

5-2015

INVESTIGATION OF GENETIC ALTERATIONS IN EMT SUPPRESSOR, DEAR1, THROUGH PAN-CANCER ANALYSIS AND ULTRA-DEEP TARGETED SEQUENCING IN DUCTAL CARCINOMA IN SITU

Jacquelyn Reuther

Follow this and additional works at: http://digitalcommons.library.tmc.edu/utgsbs_dissertations

 Part of the [Computational Biology Commons](#), [Genetics Commons](#), [Genomics Commons](#), and the [Medicine and Health Sciences Commons](#)

Recommended Citation

Reuther, Jacquelyn, "INVESTIGATION OF GENETIC ALTERATIONS IN EMT SUPPRESSOR, DEAR1, THROUGH PAN-CANCER ANALYSIS AND ULTRA-DEEP TARGETED SEQUENCING IN DUCTAL CARCINOMA IN SITU" (2015). *UT GSBS Dissertations and Theses (Open Access)*. Paper 577.

This Dissertation (PhD) is brought to you for free and open access by the Graduate School of Biomedical Sciences at DigitalCommons@The Texas Medical Center. It has been accepted for inclusion in UT GSBS Dissertations and Theses (Open Access) by an authorized administrator of DigitalCommons@The Texas Medical Center. For more information, please contact laurel.sanders@library.tmc.edu.

INVESTIGATION OF GENETIC ALTERATIONS IN EMT SUPPRESSOR, *DEARI*, THROUGH PAN-
CANCER ANALYSIS AND ULTRA-DEEP TARGETED SEQUENCING IN *DUCTAL CARCINOMA*

IN SITU

A

DISSERTATION

Presented to the Faculty of

The University of Texas

Health Science Center at Houston

and

The University of Texas

MD Anderson Cancer Center

Graduate School of Biomedical Sciences

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

by

Jacquelyn Reuther, B.S.

Houston, Texas

May 2015

Acknowledgements

First and foremost, I want to thank God- for “with man, this is impossible but with God, all things are possible” (Matthew 19:26).

All of this work would not have been possible without the guidance from Dr. Ann Killary and Dr. Steven Lott. Both of you have given me such wonderful mentoring in both science and life-lessons that will benefit me far beyond the reaches of the lab bench. I would not have wanted my time at graduate school to have been any different. I have grown so much during this time, both scientifically as well as in person, to which the two of you deserve a lot of credit. I can never repay you for what you have given me during this time. The words “Thank you” do not seem enough, but they will have to do.

Many other have played vital roles in this accomplishment. I am truly grateful for the support that I have received throughout my time in graduate school, both emotionally and academically:

Members of my Supervisory Committee: Dr. Timothy McDonnell, Dr. Swathi Arur, Dr. Subrata Sen, Dr. Khandan Keyomarsi, as well as Dr. Steven Lott and Dr. Killary.

My parents Gloria and Randy Reuther and my brother and sister-in-law Mitchell and Traci Reuther- Thank you for always being there for me and supporting me emotionally, financially, spiritually, and physically. You have been there for me in everything: during the hard times, you have all been there to pick me up and during the happy times, you were all there to celebrate with me. I couldn't have asked for a better family. I am truly blessed to have you all in my life!

My colleagues and lab mates: Nanyue Chen, Seetharaman Balasenthil, and Mimi (Uyen) Le- Thank you for constant advice, guidance, and mentoring. I appreciate the many times that you all have helped me within the lab. I have enjoyed the time and conversations we have had together.

INVESTIGATION OF GENETIC ALTERATIONS IN EMT SUPPRESSOR, *DEAR1*, THROUGH
PAN-CANCER DATABASE ANALYSIS AND ULTRA-DEEP TARGETED SEQUENCING IN
DUCTAL CARCINOMA IN SITU

Jacquelyn Nicole Reuther, Ph.D.

Advisory Professor: Ann McNeill Killary, Ph.D.

Ductal carcinoma in situ (DCIS) is thought to be the earliest pre-invasive form of and non-obligate precursor to invasive ductal carcinoma (IDC). There is an urgent need to identify predictive and prognostic biomarkers for breast cancers with a heightened risk of progression from DCIS to IDC. Our laboratory has previously discovered a novel TRIM family member, *DEAR1* (*Ductal Epithelium Associated Ring Chromosome 1, annotated as TRIM62*) within chromosome 1p35.1, that is mutated and homozygously deleted in breast cancer and whose expression is downregulated/lost in DCIS. Previous work has shown that *DEAR1* is a novel tumor suppressor that acts as a dominant regulator of polarity, tissue architecture, and TGF β -driven epithelial-mesenchymal transition (EMT)^{1,2}. Herein, I have shown by pan-cancer database analysis that chromosomal loss of *DEAR1* is a moderately frequent event in multiple epithelial cancers and that targeted deletion of *Dear1* in the mouse recapitulates the tumor spectrum of human tumor types undergoing *DEAR1* copy number losses, including mammary tumors. Therefore, results indicate the relevance of the *Dear1* mouse model to human disease and suggest that genomic alteration of *DEAR1* could play a role in the etiology of multiple cancers, including breast cancer. Because *DEAR1* is downregulated in DCIS and regulates polarity and EMT, I hypothesized *DEAR1* mutations might be driver events in the progression of DCIS to IDC. Therefore I completed targeted ultra-deep sequencing of *DEAR1* in FFPE samples of 17 Pure DCIS and 17 DCIS samples associated with invasive lesions. Deep sequencing of DCIS indicate *DEAR1* is mutated in 71% of DCIS. Within these samples, multiple mutations within *DEAR1*, including exonic variants previously discovered in IDC and novel nonsense mutations were discovered and validated. Interestingly, variants in samples of DCIS associated with an invasive component indicate few variants shared between the two

