects; thus, continued advancements to improve both clinical outcome and treatment tolerability are still needed.

Potentially Targetable Molecular Pathways Associated with Basal-like Breast Cancer

Identifying molecular pathways that drive basal-like breast cancers, and that may also explain their development provides an avenue through which to pursue potentially druggable targets for this disease. Two such pathways that show promise as basal-like involved targets that have been described in this work are the DSB DNA repair pathway and the PI3K signaling pathway.

This research provides preclinical support for clinical trials using PARP inhibitors such as the one titled "A Phase II Study of the PARP Inhibitor, INIPARIB (BSI-201), in Combination With Chemotherapy to Treat Triple Negative Breast Cancer Brain Metastasis", which is currently being run here at UNC. Successful completion of this trial, among the first of the PARP inhibitor trials, could provide improvements in the ability to provide care to women suffering from BBT and result in significant clinical benefits. Other means of exploiting the synthetic lethality conferred by DNA repair defects could also prove promising.

The PI3K signaling pathway is also a proposed therapeutic target in a range of cancers and there are currently a number of phase I–II clinical trials in progress investigating the efficacy of PI3K inhibitors. Indeed, it has been proposed that the DSB repair defects of BBT may precipitate PI3K signaling pathway dysfunction through genetic disruption of *PTEN* (Saal et al. 2008), suggesting that therapy targeted to this pathway may be an effective means to treat some sporadic and hereditary breast cancers. The level of *INPP4B* expression may also prove useful as

in cell lines, loss of INPP4B sensitized cells to PIK3CA inhibitors. Thus, loss of INPP4B, or simply being a Basal-like tumor may be a marker in predicting the patients that will respond to PIK3CA inhibitors.

Improving Ability to Differentiate Basal-like Breast Cancer

An increase in the ability to differentiate more clearly between the known breast cancer subtypes is also clinically relevant, given what is already known about prognostic associations with each subtype. By knowing that a given cancer is of a more aggressive form, this potentially informs the clinician to pursue more aggressive therapy at the start; conversely, by knowing that a given cancer is unlikely to respond further to aggressive therapy, it allows the patient to be spared certain unwanted side effects of treatment. Clear knowledge of the type of breast cancer one has also allows the patient to be more informed of one's disease.

In this work, we identified a number of novel basal-like specific features from copy number analyses. Indeed, we find that loss of the 5q genes (*RAD17, RAD50, RAP80*) is very specific (>90%) at predicting BBT over other subtypes. When coupled with gain of chromosome 10p, it becomes even more specific for BBT, and these factors could allow the clinician to be confident in the diagnosis. *INPP4B* loss was also determined to be a basal-specific feature and with our collaborators at UBC, it was found to possess the best combination of sensitivity (61%) and specificity (99%) of any single marker they surveyed for identifying BBTs using IHC. Continued discovery of such factors and improved understanding of the differences between the subtypes helps improve diagnostic power.

5.3 Long-term Directions

Identify Other Subtype-specific Markers with New Datasets and Tools

More data regarding breast cancer is being generated daily, whether this occurs through samples collected in new datasets, or analyses being performed using new technologies. Though we identified a number of novel findings of functional significance in this work, the drive to progress certainly does not cease.

A collaborative effort that has been underway in recent years is The Cancer Genome Atlas (TCGA), which aims to fully analyze a large number of cancers on multiple levels, including full genome sequencing, mRNA-sequencing, methylation, and protein expression. One cancer type to be examined is breast cancer and already we have seen a number of the findings in this work recapitulated in this large dataset; importantly, the loss of 4q and 5q as basal-like subtype specific events was rediscovered on TCGA breast data (C.Perou, personal communications). The expanded information available through this pursuit allows us to more fully examine our findings and draw more complete conclusions that tie in information from this cancer on all levels. For example, we put forth the hypothesis that all BBT may have some form of BRCA1 pathway dysfunction, but have been limited to identifying the dysfunction only on a DNA and RNA level. The TCGA allows us to study these tumors on all levels and see other forms of inactivation to test this hypothesis which have simply not been available to us previously.

Another continued goal is to develop a greater understanding of the nature of micro-instability within breast cancers and other tumors. Though we characterized them in Chapter 3, we are as yet unable to describe them fully on a sequence level

and explain the exact nature of certain micro-amplifications. This represents an area of ongoing study and indeed the full-genome sequencing information from the TCGA may be one means to answer this question.

Continued Analysis of Basal-like Specific Targets

Aside from the basal-like specific regions of aberration discussed in this work, there are a number of other areas of both gain and loss identified for this subtype that may also contain promising candidates for study. Although in Chapter 2, we examined the patterns of loss and functional effects of three BRCA1-dependent DNA repair pathway genes located on chromosome 5q, we did in fact identify other components of this pathway in our initial analyses. A few examples are *RAD51, RAD51B*, and *RAD51C*, located on chromosome 15q15, 14q12, and 17q23 respectively. We attempted to prioritize what we believed to be the most important players, but these other candidates may also make key contributions to the genomic instability phenotype of BBT.

Another potentially promising region to further investigate is an amplified region on chromosome 10p also found to be basal-specific. This region notably contained the genes *MAP3K8, ZEB1*, and *FAM107B*. Though this region was not robustly investigated in the manner in which chromosome 5q11-35 was, the fact that it is gained instead of lost is potentially useful clinically as it is currently easier to inhibit an overactive target than replace a missing factor. MAP3K8, in particular, is a kinase that could be effectively targeted if it does turn out to be a critical driver of this particular amplicon.

Beyond the two regions highlighted here, there are still several regions of aberration that have not been extensively examined. There may still be important targets yet to be discovered that drive BBT etiology that will reveal themselves with more detailed analyses. The need and potential for continued improvements in BBT treatment means that this remains an area of active research.