components, possibly supporting an independent yet parallel evolution between DCIS and IDC. Further, functional screens were performed on a subset of mutations identified and demonstrated that indicated missense mutations can affect DEAR1's regulation of tissue architecture and TGF β signaling. Altogether, this data suggests that genomic alteration of *DEAR1* is an important mechanism for its loss of function and may be of significance in early breast cancer.

Table of Contents

List of Illustrations	x
List of Tables	xi
1. CHAPTER 1: INTRODUCTION	
1.1. Role of Genetic Alterations in the Progression of Cancer	Pg.11
1.1.1. Cancer as a Genetic Disease	Pg.11
1.1.2. Chromosomal Alterations Can Be Integral Players in Tumor Progression	Pg.15
1.1.3. The Cancer Genome Era Defines the Landscape of Cancer	Pg.17
1.1.4. Genetic Heterogeneity of Tumors Can Drive Tumor Progression and Can Have Major Implications in Therapeutic and Survival Outcomes	Pg.19
1.2. The Proteasome Pathway is integral to the maintenance of cellular homeostasis	Pg.23
1.3. The Role of the TRIM Family Proteins as Tumor Suppressors in Cancer	Pg.24
1.3.1. The TRIM Family of E3 Ligases Play Central Roles in Cell Cycle Regulation and Maintenance of Genomic Integrity	Pg.25
1.3.2. Regulation of Apoptosis by TRIM Family Members Can Promote Tumor Suppression	Pg.27
1.3.3. TRIM Proteins' Modulate Pro-Tumorigenic Inflammatory Response Driven by NFκB	Pg.28
1.3.4. TRIM Proteins' Regulation of Differentiation and Migration are Integral to Tumor Suppression	Pg.30
1.3.5. Tumor Suppressor and TRIM Family Protein DEAR1 (TRIM62) is a Vital Regulator of Polarity and EMT, Whose Loss is Correlated with Important Clinical Parameters	Pg.32

1.4. Investigation of Genomic Alterations in <i>DEAR1</i> (TRIM62) Using Pan-Cancer Analysis and Ultra-Deep Targeted Sequencing in Ductal Carcinoma In Situ (DCIS)	Pg.37
2. CHAPTER 2: IDENTIFICATION OF <i>DEAR1</i> ALTERATIONS ACROSS CANCERS USING PAN-CANCER DATABASE ANALYSIS	
2.1. Introduction	Pg.39
2.2. Methods	Pg.42
2.3. Results	
2.3.1. <i>DEAR1</i> Displays Chromosomal Loss and mRNA Downregulation in Humans in a Similar Tumor Spectrum as Tumors Found in <i>DEAR1</i> Knockout Mice	Pg.42
2.3.2. <i>DEAR1</i> Exhibits Rare Mutation in Multiple Epithelial Cancers	Pg.48
2.4. Discussion	Pg.50
3. CHAPTER 3: DEVELOPMENT AND PERFORMANCE EVALUATION OF A <i>DEAR1</i> ULTRA DEEP TARGETED SEQUENCING ASSAY FOR ION TORRENT NEXT GENERATION SEQUENCING PLATFORM	
3.1. Introduction	
3.1.1. Next Generation Sequencing and Precision Medicine in Oncology Care	Pg.56
3.1.2. Sensitivity of NGS Platforms is Essential for the Identification of Clinically Important Rare Variants	Pg.57
3.2. Methods	
3.2.1. Design of Custom <i>DEAR1</i> Targeted Ampliseq Panel	Pg.58
3.2.2. Creation of Ampliseq Spike-In for determination of Accuracy	Pg.58
3.2.3. Ampliseq Library Construction and Sequencing	Pg.61
3.2.4. Sequence Analysis	Pg.63
3.2.5. Digital PCR	Pg.64
3.2.6. Read Count Accuracy	Pg.64
3.2.7. Sensitivity and Specificity	Pg.65
3.3. Results	

3.3.1. Determination of Appropriate NGS Platform for Experimental Procedure	Pg.66
3.3.2. Design of Ion Torrent Custom <i>DEARI</i> Targeted Ampliseq Panel	Pg.68
3.3.3. Control Artificial Spike-in	Pg.68
3.3.4. Determination of Read Count Accuracy	Pg.70
3.3.5. Sensitivity and Specificity of the Custom <i>DEARI</i> Ampliseq Panel	Pg.76
3.4. Discussion	Pg.79
4. CHAPTER 4:CHARACTERIZING <i>DEARI</i> 'S GENETIC ROLE IN DUCTAL CARCINOMA IN SITU PROGRESSION	
4.1. Introduction	
4.1.1. <i>Ductal Carcinoma In Situ</i> and its Progression to Invasive Disease	Pg.86
4.1.2. <i>DEARI</i> is an Important Regulator of Polarity and EMT	Pg.88
4.2. Methods	
4.2.1. Human Specimens Collection	Pg.89
4.2.2. DNA Extraction and Quantification	Pg.91
4.2.3. Ampliseq Library Construction and Sequencing	Pg.93
4.2.4. Sequence Analysis	Pg.95
4.2.5. Digital PCR	Pg.96
4.2.6. Variant Functional Studies	Pg.97
4.3. Results	
4.3.1. Ultra-Deep Targeted Sequencing Reveals High Frequency of Alteration of <i>DEARI</i> in DCIS	Pg.98
4.3.2. Sequencing of Pure DCIS Indicated the Presence of <i>DEARI</i> Exonic and Regulatory Variants	Pg.102
4.3.3. Spectrum of Variants Found in DCIS with Invasive Components is More Complex than Pure DCIS and Share Relatively Few Variants between the Adjacent <i>in Situ</i> and Invasive Lesions	Pg.113
4.3.4. <i>DEARI</i> Exhibits Functional Mutations in DCIS as Shown by Functional Assays	Pg.115

4.4. Discussion	Pg.117
5. CHAPTER 5: DISCUSSION	
5.1. Introduction	Pg.128
5.2. Discussion	Pg.129
5.3. Future Directions	Pg.135
6. Appendix	
6.1. Spike-in Plasmid Sequence	Pg.139
6.2. R Code to Extract Barcoded Reads from Unmapped BAM Files	Pg.140-141
6.3. Differential Testing of Various Variant Filtering Stringencies	Pg.142
6.4. R Code to Calculate Coverage By Amplicon	Pg.143-146
6.5. R Code to Graph Digital PCR Data	Pg.147
6.6. R Code for Determination of Sensitivity and Specificity	Pg.148
6.7. Catalogue of Clinical Samples Sequenced	Pg.149
6.8. R Code for Creation of Venn Diagram Using the VennDiagram R Package	Pg.150
6.9. R Code for Variant Count Per Sample Boxplots	Pg.151
6.10. Full List of All Variants Harbored Within the 48kb Locus Encompassing <i>DEAR1</i> Found in Ductal Carcinoma <i>in Situ</i>	Pg.152-184
6.11. Detailed List of <i>DEAR1</i> Variants Shared Between the <i>In Situ</i> and Adjacent Invasive Components of Ductal Carcinoma <i>in Situ</i>	Pg.185- 188
7. BIBLIOGRAPHY	Pg.189-238
8. VITA	Pg.239

List of Illustrations

Figure 1- <i>DEAR1</i> LOH in Various Cancer Cell Lines as Shown By CONAN-Copy Number Analysis (Sanger Institute) Software.	Pg.44
Figure 2- <i>DEAR1</i> Heterozygous Loss and <i>SNAI2</i> Alterations Can Predict Overall Survival.	Pg.46
Figure 3-Illustration of the Spike-in Plasmid.	Pg.59
Figure 4- Illustration of Custom <i>DEAR1</i> Ampliseq Design.	Pg.69
Figure 5- Comparison of the Artificial Spike-In Frequency to Observed Allele Frequency in Next Generation Sequencing.	Pg.71
Figure 6- Coverage of the <i>DEAR1</i> Ampliseq Panel Amplicons Within the <i>DEAR1</i> Promoter and Exonic Regions.	Pg.75
Figure 7- Variant Counts Per Sample in <i>DEAR1</i> in DCIS.	Pg.100-101
Figure 8- Distribution of Variants in <i>DEAR1</i> .	Pg.103
Figure 9- Validation of <i>DEAR1</i> Variants Found Through Ultra-Deep Targeted Next Generation Sequencing by Digital PCR.	Pg.110
Figure 10- <i>DEAR1</i> Experiences Exonic Variants in DCIS of Which Few are Shared Within Lesion Components.	Pg.111
Figure 11- Mutation of <i>DEAR1</i> Affects TGF- β and SMAD3 Signal Transduction.	Pg.118
Figure 12- <i>DEAR1</i> Variants Can Effect Acinar Morphology in SKBR3 in 3D Culture.	Pg.119
Figure 13- Model of <i>DEAR1</i> Variants Based on DCIS Progression Models	Pg.126

List of Tables

Table 1- <i>DEAR1</i> Exhibits Heterozygous Loss in Multiple Human Tumor Types.	Pg.45
Table 2- <i>DEAR1</i> 's Expression is Downregulated in Multiple Tumor Types.	Pg.47
Table 3- <i>DEAR1</i> Undergoes Rare Mutation in Multiple Tumor Types.	Pg.49
Table 4- Correlation between <i>DEAR1</i> Mutation and Clinical Outcome.	Pg.51
Table 5- Accuracy of Sequencing Data as Determined by Comparison to Predicted Read Counts	Pg.73
Table 6- Sensitivity and Specificity of Custom <i>DEAR1</i> Ampliseq Panel.	Pg.77-78
Table 7- Patient Characteristics of Sampled Pure DCIS and DCIS with IDC Components.	Pg.99
Table 8- Full List of <i>DEAR1</i> Variants in Pure DCIS FFPE Samples Within Exonic and Regulatory Regions.	Pg.104
Table 9- Full List of <i>DEAR1</i> Variants in DCIS FFPE Samples with Adjacent Invasive Components Within Exonic and Regulatory Regions.	Pg.105-107
Table 10- List of High Variant Frequency (>10%) <i>DEAR1</i> Variants in Exonic and Regulatory Regions.	Pg.108-109