Identify Importance of Other Subtype-specific Regions of Aberration

Though the focus of this work was on the basal-like subtype due to its poor clinical prognosis and paucity of treatment options, we do still face challenges in treating the other breast cancer subtypes. Even though luminal breast cancers have relatively good outcomes compared to other breast subtypes, they actually result in the highest number of total deaths due simply to the fact that most breast cancers are luminal breast cancers (O'Brien et al.). Thus, the other subtypes should not simply be ignored as there is much benefit to be derived from a greater understanding of disease processes and treatment improvements for those groups as well.

We focused on basal-like specific aberrations here, but the Luminal A, Luminal B, and Her2-enriched subtypes all displayed specific regions of aberration as well. Though initial analyses did not expose any obvious targets within these regions, useful findings may still be derived from mining this data. It would be of great utility if relevant targets could be found within every subtype that help both uniquely identify the subtype and play critical roles in their development that were exploitable for therapy.

5.4 Concluding Remarks

In summary, the work described here examines the role of genomic instability in basal-like breast tumors (BBT), primarily in the context of the BRCA1-dependent DNA repair pathway. The analyses identified several regions of aberration specific to BBT, and associated functional studies showed possible effects of these disruptions that may explain their DNA repair defect and other contributions to overall genomic instability. These findings provide insights into BBT biology and lay the groundwork for many future studies.

REFERENCES

Integrated genomic analyses of ovarian carcinoma. *Nature* **474**(7353): 609-615.
2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature* **455**(7216): 1061-1068.

Adelaide, J., Finetti, P., Bekhouche, I., Repellini, L., Geneix, J., Sircoulomb, F., Charafe-Jauffret, E., Cervera, N., Desplans, J., Parzy, D. et al. 2007. Integrated Profiling of Basal and Luminal Breast Cancers. *Cancer Res* **67**(24): 11565-11575.

Anbazhagan, R., Bartek, J., Monaghan, P., and Gusterson, B.A. 1991. Growth and development of the human infant breast. *Am J Anat* **192**(4): 407-417.

Andrews, D.W.K. 1993. Tests for Parameter Instability and Structural Change With Unknown Change Point. *Econometrica* **61**: 821-856.

Arnes, J.B., Brunet, J.S., Stefansson, I., Begin, L.R., Wong, N., Chappuis, P.O., Akslen, L.A., and Foulkes, W.D. 2005. Placental cadherin and the basal epithelial phenotype of BRCA1-related breast cancer. *Clin Cancer Res* **11**(11): 4003-4011.

Bai, J., Perron, B. 1998. Estimating and Testing Linear Models with Multiple Structural Changes. *Econometrica* **66**: 47-78.

Baker, S.J. 2007. PTEN enters the nuclear age. *Cell* **128**(1): 25-28.

Bamford, S., Dawson, E., Forbes, S., Clements, J., Pettett, R., Dogan, A., Flanagan, A., Teague, J., Futreal, P.A., Stratton, M.R. et al. 2004. The COSMIC (Catalogue of Somatic Mutations in Cancer) database and website. *Br J Cancer* **91**(2): 355-358.

Bao, S., Tibbetts, R.S., Brumbaugh, K.M., Fang, Y., Richardson, D.A., Ali, A., Chen, S.M., Abraham, R.T., and Wang, X.F. 2001. ATR/ATM-mediated phosphorylation of human Rad17 is required for genotoxic stress responses. *Nature* **411**(6840): 969-974.

Benito, M., Parker, J., Du, Q., Wu, J., Xiang, D., Perou, C.M., and Marron, J.S. 2004. Adjustment of systematic microarray data biases. *Bioinformatics* **20**(1): 105-114.

Benjamini, Y. and Hochberg, Y. 1995. Controlling the False Discovery Rate: A Practical and Powerful Approach to Multiple Testing. *Journal of the Royal Statistical Society Series B (Methodological)* **57**(1): 289-300.

Bergamaschi, A. 2006. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression

subtypes of breast cancer. *Genes, Chromosomes and Cancer* **45**(11): 1033-1040.

Bergamaschi, A., Kim, Y.H., Kwei, K.A., La Choi, Y., Bocanegra, M., Langerod, A., Han, W., Noh, D.Y., Huntsman, D.G., Jeffrey, S.S. et al. 2008. CAMK1D amplification implicated in epithelial-mesenchymal transition in basal-like breast cancer. *Mol Oncol* **2**(4): 327-339.

Bergamaschi, A., Kim, Y.H., Wang, P., Sorlie, T., Hernandez-Boussard, T., Lonning, P.E., Tibshirani, R., Borresen-Dale, A.L., and Pollack, J.R. 2006. Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. *Genes Chromosomes Cancer* **45**(11): 1033-1040.

Brodie, S.G. and Deng, C.-X. 2001. BRCA1-associated tumorigenesis: what have we learned from knockout mice? *Trends in Genetics* **17**(10): S18.

Burris, H.A., 3rd. 2001. Docetaxel (Taxotere) plus trastuzumab (Herceptin) in breast cancer. *Seminars in oncology* **28**(1 Suppl 3): 38-44.

Carey, L.A., Dees, E.C., Sawyer, L., Gatti, L., Moore, D.T., Collichio, F., Ollila, D.W., Sartor, C.I., Graham, M.L., and Perou, C.M. 2007. The Triple Negative Paradox: Primary Tumor Chemosensitivity of Breast Cancer Subtypes. *Clin Cancer Res* **13**(8): 2329-2334.

Carey, L.A., Perou, C.M., Livasy, C.A., Dressler, L.G., Cowan, D., Conway, K., Karaca, G., Troester, M.A., Tse, C.K., Edmiston, S. et al. 2006. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *Jama* **295**: 2492 - 2502.