Chapter 1

Introduction

Role of Genetic Alterations in the Progression of Cancer

Cancer as a Genetic Disease

Cancer is in essence a genetic disease with genomic alterations being vital to the initiation and progression of cancer. The development of cancer is dependent on the accumulation of mutations and copy number changes across multiple years or decades, inducing changes which allow normal tissue to slowly progress to neoplastic disease, and in later stages to advanced metastatic disease¹. Genetic alterations, which occur during tumor development and progression, can happen in either genes promoting tumorigenesis (oncogenes) or in genes that act to suppress cancer development and its progression (tumor suppressor genes). Oncogenes were first discovered through work with retroviruses. Peyton Rous described the first transforming retrovirus, Rous Sarcoma Virus (RSV), which caused the formation of sarcomas in chickens³. Subsequent molecular studies showed that the gene responsible for RSV's transforming activity was v-Src, which was later identified as an integral oncogene in a variety of human solid tumors^{4,5}. This finding inspired the discovery of multiple other transforming genes, termed oncogenes by Huebner and Todaro, through gene transfer assays into NIH3T3 cells⁶⁻⁸. Oncogenes differ from their normal cellular counterparts through the acquisition of alterations such as point mutations, gene reduplication/amplification, or by translocation. These changes can alter the activity of an oncogene (point mutations/translocations) or increase its expression (gene reduplication/amplification)⁹. Oncogenes typically fall into four categories: growth factors and their receptors, signal transducers, transcription factors, and others, including anti-apoptotic cell death regulators, oncogenic miRNAs, as well as the newly described oncogenic role that members of the glycolytic pathway can play⁹. Genes which suppress oncogenic development and activity are called tumor suppressor genes. Tumor suppressor genes typically fall into five different categories: DNA repair genes, cell cycle regulators, inhibitors of growth factor receptors, pro-apoptotic genes, and intracellular proteins of which many

inhibit pro-tumorigenic cell signaling⁹. The concept of tumor suppressors came from the observation that non-tumorigenic clones resulted from the fusion of normal cells with tumor cells by somatic cell hybridization^{10, 11}. The normal cells were hypothesized to be contributing genes which could suppress tumorigenesis and bring the cellular state back to homeostasis¹¹⁻¹⁷. The first tumor suppressor gene identified was Retinoblastoma (Rb)^{18, 19}. Rb was found to be mutated not only in the childhood eye cancer called by the same name, but also in lung, breast, esophageal, prostate and renal carcinomas, as well as sarcomas and leukemias²⁰. Rb has been shown to be important in cell cycle regulation and in the suppression of tumorigenesis²⁰. Deregulation of the Rb pathway has been described as a quintessential mechanism in the acquisition of the insensitivity to anti-growth signals, a known hallmark of tumorigenesis²¹. After the identification of Rb, the second tumor suppressor identified was p53. p53 is known to play major roles in cellular stress mitigation through its regulation of the cell cycle, senescence, and apoptosis^{22, 23}. Interestingly, p53 was originally identified as an oncogene, as the gene was first cloned from a cancer cell harboring a dominant negative mutation²⁴. Later, studies using wildtype p53 revealed the tumor suppressive activities of the gene. p53 is mutated in up to 50% of all cancers and is also associated with the familial cancer syndrome Li Fraumeni, which has a very diverse tumor spectrum, including sarcomas, breast and brain cancers, as well as leukemias^{22, 25}. The identification of Rb and p53 reflected major strides in understanding cellular mechanisms to prevent the onset of tumorigenesis. The formative work previously accomplished within the past century has not only informed of the general processes involved in cancer but has laid the groundwork for the current and future discoveries related to the complex nature of cancer genetics.

Genes which undergo genetic alterations in cancer have been historically classified in one of three groups: caretakers, gatekeepers, and landscaper genes^{26, 27}. Gatekeepers regulate the balance between cellular growth and death. This group is known to contain both tumor suppressor genes and oncogenes. One example of a gatekeeper gene is Retinoblastoma (Rb) which is vital to the regulation of the cell cycle. Caretakers are maintainers of genetic integrity with many members involved in the regulation of DNA damage response, such as the mismatch repair genes. Inactivation of caretaker genes