Chang, H.R., Glaspy, J., Allison, M.A., Kass, F.C., Elashoff, R., Chung, D.U., and Gornbein, J. Differential response of triple-negative breast cancer to a docetaxel and carboplatin-based neoadjuvant treatment. *Cancer* **116**(18): 4227-4237.

Chin, K., DeVries, S., Fridlyand, J., Spellman, P.T., Roydasgupta, R., Kuo, W.L., Lapuk, A., Neve, R.M., Qian, Z., Ryder, T. et al. 2006. Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. *Cancer Cell* **10**(6): 529-541.

Chin, S.F., Teschendorff, A.E., Marioni, J.C., Wang, Y., Barbosa-Morais, N.L., Thorne, N.P., Costa, J.L., Pinder, S.E., van de Wiel, M.A., Green, A.R. et al. 2007. High-resolution aCGH and expression profiling identifies a novel genomic subtype of ER negative breast cancer. *Genome Biol* **8**(10): R215.

Cressman, V.L., Backlund, D.C., Hicks, E.M., Gowen, L.C., Godfrey, V., and Koller, B.H. 1999. Mammary tumor formation in p53- and BRCA1-deficient mice. *Cell Growth Differ* **10**(1): 1-10.

Date, O., Katsura, M., Ishida, M., Yoshihara, T., Kinomura, A., Sueda, T., and Miyagawa, K. 2006. Haploinsufficiency of RAD51B causes centrosome fragmentation and aneuploidy in human cells. *Cancer Res* **66**(12): 6018-6024.

Deng, C.-X. and Wang, R.-H. 2003. Roles of BRCA1 in DNA damage repair: a link between development and cancer. *Hum Mol Genet* **12**(suppl_1): R113-123.

Di Cristofano, A., Pesce, B., Cordon-Cardo, C., and Pandolfi, P.P. 1998. Pten is essential for embryonic development and tumour suppression. *Nat Genet* **19**(4): 348-355.

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. Genome remodelling in a basal-like breast cancer metastasis and xenograft. *Nature* **464**(7291): 999-1005.

Donawho, C.K., Luo, Y., Luo, Y., Penning, T.D., Bauch, J.L., Bouska, J.J., Bontcheva-Diaz, V.D., Cox, B.F., DeWeese, T.L., Dillehay, L.E. et al. 2007. ABT-888, an orally active poly(ADP-ribose) polymerase inhibitor that potentiates DNA-damaging agents in preclinical tumor models. *Clin Cancer Res* **13**(9): 2728-2737.

Dosanjh, M.K., Collins, D.W., Fan, W., Lennon, G.G., Albala, J.S., Shen, Z., and Schild, D. 1998. Isolation and characterization of RAD51C, a new human member of the RAD51 family of related genes. *Nucleic Acids Res* **26**(5): 1179-1184.

Engelman, J.A., Luo, J., and Cantley, L.C. 2006. The evolution of phosphatidylinositol 3-kinases as regulators of growth and metabolism. *Nat Rev Genet* **7**(8): 606-619.

Evers, B. and Jonkers, J. 2006. Mouse models of BRCA1 and BRCA2 deficiency: past lessons, current understanding and future prospects. *Oncogene* **25**(43): 5885-5897.

Fan, C., Oh, D.S., Wessels, L., Weigelt, B., Nuyten, D.S.A., Nobel, A.B., van't Veer, L.J., and Perou, C.M. 2006. Concordance among Gene-Expression-Based Predictors for Breast Cancer. *N Engl J Med* **355**(6): 560-569.

Farmer, H., McCabe, N., Lord, C.J., Tutt, A.N., Johnson, D.A., Richardson, T.B., Santarosa, M., Dillon, K.J., Hickson, I., Knights, C. et al. 2005. Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy. *Nature* **434**(7035): 917-921.

Fedele, C.G., Ooms, L.M., Ho, M., Vieusseux, J., O'Toole, S.A., Millar, E.K., Lopez-Knowles, E., Sriratana, A., Gurung, R., Baglietto, L. et al. Inositol polyphosphate 4-phosphatase II regulates PI3K/Akt signaling and is lost in human basal-like breast cancers. *Proc Natl Acad Sci U S A*.

Foulkes, W.D., Stefansson, I.M., Chappuis, P.O., Begin, L.R., Goffin, J.R., Wong, N., Trudel, M., and Akslen, L.A. 2003. Germline BRCA1 mutations and a basal epithelial phenotype in breast cancer. *J Natl Cancer Inst* **95**(19): 1482-1485.

Franke, T.F., Kaplan, D.R., Cantley, L.C., and Toker, A. 1997. Direct regulation of the Akt proto-oncogene product by phosphatidylinositol-3,4-bisphosphate. *Science* **275**(5300): 665-668.

French, C.A., Tambini, C.E., and Thacker, J. 2003. Identification of functional domains in the RAD51L2 (RAD51C) protein and its requirement for gene conversion. *J Biol Chem* **278**(46): 45445-45450.

Fridlyand, J., Snijders, A.M., Ylstra, B., Li, H., Olshen, A., Segraves, R., Dairkee, S., Tokuyasu, T., Ljung, B.M., Jain, A.N. et al. 2006. Breast tumor copy number aberration phenotypes and genomic instability. *BMC Cancer* **6**(1): 96.

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**(2): 115.

Goldenberg, M.M. 1999. Trastuzumab, a recombinant DNA-derived humanized monoclonal antibody, a novel agent for the treatment of metastatic breast cancer. *Clin Ther* **21**(2): 309-318.

Gonzalez, M.A., Tachibana, K.E., Chin, S.F., Callagy, G., Madine, M.A., Vowler, S.L., Pinder, S.E., Laskey, R.A., and Coleman, N. 2004. Geminin predicts adverse clinical outcome in breast cancer by reflecting cell-cycle progression. *J Pathol* **204**(2): 121-130.

Graeser, M., McCarthy, A., Lord, C.J., Savage, K., Hills, M., Salter, J., Orr, N., Parton, M., Smith, I.E., Reis-Filho, J.S. et al. A marker of homologous recombination predicts pathologic complete response to neoadjuvant chemotherapy in primary breast cancer. *Clin Cancer Res* **16**(24): 6159-6168.

Gruvberger, S., Ringner, M., Chen, Y., Panavally, S., Saal, L.H., Borg, A., Ferno, M., Peterson, C., and Meltzer, P.S. 2001. Estrogen receptor status in breast cancer is associated with remarkably distinct gene expression patterns. *Cancer Res* **61**(16): 5979-5984.

Gunderson, K.L., Steemers, F.J., Lee, G., Mendoza, L.G., and Chee, M.S. 2005. A genome-wide scalable SNP genotyping assay using microarray technology. *Nat Genet* **37**(5): 549-554.

Hatzis, C., Pusztai, L., Valero, V., Booser, D.J., Esserman, L., Lluch, A., Vidaurre, T., Holmes, F., Souchon, E., Wang, H. et al. A genomic predictor of response and survival following taxane-anthracycline chemotherapy for invasive breast cancer. *JAMA* **305**(18): 1873-1881.

Haverty, P.M., Fridlyand, J., Li, L., Getz, G., Beroukhim, R., Lohr, S., Wu, T.D., Cavet, G., Zhang, Z., and Chant, J. 2008. High-resolution genomic and expression analyses of copy number alterations in breast tumors. *Genes Chromosomes Cancer* **47**(6): 530-542.

Herschkowitz, J.I., He, X., Fan, C., and Perou, C.M. 2008. The functional loss of the retinoblastoma tumor suppressor is a common event in Basal-like and Luminal B breast carcinomas. *Breast Cancer Res* **10**(5): R75.

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.

Hess, K.R., Anderson, K., Symmans, W.F., Valero, V., Ibrahim, N., Mejia, J.A., Booser, D., Theriault, R.L., Buzdar, A.U., Dempsey, P.J. et al. 2006. Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer. *J Clin Oncol* **24**: 4236 - 4244.

Hillenmeyer, M.E., Fung, E., Wildenhain, J., Pierce, S.E., Hoon, S., Lee, W., Proctor, M., St Onge, R.P., Tyers, M., Koller, D. et al. 2008. The chemical genomic portrait of yeast: uncovering a phenotype for all genes. *Science* **320**(5874): 362-365.

Hoadley, K.A., Weigman, V.J., Fan, C., Sawyer, L.R., He, X., Troester, M.A., Sartor, C.I., Rieger-House, T., Bernard, P.S., Carey, L.A. et al. 2007. EGFR associated expression profiles vary with breast tumor subtype. *BMC Genomics* **8**(1): 258.

Howard, B.A. and Gusterson, B.A. 2000. Human breast development. *J Mammary Gland Biol Neoplasia* **5**(2): 119-137.

Hu, Z., Fan, C., Oh, D., Marron, J.S., He, X., Qaqish, B., Livasy, C., Carey, L., Reynolds, E., Dressler, L. et al. 2006. The molecular portraits of breast tumors are conserved across microarray platforms. *BMC Genomics* **7**(1): 96.

Huang, E., Cheng, S.H., Dressman, H., Pittman, J., Tsou, M.H., Horng, C.F., Bild, A., Iversen, E.S., Liao, M., and Chen, C.M. 2003. Gene expression predictors of breast cancer outcomes. *Lancet* **361**(9369): 1590 - 1596.

Janzen, V. and Scadden, D.T. 2006. Stem cells: good, bad and reformable. *Nature* **441**(7092): 418-419.

Jarvinen, A.K., Autio, R., Kilpinen, S., Saarela, M., Leivo, I., Grenman, R., Makitie, A.A., and Monni, O. 2008. High-resolution copy number and gene expression microarray analyses of head and neck squamous cell carcinoma cell lines of tongue and larynx. *Genes Chromosomes Cancer* **47**(6): 500-509.

Jiang, Z., Deng, T., Jones, R., Li, H., Herschkowitz, J.I., Liu, J.C., Weigman, V.J., Tsao, M.S., Lane, T.F., Perou, C.M. et al. 2010. Rb deletion in mouse mammary progenitors induces luminal-B or basal-like/EMT tumor subtypes depending on p53 status. *J Clin Invest*.

Johannsdottir, H.K., Jonsson, G., Johannesdottir, G., Agnarsson, B.A., Eerola, H., Arason, A., Heikkila, P., Egilsson, V., Olsson, H., Johannsson, O.T. et al. 2006. Chromosome 5 imbalance mapping in breast tumors from BRCA1 and BRCA2 mutation carriers and sporadic breast tumors. *Int J Cancer* **119**(5): 1052-1060.

Jonsson, G., Naylor, T.L., Vallon-Christersson, J., Staaf, J., Huang, J., Ward, M.R., Greshock, J.D., Luts, L., Olsson, H., Rahman, N. et al. 2005. Distinct Genomic Profiles in Hereditary Breast Tumors Identified by Array-Based Comparative Genomic Hybridization. *Cancer Res* **65**(17): 7612-7621.

Jonsson, G., Staaf, J., Vallon-Christersson, J., Ringner, M., Holm, K., Hegardt, C., Gunnarsson, H., Fagerholm, R., Strand, C., Agnarsson, B.A. et al. Genomic subtypes of breast cancer identified by array-comparative genomic hybridization display distinct molecular and clinical characteristics. *Breast Cancer Res* **12**(3): R42.