indirectly promotes tumorigenesis as loss of these genes can induce genetic instability leading to an increased mutation rate, causing further deregulation of pathways involved in cancer²⁷. Another category, landscaper genes, are involved in processes related to the microenvironment conditions, especially in stromal-epithelial interactions²⁶. Changes in these genes, like TGF β , can upset homeostasis within the microenvironment and produce conditions conducive for tumor progression. The accumulation of mutations within each of these classes of genes are essential for cancer's initiation and progression²⁷. Besides these classifications, genetic mutations in cancer can also be classified as being either passenger and driver mutations²⁸. Driver mutations are within genes which are known to be integral to growth processes and whose alterations can become fixed within the tumor population due to their strong nature in promoting tumor growth. Adult sporadic tumors have been found to contain, on average, alterations in 3-6 driver genes¹. The overall effect of these driver mutations, however, is estimated to be small, on the order of a 0.4% increase in cellular growth over cell death for each gene mutated¹. Passenger mutations were originally defined as mutations which do not confer a selective advantage for the tumor and may only be fixed within the tumor populations through their co-presence with other strong driver mutations, essentially hitchhiking with these drivers to become fixed during tumor progression. This paradigm, however, is starting to change as research has shown that the accumulation of passenger mutations with moderate deleterious nature, whom by themselves cannot drive cancer progression, may cooperate with other moderately deleterious passenger mutations to produce an overall damaging effect²⁸. These deleterious "passenger" mutations have a higher probability of becoming fixed within the tumor population than their less deleterious counterparts and the accumulation of these variants can help to prolong progression leading for the ability of the tumor to acquire many more mutations. Moreover, some of these deleterious "passenger" mutations may in their own right act as drivers during certain times during tumor progression or act synergistically with the alteration of certain other pathways in order to drive tumorigenesis²⁹. For examples, landscaper genes, in synergism with other alterations, can drive cancer progression, as they help to make the environment more permissible for tumor progression despite not necessarily affecting tumor growth themselves. No matter what category particular cancer

genes exist in, tumor progression relies on the combination and synergism of multiple alterations to provide a selective advantage to the tumor and drive its progression.

Tumor suppressors genes can fit into any of the previously mentioned classic categories of cancer genes and in many respects, inactivation of these genes can act as strong drivers of tumor initiation and progression. Tumor suppressor genes can be inactivated in multiple ways, such as through mutations and deletions. Often in cancer, these alterations occur through defective DNA repair processes (mutations) or by aberrant mitotic recombination and nondisjunction (deletions). A major development in the understanding of the genetic mechanisms of tumorigenesis came through the study of retinoblastoma. The observation that sporadic forms of retinoblastoma developed later in life than familial forms led to the development of the groundbreaking definition of the “two-hit hypothesis” by Alfred Knudson³⁰. The “two-hit” model described a mechanism for tumor initiation in which two alleles are required to be inactivated, whether by mutation or chromosomal loss, for the initiation of tumorigenesis. In familial cancer, one of the inactivating hits is inherited in the germline while the second hit occurs somatically; whereas for sporadic cancer, both inactivating hits are acquired somatically. The classical definition of tumor suppressors, defined by the “two hit” model, has been verified for many tumor suppressors genes. In fact, a hallmark of tumor suppressor genes is the identification of deletions or mutations that completely knock-out the function of the gene, such as homozygous deletions and nonsense or frameshift mutations³¹. Of these, one of the more common mechanisms of tumor suppressor inactivation is through chromosomal deletion which has been shown to be important in tumorigenesis^{32,33}. However, as homozygous deletion is rare, often a combination of chromosomal deletion and gene mutation is needed in order to inactivate the tumor suppressor. For example, when the deletion of a wildtype allele occurs in a concomitant fashion with mutation of the other allele, the event is known as loss of heterozygosity (LOH)⁹. As chromosomal deletions are a common event in cancer, LOH can be critical to the inactivation of classical tumor suppressors. Most tumor suppressor genes have been shown to follow the classic “two-hit” model, yet another class of genes have been found to exist in which the alteration of only one allele is sufficient to inactivate the

tumor suppressor^{34, 35}. This condition is known as “haploinsufficiency”^{34, 35}. Classic examples of haploinsufficient genes are Arf, p27^{kip1}, and PTEN³⁶. It has been hypothesized that strong acting haploinsufficient genes may be more frequently inactivated than classical tumor suppressors because they only require the alteration of a single allele to become inactivated³⁷. These haploinsufficient genes, through their higher propensity to become inactivated, have been proposed to contribute to an increased rate of tumor development and progression³⁷. Nonetheless, no matter the mechanism of inactivation, the loss of tumor suppressor genes is critical to the onset of cancer and the progression of the disease.

Multiple studies have shown the clinical importance of inactivating mutations and chromosomal alterations in tumor suppressors to cancer progression. For example, heterozygous or homozygous loss of *Pten* causes greatly reduced survival in a prostate cancer mouse model³⁷. Dramatic reductions in survival as well as promotion of tumor progression have also been found to be associated with LKB1 inactivation via mutation or chromosomal deletions in cervical cancer patients³⁸. Moreover, it's been shown that amplification of the chromosomal location containing the Androgen Receptor (AR) gene, occurring in about 30% of prostate cancers, can be acquired during androgen deprivation therapy, effectively allowing for the continuation of tumor growth in the presence of low androgen conditions³⁹. Particular alterations can also be associated with the acquisition of other certain genetic deficiencies in cancer as is typified by carriers of germline BRCA1 or BRCA2 mutations being associated with multiple specific chromosomal alterations in breast cancers⁴⁰. Further, BRCA2 mutations are correlated with increased stage/grade of prostate cancer as well as reduced survival rates⁴¹. Lastly, matrix metalloproteinase (MMP) family member, MMP8, which is a landscaper gene important microenvironment regulation through its proteinase activities within the extracellular matrix, has been shown to be a tumor suppressor in melanoma and loss of function mutations within this gene can promote melanoma metastasis *in vivo*⁴². It is clear that genetic alterations in various types of pathways can influence the progression of tumorigenesis and subsequent survival rates.