Kao, J., Salari, K., Bocanegra, M., Choi, Y.L., Girard, L., Gandhi, J., Kwei, K.A., Hernandez-Boussard, T., Wang, P., Gazdar, A.F. et al. 2009. Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. *PLoS One* **4**(7): e6146.

Kauraniemi, P., Hautaniemi, S., Autio, R., Astola, J., Monni, O., Elkahloun, A., and Kallioniemi, A. 2004. Effects of Herceptin treatment on global gene expression patterns in HER2-amplified and nonamplified breast cancer cell lines. *Oncogene* **23**(4): 1010-1013.

Kim, H., Chen, J., and Yu, X. 2007. Ubiquitin-binding protein RAP80 mediates BRCA1-dependent DNA damage response. *Science* **316**(5828): 1202-1205.

Kishimoto, H., Hamada, K., Saunders, M., Backman, S., Sasaki, T., Nakano, T., Mak, T.W., and Suzuki, A. 2003. Physiological functions of Pten in mouse tissues. *Cell Struct Funct* **28**(1): 11-21.

Kristensen, G.B., Kildal, W., Abeler, V.M., Kaern, J., Vergote, I., Trope, C.G., and Danielsen, H.E. 2003. Large-scale genomic instability predicts long-term outcome for women with invasive stage I ovarian cancer. *Ann Oncol* **14**(10): 1494-1500.

Liu, H.X., Cartegni, L., Zhang, M.Q., and Krainer, A.R. 2001. A mechanism for exon skipping caused by nonsense or missense mutations in BRCA1 and other genes. *Nat Genet* **27**(1): 55-58.

Loo, L.W.M., Grove, D.I., Williams, E.M., Neal, C.L., Cousens, L.A., Schubert, E.L., Holcomb, I.N., Massa, H.F., Glogovac, J., Li, C.I. et al. 2004. Array Comparative Genomic Hybridization Analysis of Genomic Alterations in Breast Cancer Subtypes. *Cancer Res* **64**(23): 8541-8549.

Ma, K., Cheung, S.M., Marshall, A.J., and Duronio, V. 2008. PI(3,4,5)P3 and PI(3,4)P2 levels correlate with PKB/akt phosphorylation at Thr308 and Ser473, respectively; PI(3,4)P2 levels determine PKB activity. *Cell Signal* **20**(4): 684-694.

Manning, B.D. and Cantley, L.C. 2007. AKT/PKB signaling: navigating downstream. *Cell* **129**(7): 1261-1274.

Matros, E., Wang, Z.C., Lodeiro, G., Miron, A., Iglehart, J.D., and Richardson, A.L. 2005. BRCA1 promoter methylation in sporadic breast tumors: relationship to gene expression profiles. *Breast Cancer Research and Treatment* **91**(2): 179.

Miller, K.A., Yoshikawa, D.M., McConnell, I.R., Clark, R., Schild, D., and Albala, J.S. 2002. RAD51C interacts with RAD51B and is central to a larger protein complex in vivo exclusive of RAD51. *J Biol Chem* **277**(10): 8406-8411.

Monaco, A.P., Bertelson, C.J., Liechti-Gallati, S., Moser, H., and Kunkel, L.M. 1988. An explanation for the phenotypic differences between patients bearing partial deletions of the DMD locus. *Genomics* **2**(1): 90-95.

Nathanson, K.L., Shugart, Y.Y., Omaruddin, R., Szabo, C., Goldgar, D., Rebbeck, T.R., and Weber, B.L. 2002. CGH-targeted linkage analysis reveals a possible BRCA1 modifier locus on chromosome 5q. *Hum Mol Genet* **11**(11): 1327-1332.

Natrajan, R., Weigelt, B., Mackay, A., Geyer, F.C., Grigoriadis, A., Tan, D.S., Jones, C., Lord, C.J., Vatcheva, R., Rodriguez-Pinilla, S.M. et al. An integrative genomic

and transcriptomic analysis reveals molecular pathways and networks regulated by copy number aberrations in basal-like, HER2 and luminal cancers. *Breast Cancer Res Treat* **121**(3): 575-589.

Naume, B., Zhao, X., Synnestvedt, M., Borgen, E., Russnes, H.G., Lingjaerde, O.C., Stromberg, M., Wiedswang, G., Kvalheim, G., Karesen, R. et al. 2007. Presence of bone marrow micrometastasis is associated with different recurrence risk within molecular subtypes of breast cancer. *Mol Oncol* **1**(2): 160-171.

Naylor, T.L., Greshock, J., Wang, Y., Colligon, T., Yu, Q.C., Clemmer, V., Zaks, T.Z., and Weber, B.L. 2005. High resolution genomic analysis of sporadic breast cancer using array-based comparative genomic hybridization. *Breast Cancer Res* **7**(6): R1186-1198.

Netto, G.J., Saad, R.D., and Dysert, P.A., 2nd. 2003. Diagnostic molecular pathology: current techniques and clinical applications, part I. *Proc (Bayl Univ Med Cent)* **16**(4): 379-383.

Nielsen, T.O., Hsu, F.D., Jensen, K., Cheang, M., Karaca, G., Hu, Z., Hernandez-Boussard, T., Livasy, C., Cowan, D., Dressler, L. et al. 2004. Immunohistochemical and Clinical Characterization of the Basal-Like Subtype of Invasive Breast Carcinoma. *Clin Cancer Res* **10**(16): 5367-5374.

Nordgard, S.H., Johansen, F.E., Alnaes, G.I., Bucher, E., Syvanen, A.C., Naume, B., Borresen-Dale, A.L., and Kristensen, V.N. 2008. Genome-wide analysis identifies 16q deletion associated with survival, molecular subtypes, mRNA expression, and germline haplotypes in breast cancer patients. *Genes Chromosomes Cancer* **47**(8): 680-696.