Chromosomal Alterations Can Be Integral Players in Tumor Progression

As mentioned previously, chromosomal alterations can act as drivers in tumor progression and are common across the majority of tumor types⁴³. Cancer genomes experience a multitude of different types of chromosome alterations including aneuploidy, amplifications, double minutes, deletions, inversions, and translocations³³. Structural changes often result in the gain or loss of specific genomic loci³³. Further, these structural changes may also be balanced, resulting from either equal exchange of genetic material, or neutral, in which a deletion or amplification of chromosomal material is followed by an equal but opposing structural change³³. Across cancers, the average number of chromosomal alterations can vary greatly, with some tumors known to contain dozens of different alterations while other cancers may exhibit only a few³³. Moreover, the spectrum of chromosomal alterations can vary between inherited and sporadic tumors of the same tissue origin. For example, hereditary non-polyposis colorectal cancer is associated with a general diploid genome yet sporadic colorectal cancer often tends to exhibit aneuploidy³³. As the areas encompassing recurrent karyotypic abnormalities can demarcate loci encompassing genes important for tumor progression, chromosomal alterations have often been used to identify new oncogenes and tumor suppressor genes in both solid and hematological tumors⁴⁴. The first chromosome abnormality associated with cancer was the Philadelphia chromosome translocation t(9;22) in Chronic Myelogenous Leukemia (CML), in which a portion of chromosome 9 is translocated onto chromosome 22^{45,36}. This translocation creates the oncogenic fusion of genes Bcr and Abl. Imatinib, a specific inhibitor for the Bcr-Abl gene fusion, was one of the first successful targeted therapies created⁴⁶. The description of the Philadelphia chromosome in CML provided the first support that chromosomal aberrations can be tumor initiating events. Since this discovery, many other oncogenes and tumor suppressors have been identified that occur at or near chromosomal translocations and within areas found to be amplified or deleted in cancer. Chromosomal alterations can be focal, involving one to two genes, or they can affect large areas of the chromosome, with loss of entire chromosome arms being more common than focal alterations in cancer⁴³. The limited number of genes involved in focal alterations can allow for easier identification of genes important for tumor progression, whereas larger chromosomal aberrations often involve tens to hundreds of genes, which make it hard to determine

which particular gene or genes are driving tumor progression. However, recent research has suggested that alteration of multiple genes involved in large chromosomal alterations can potentially act synergistically to cause changes in the expression of multiple driver genes³³. In support of this, it has been shown that the loss of a single copy of a chromosome locus containing multiple tumor suppressor genes can confer a selective advantage to the tumor^{47, 48}. Large chromosomal alterations can occur either by the alteration of a single unified region or through the newly discovered mutational pattern of chromosome shattering, termed “chromothripsis”⁴⁹. Chromothripsis is thought to occur via aberrations in chromosomal segregation, exposure to ionizing radiation, or micronuclear fragmentation^{50, 51}. Chromosome shattering can contribute to rapid tumor evolution and has been shown to be associated with poor prognosis in acute myeloid leukemia (AML), multiple myeloma, and melanoma^{49, 51}. These large alterations, whether through loss or gain of a single chromosomal section or through chromosomal shattering, can be integral for tumor progression by their ability to effect gene expression of both oncogenes, such as ERBB and MYC, as well as in tumor suppressors, including PTEN and CDKN2A⁹. Changes in gene expression and inactivation of important tumor suppressors through these chromosomal alterations can increase oncogenic potential by promoting cellular changes consistent with the hallmarks of cancer⁵². As indicated, chromosomal alterations can play major roles in the deregulation of important genetic drivers in cancer and in the promotion of tumor progression.

The Cancer Genome Era Defines the Landscape of Cancer

As genetic aberrations are critical to cancer initiation and progression, more emphasis during this current era of genome sequencing has been placed on fully characterizing the genetic nature of cancer. Large consortium sequencing projects, such as The Cancer Genome Atlas (TCGA) and the International Cancer Genome Consortium (ICGC), have become prominent within the scientific community. These projects have been largely successful in their identification of novel cancer related genes and subsequent clinical biomarkers, as well as the establishment of new cancer subtypes based on molecular and genetic data⁵³. Analysis of the large genomic data sets that have come out of these projects have also revealed the mutational frequencies and the prevalence of somatic mutations across human

tumor types. Pediatric tumors have been shown to experience the lowest mutation frequency, where as environmentally driven tumors, such as lung cancer and melanoma, are associated with the highest mutation rate⁵⁴. Pan-cancer analysis has shown that some cancers unexpectedly share common alterations across tumor types. For example, p53 mutations, the most altered gene in cancer, have been found to be a driver of high-grade serous ovarian, serous endometrial, and basal breast cancer^{53, 55}. PIK3CA is the second most common gene mutated in cancer with an incidental frequency greater than 10% in most cancers⁵⁵. HER2 (ERBB2) mutations and amplification, which are known driver events in breast cancer and for which a targeted therapy has already been developed for, has also been found to occur in glioblastoma multiforme, gastric, serous endometrial, bladder and lung cancer⁵³. The determination that targetable alterations can occur across tumor types increases the range of patients able to benefit from a targeted drug originally developed for another tumor type. Pan cancer sequencing analysis has also shown that genes which are significantly mutated in cancer fall into 20 different pathways or functional groups, with the largest categories being transcription factors/regulators, histone modifiers, genome integrity, receptor tyrosine kinase signaling, and cell cycle, as well as, signaling pathways mitogen-activated protein kinases (MAPK), phosphatidylinositol-3-OH kinase (PI(3)K), transforming growth factor β (TGF β), and Wnt/ β -catenin⁵⁵. Of these, significantly mutated gene categorized within the histone modifiers, PI(3)K signaling and genome integrity categories are found to be mutated across cancer types. Further, mutational signatures have also been described that occur across cancer types as well. Most cancer types have been found to exhibit 2-6 mutational signatures, with each individual tumor displaying more than one signature⁵⁴. For example, mutations characterized by C>T transitions at NpCpG trinucleotides have been linked to increased cancer age of onset and was found to be almost ubiquitous in all cancer types⁵⁴. Moreover, some mutational signatures are associated with transcription coupled repair which tend to be enriched in C>T and C>A mutations in environmentally driven cancers such as lung and head & neck cancers that are linked to smoking, as well as UV driven melanoma⁵⁴. Another mutation signature has been shown to be caused by defective DNA repair or overly active DNA editing by members of the APOBEC/AID family which are associated with mutation signatures characterized by either small deletions and C>T substitutions linked to mismatch repair

deficiencies or C>T/C>G substitutions which are connected to the cytosine deaminase activity of APOPECs⁵⁴. A surprising find which has emerged from the sequencing efforts by the large consortiums is the “long tailed” nature of cancer mutations, in which few repetitively altered driver genes are found to be mutated and implicated in cancer⁵⁶. In accordance, the cancer genome landscape has been described as a collection of “mountains” and “hills”⁵⁷. With “mountains” defining genes indicated to have a high frequency of alterations in cancer and “hills” describing genes which are known to experience only few mutations. As shown by whole genome sequencing of many cancers, rare mutation “hills” dominate the cancer genome landscape. However, through aggregation of the data across tumor types by pan-cancer analysis, these rare alterations have been able to be implicated as drivers as well⁵³. Sequencing of large cohorts of cancer samples have shown that most solid tumors, though they are estimated to be driven only by 2-6 dominant driver mutations, also carry, on average, mutations in 33-66 genes that are predicted to be deleterious^{1,55}. Ninety-five percent of these mutations are single base pair substitutions, with 91% of these substitutions being missense mutations¹. Only 8% of the substitutions typically result in nonsense mutations, causing an inappropriate and early stop codon within the protein coding region. In all, these large studies have revealed the striking patterns seen across cancer, with unexpected cancers sharing similar genetic alterations and signatures, as well as the dominance of mutational “hills” across the landscape of cancer.

Genetic Heterogeneity of Tumors Can Drive Tumor Progression and Can Have Major Clinical Implications in Therapeutic and Survival Outcomes

One of the more surprising aspects to come out of the genome sequencing era, concerning cancer, is the degree of heterogeneity that exists in cancer. Four types of heterogeneity have been reported: intertumoral, intratumoral, intermetastatic, and intrametastatic¹. Intertumoral heterogeneity describes the heterogeneity that exists among patients¹. Most alterations within an individual’s tumor are distinct from alterations occurring within other individuals of the same cancer⁵³. Those few mutations that are typically shared amongst patients tend to be in the significantly mutated genes that are classified as “mountains” within cancer. Intratumoral heterogeneity is defined as the heterogeneity that occurs

within an individual's primary tumor¹. Intratumor heterogeneity can exist in both a spatial and temporal manner and can develop as a result of newly acquired mutations that occur as the tumor progresses, with distinct and private mutations occurring in small subpopulations termed subclones. These subclones can be intermingled with each other or reside in distinct regions of the tumor⁵¹. Evidence for intratumor heterogeneity has been established in multiple tumor types including renal cell carcinoma, glioblastoma, and breast and pancreatic cancer^{51, 58}. It has been noted that by the time of diagnosis, the vast majority of tumors consist of multiple different clones⁵⁹. The degree of heterogeneity amongst these subclones can be rather extensive. A seminal paper by Gerlinger et al. found that only 34% of mutations within renal cell carcinoma patients were present throughout the entire tumor⁵⁸. Further, Kandath and colleagues determined that 35% of breast cancers and 44% of endometrial cancers contain the presence of subclones, though this is most likely an underestimate as only coding mutations were used in this particular study⁵⁵. Moreover, there are times in which the intratumor heterogeneity is so profound that a single regional biopsy of tumor may be more similar to another patient than to biopsies of other regions within the same tumor⁵¹. The predominance of a particular individual subclone is dependent on principals similar to those found in evolutionary ecology, with the subclone exhibiting the greatest fitness establishing dominance⁶⁰. In respect to tumors, these principals establish that the dominant subclone will have acquired substantial phenotypic advantage compared to its surrounding subclones within a specific temporal and environmental context, and thus have outgrown the other tumor subpopulations⁵¹. The degree of heterogeneity as well as which individual subclone is predominant at a particular time, changes throughout the course of the disease⁵¹. Multiple theories have been proposed to account for the maintenance of intratumor heterogeneity throughout tumor progression. One theory suggests that each particular subclone exists due to its particular fitness within a given spatial niche in the tumor, whereby physical separation boundaries such as vasculature or radial distance across the tumor can serve to induce its own version of "allosteric speciation"⁵¹. Another theory has suggested that cancer stem cells can propagate heterogeneity as extensive heterogeneity has been reported in purported cancer stem cell populations²⁹. Intratumor heterogeneity has been suggested to provide the seeds which may drive intermetastatic heterogeneity.