Norris, F.A. and Majerus, P.W. 1994. Hydrolysis of phosphatidylinositol 3,4-bisphosphate by inositol polyphosphate 4-phosphatase isolated by affinity elution chromatography. *J Biol Chem* **269**(12): 8716-8720.

O'Brien, K.M., Cole, S.R., Tse, C.K., Perou, C.M., Carey, L.A., Foulkes, W.D., Dressler, L.G., Geradts, J., and Millikan, R.C. Intrinsic breast tumor subtypes, race, and long-term survival in the Carolina Breast Cancer Study. *Clin Cancer Res* **16**(24): 6100-6110.

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

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

Perou, C.M., Jeffrey, S.S., van de Rijn, M., Rees, C.A., Eisen, M.B., Ross, D.T., Pergamenschikov, A., Williams, C.F., Zhu, S.X., and Lee, J.C. 1999. Distinctive gene expression patterns in human mammary epithelial cells and breast cancers. *Proc Natl Acad Sci U S A* **96**(16): 9212 - 9217.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., and Akslen, L.A. 2000. Molecular portraits of human breast tumours. *Nature* **406**(6797): 747 - 752.

Pinkel, D. and Albertson, D.G. 2005. Array comparative genomic hybridization and its applications in cancer. *Nat Genet* **37 Suppl**: S11-17.

Plas, D.R. and Thompson, C.B. 2005. Akt-dependent transformation: there is more to growth than just surviving. *Oncogene* **24**(50): 7435-7442.

Pollack, J.R., Perou, C.M., Alizadeh, A.A., Eisen, M.B., Pergamenschikov, A., Williams, C.F., Jeffrey, S.S., Botstein, D., and Brown, P.O. 1999. Genome-wide analysis of DNA copy-number changes using cDNA microarrays. *Nat Genet* **23**(1): 41-46.

Pollack, J.R., Sorlie, T., Perou, C.M., Rees, C.A., Jeffrey, S.S., Lonning, P.E., Tibshirani, R., Botstein, D., Borresen-Dale, A.L., and Brown, P.O. 2002. Microarray analysis reveals a major direct role of DNA copy number alteration in the transcriptional program of human breast tumors. *Proc Natl Acad Sci U S A* **99**(20): 12963-12968.

Post, S.M., Tomkinson, A.E., and Lee, E.Y. 2003. The human checkpoint Rad protein Rad17 is chromatin-associated throughout the cell cycle, localizes to DNA replication sites, and interacts with DNA polymerase epsilon. *Nucleic Acids Res* **31**(19): 5568-5575.

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

Puc, J., Keniry, M., Li, H.S., Pandita, T.K., Choudhury, A.D., Memeo, L., Mansukhani, M., Murty, V.V., Gaciong, Z., Meek, S.E. et al. 2005. Lack of PTEN sequesters CHK1 and initiates genetic instability. *Cancer Cell* **7**(2): 193-204.

Quinn, J.E., Kennedy, R.D., Mullan, P.B., Gilmore, P.M., Carty, M., Johnston, P.G., and Harkin, D.P. 2003. BRCA1 functions as a differential modulator of chemotherapy-induced apoptosis. *Cancer Res* **63**(19): 6221-6228.

R Development Core Team. 2009. R: A Language and Environment for Statistical Computing. In. R Foundation for Statistical Computing, Vienna, Austria.

Rauen, M., Burtelow, M.A., Dufault, V.M., and Karnitz, L.M. 2000. The human checkpoint protein hRad17 interacts with the PCNA-like proteins hRad1, hHus1, and hRad9. *J Biol Chem* **275**(38): 29767-29771.

Richardson, A.L., Wang, Z.C., De Nicolo, A., Lu, X., Brown, M., Miron, A., Liao, X., Iglehart, J.D., Livingston, D.M., and Ganesan, S. 2006. X chromosomal abnormalities in basal-like human breast cancer. *Cancer Cell* **9**(2): 121-132.

Roos-Mattjus, P., Vroman, B.T., Burtelow, M.A., Rauen, M., Eapen, A.K., and Karnitz, L.M. 2002. Genotoxin-induced Rad9-Hus1-Rad1 (9-1-1) chromatin association is an early checkpoint signaling event. *J Biol Chem* **277**(46): 43809-43812.

Rouzier, R., Perou, C.M., Symmans, W.F., Ibrahim, N., Cristofanilli, M., Anderson, K., Hess, K.R., Stec, J., Ayers, M., Wagner, P. et al. 2005. Breast cancer molecular subtypes respond differently to preoperative chemotherapy. *Clin Cancer Res* **11**(16): 5678-5685.

Rowley, J.D. 1973. Letter: A new consistent chromosomal abnormality in chronic myelogenous leukaemia identified by quinacrine fluorescence and Giemsa staining. *Nature* **243**(5405): 290-293.

Russnes, H.G., Vollan, H.K., Lingjaerde, O.C., Krasnitz, A., Lundin, P., Naume, B., Sorlie, T., Borgen, E., Rye, I.H., Langerod, A. et al. Genomic architecture characterizes tumor progression paths and fate in breast cancer patients. *Sci Transl Med* **2**(38): 38ra47.

Russo, J. and Russo, I.H. 2004. Development of the human breast. *Maturitas* **49**(1): 2-15.

Saal, L.H., Gruvberger-Saal, S.K., Persson, C., Lovgren, K., Jumppanen, M., Staaf, J., Jonsson, G., Pires, M.M., Maurer, M., Holm, K. et al. 2008. Recurrent gross mutations of the PTEN tumor suppressor gene in breast cancers with deficient DSB repair. *Nat Genet* **40**(1): 102-107.

Saal, L.H., Holm, K., Maurer, M., Memeo, L., Su, T., Wang, X., Yu, J.S., Malmstrom, P.O., Mansukhani, M., Enoksson, J. et al. 2005. PIK3CA mutations correlate with hormone receptors, node metastasis, and ERBB2, and are mutually exclusive with PTEN loss in human breast carcinoma. *Cancer Res* **65**(7): 2554-2559.

Sakakura, C., Mori, T., Sakabe, T., Ariyama, Y., Shinomiya, T., Date, K., Hagiwara, A., Yamaguchi, T., Takahashi, T., Nakamura, Y. et al. 1999. Gains, losses, and amplifications of genomic materials in primary gastric cancers analyzed by comparative genomic hybridization. *Genes Chromosomes Cancer* **24**(4): 299-305.

Sauer, T. 2007. Cytologic findings in malignant myoepithelioma: a case report and review of the literature. *Cytojournal* **4**: 3.

Scheid, M.P., Huber, M., Damen, J.E., Hughes, M., Kang, V., Neilsen, P., Prestwich, G.D., Krystal, G., and Duronio, V. 2002. Phosphatidylinositol (3,4,5)P3 is essential but not sufficient for protein kinase B (PKB) activation; phosphatidylinositol (3,4)P2 is required for PKB phosphorylation at Ser-473: studies using cells from SH2-containing inositol-5-phosphatase knockout mice. *J Biol Chem* **277**(11): 9027-9035.

Schorr, K., Li, M., Krajewski, S., Reed, J.C., and Furth, P.A. 1999. Bcl-2 gene family and related proteins in mammary gland involution and breast cancer. *J Mammary Gland Biol Neoplasia* **4**(2): 153-164.

Seshadri, R., Firgaira, F.A., Horsfall, D.J., McCaul, K., Setlur, V., and Kitchen, P. 1993. Clinical significance of HER-2/neu oncogene amplification in primary breast cancer. The South Australian Breast Cancer Study Group. *J Clin Oncol* **11**(10): 1936-1942.

Seshadri, R., Matthews, C., Dobrovic, A., and Horsfall, D.J. 1989. The significance of oncogene amplification in primary breast cancer. *Int J Cancer* **43**(2): 270-272.

Shen, Z. Genomic instability and cancer: an introduction. *J Mol Cell Biol* **3**(1): 1-3.

Silver, D.P., Richardson, A.L., Eklund, A.C., Wang, Z.C., Szallasi, Z., Li, Q., Juul, N., Leong, C.O., Calogrias, D., Buraimoh, A. et al. Efficacy of neoadjuvant Cisplatin in triple-negative breast cancer. *J Clin Oncol* **28**(7): 1145-1153.

Sjoblom, 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**(5797): 268-274.

Slamon, D.J., Leyland-Jones, B., Shak, S., Fuchs, H., Paton, V., Bajamonde, A., Fleming, T., Eiermann, W., Wolter, J., Pegram, M. et al. 2001. Use of chemotherapy plus a monoclonal antibody against HER2 for metastatic breast cancer that overexpresses HER2. *N Engl J Med* **344**(11): 783-792.

Sobhian, B., Shao, G., Lilli, D.R., Culhane, A.C., Moreau, L.A., Xia, B., Livingston, D.M., and Greenberg, R.A. 2007. RAP80 targets BRCA1 to specific ubiquitin structures at DNA damage sites. *Science* **316**(5828): 1198-1202.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S. et al. 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.

Proceedings of the National Academy of Sciences of the United States of America **98**(19): 10869-10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S. et al. 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. *Proc Natl Acad Sci U S A* **100**: 8418 - 8423.

Sotiriou, C., Powles, T.J., Dowsett, M., Jazaeri, A.A., Feldman, A.L., Assersohn, L., Gadisetti, C., Libutti, S.K., and Liu, E.T. 2002. Gene expression profiles derived from fine needle aspiration correlate with response to systemic chemotherapy in breast cancer. *Breast Cancer Res* **4**(3): R3.

Starita, L.M. and Parvin, J.D. 2003. The multiple nuclear functions of BRCA1: transcription, ubiquitination and DNA repair. *Current Opinion in Cell Biology* **15**(3): 345.

Strange, R., Metcalfe, T., Thackray, L., and Dang, M. 2001. Apoptosis in normal and neoplastic mammary gland development. *Microsc Res Tech* **52**(2): 171-181.

Sun, W., Wright, F.A., Tang, Z., Nordgard, S.H., Van Loo, P., Yu, T., Kristensen, V.N., and Perou, C.M. 2009. Integrated study of copy number states and genotype calls using high-density SNP arrays. *Nucleic Acids Res* **37**(16): 5365-5377.

Takata, M., Sasaki, M.S., Sonoda, E., Fukushima, T., Morrison, C., Albala, J.S., Swagemakers, S.M., Kanaar, R., Thompson, L.H., and Takeda, S. 2000. The Rad51 paralog Rad51B promotes homologous recombinational repair. *Mol Cell Biol* **20**(17): 6476-6482.

Tran, L.M., Zhang, B., Zhang, Z., Zhang, C., Xie, T., Lamb, J.R., Dai, H., Schadt, E.E., and Zhu, J. Inferring causal genomic alterations in breast cancer using gene expression data. *BMC Syst Biol* **5**: 121.

Troester, M., Herschkowitz, J., Oh, D., He, X., Hoadley, K., Barbier, C., and Perou, C. 2006. Gene expression patterns associated with p53 status in breast cancer. *BMC Cancer* **6**(1): 276.

Troester, M.A., Hoadley, K.A., Sorlie, T., Herbert, B.S., Borresen-Dale, A.L., Lonning, P.E., Shay, J.W., Kaufmann, W.K., and Perou, C.M. 2004. Cell-type-specific responses to chemotherapeutics in breast cancer. *Cancer Res* **64**: 4218 - 4226.

Trope, C., Kaern, J., Hogberg, T., Abeler, V., Hagen, B., Kristensen, G., Onsrud, M., Pettersen, E., Rosenberg, P., Sandvei, R. et al. 2000. Randomized study on adjuvant chemotherapy in stage I high-risk ovarian cancer with evaluation of DNA-ploidy as prognostic instrument. *Ann Oncol* **11**(3): 281-288.

Turner, N.C., Reis-Filho, J.S., Russell, A.M., Springall, R.J., Ryder, K., Steele, D., Savage, K., Gillett, C.E., Schmitt, F.C., Ashworth, A. et al. 2006. BRCA1 dysfunction in sporadic basal-like breast cancer. *Oncogene* **26**(14): 2126.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., and Witteveen, A.T. 2002. Gene expression profiling predicts clinical outcome of breast cancer. *Nature* **415**(6871): 530 - 536.

van de Vijver, M.J., He, Y.D., van't Veer, L.J., Dai, H., Hart, A.A., Voskuil, D.W., Schreiber, G.J., Peterse, J.L., Roberts, C., and Marton, M.J. 2002. A gene-expression signature as a predictor of survival in breast cancer. *N Engl J Med* **347**(25): 1999 - 2009.

Van Ewijk, P.H. and Hoekstra, J.A. 1993. Calculation of the EC50 and its confidence interval when subtoxic stimulus is present. *Ecotoxicol Environ Saf* **25**(1): 25-32.

Van Loo, P., Nordgard, S.H., Lingjaerde, O.C., Russnes, H.G., Rye, I.H., Sun, W., Weigman, V.J., Marynen, P., Zetterberg, A., Naume, B. et al. 2010. Allele-specific copy number analysis of tumors. *Proc Natl Acad Sci U S A*.

Vazquez, F. and Sellers, W.R. 2000. The PTEN tumor suppressor protein: an antagonist of phosphoinositide 3-kinase signaling. *Biochim Biophys Acta* **1470**(1): M21-35.

Venkitaraman, A.R. 2002. Cancer Susceptibility and the Functions of BRCA1 and BRCA2. **108**(2): 171.

Villadsen, R. 2005. In search of a stem cell hierarchy in the human breast and its relevance to breast cancer evolution. *APMIS* **113**(11-12): 903-921.

Villadsen, R., Fridriksdottir, A.J., Ronnov-Jessen, L., Gudjonsson, T., Rank, F., LaBarge, M.A., Bissell, M.J., and Petersen, O.W. 2007. Evidence for a stem cell hierarchy in the adult human breast. *J Cell Biol* **177**(1): 87-101.

Vivanco, I. and Sawyers, C.L. 2002. The phosphatidylinositol 3-Kinase AKT pathway in human cancer. *Nat Rev Cancer* **2**(7): 489-501.

Wang, B., Matsuoka, S., Ballif, B.A., Zhang, D., Smogorzewska, A., Gygi, S.P., and Elledge, S.J. 2007. Abraxas and RAP80 form a BRCA1 protein complex required for the DNA damage response. *Science* **316**(5828): 1194-1198.

Wang, Y., Cortez, D., Yazdi, P., Neff, N., Elledge, S.J., and Qin, J. 2000. BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes Dev* **14**(8): 927-939.

Weigman, V.J., Chao, H.H., Shabalin, A.A., He, X., Parker, J.S., Nordgard, S.H., Grushko, T., Huo, D., Nwachukwu, C., Nobel, A. et al. Basal-like Breast cancer DNA copy number losses identify genes involved in genomic instability, response to therapy, and patient survival. *Breast Cancer Res Treat*.

Weinstein, I.B. 2002. Cancer. Addiction to oncogenes--the Achilles heal of cancer. *Science* **297**(5578): 63-64.

Westbrook, T.F., Martin, E.S., Schlabach, M.R., Leng, Y., Liang, A.C., Feng, B., Zhao, J.J., Roberts, T.M., Mandel, G., Hannon, G.J. et al. 2005. A genetic screen for candidate tumor suppressors identifies REST. *Cell* **121**(6): 837-848.

Whitfield, M.L., Sherlock, G., Saldanha, A.J., Murray, J.I., Ball, C.A., Alexander, K.E., Matese, J.C., Perou, C.M., Hurt, M.M., and Brown, P.O. 2002. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell* **13**(6): 1977 - 2000.

Wiseman, B.S. and Werb, Z. 2002. Stromal effects on mammary gland development and breast cancer. *Science* **296**(5570): 1046-1049.

Wood, L.D., Parsons, D.W., Jones, S., Lin, J., Sjoblom, T., Leary, R.J., Shen, D., Boca, S.M., Barber, T., Ptak, J. et al. 2007. The Genomic Landscapes of Human Breast and Colorectal Cancers. *Science*.

Xing, F., Persaud, Y., Pratilas, C.A., Taylor, B.S., Janakiraman, M., She, Q.B., Gallardo, H., Liu, C., Merghoub, T., Hefter, B. et al. Concurrent loss of the PTEN and RB1 tumor suppressors attenuates RAF dependence in melanomas harboring (V600E)BRAF. *Oncogene* **31**(4): 446-457.

Zhong, Q., Chen, C.F., Li, S., Chen, Y., Wang, C.C., Xiao, J., Chen, P.L., Sharp, Z.D., and Lee, W.H. 1999. Association of BRCA1 with the hRad50-hMre11-p95 complex and the DNA damage response. *Science* **285**(5428): 747-750.