Metastatic lesions can exhibit heterogeneity in regards to the primary tumor, to other metastatic lesions within the same patient, as well as heterogeneity within the individual metastatic lesion¹. Many times the primary and metastatic lesions show, to some degree, genetic similarities with additional mutations developing in the metastasis as well as enrichment of variant allele frequencies of shared mutations within the metastasis compared to the primary tumor. For example, ESR1 mutations, which can promote endocrine treatment resistance, have been found to be enriched, as assayed by variant allele frequencies, in metastatic lesions compared to the primary tumor⁶¹. Further, evidence has shown that, in many cases, the subclone which gave rise to the metastatic lesion often existed as a small population within the primary²⁹. Intermetastatic heterogeneity describes the heterogeneity that occurs between multiple metastatic lesions occurring in an individual patient. The presence of multiple metastatic lesions in advanced cancer patients tends to be quite common¹. It is highly plausible that each metastatic lesion is founded by different subclonal cells with very dissimilar genetic alterations. In this case, curing the patient by chemotherapeutic agents would be impossible¹. Individual metastatic lesions can also contain heterogeneity, which is called intrametastatic heterogeneity. Intrametastatic heterogeneity is similar to that which occurs in the primary tumor and, as such, develops similarly as intratumoral heterogeneity does. This is because the metastatic lesion is also under evolutionary pressure in its new environmental contexts. The acquired mutations which transpire after the seeding of the metastasis provide for the development of drug resistance¹. The large degree to which heterogeneity exists across cancers, across patients, and within individual tumors have been found to play a large part in tumor progression, in adverse outcomes, as well as in treatment resistance in patients⁶². Subclonal mutations within genes, such as p53 mutations, have been found to be able to predict adverse clinical outcome in chronic lymphocytic leukemia and myelodysplastic syndromes⁶³⁻⁶⁵. Maley and colleagues found that the number of clones harbored within premalignant Barrett's esophagus lesions were able to predict relative risk to progression to esophageal adenocarcinoma⁶⁶. Moreover, analysis completed retrospectively found progression of Ductal Carcinoma *In Situ* of the breast to invasive ductal carcinoma was associated with the presence of subclones harboring specific alterations, such as amplification of MYC, CCND1, and FGFR1⁶⁷. Further, intratumor heterogeneity has been identified as one of the key

factors in primary drug resistance and relapse⁶⁷. The degree of complexity as well as the presence of rare drug resistant mutations existing within the subclonal fractions of the tumor has been associated with the duration of treatment response and cancer progression^{59, 60, 62}. Ding and colleagues showed that the majority of relapse tumors of acute myeloid leukemia (AML) resulted from the selection of a minor subclone harbored within the primary tumor⁶⁸. Current clinical therapies have been found to often be a major driver in the clonal selection of subclones which harbor drug resistant mutations⁶⁰. Further, it has been suggested that subclones with drug resistant mutations may not always drive clonal outgrowth of the particular subclone harboring the mutation but rather may act via paracrine signaling to promote the resistance of the bulk tumor to the treatment⁵⁹. Better assessment and understanding of the nature of these subclones as well as their relationship to clinical outcomes and treatment will be able to help in the future to possibly circumvent drug resistance⁵⁹.

In accordance with alterations modulating tumor progression, a better understanding of the mutational landscape of cancer via cancer genome sequencing has been found to have major clinical implications by their ability to identify recurrent genetic alterations which can serve as prognostic or predictive biomarkers, or as therapeutic targets themselves⁶⁹. One example is the development of Herceptin, a monoclonal antibody against the HER2 receptor that is relatively quite effective in patients experiencing the recurrent chromosomal amplification of the HER2 locus in breast cancer, especially in combination with radiation⁷⁰. Early results of targeted sequencing trials have revealed that an estimated 40-60% of tumors contain at least one genetic alteration which may influence treatment decision making⁵⁶. However, most targeted therapies against select genetic alterations have had varied response rates or tumors have shown a resurgence after initial successful response. This has led to multiple studies focusing on the discovery of mutations which have the ability to alter or predict sensitivity to particular targeted therapies⁷¹. Moreover, selective pressures like therapeutic treatments, as previously discussed, can cause tumors to acquire or select for mutations during the course of the treatment which can promote resistance to the targeted therapies and recurrence of the disease⁷⁰. To this end, multiple sequencing efforts have begun to define the extensive nature of intra-tumor heterogeneity, as previously discussed,

which can have major implications for “precision” techniques focusing on targeted therapies, as the presence of sub-clones can moderate the degree of therapeutic efficacy or potentially can confer drug resistance and tumor relapse^{56, 58}. It is of utmost importance for us to understand the acquisition of genetic alterations across various cancers, and at all tumor stages, to help with stratification of patients for prognosis and therapy, for the creation of new targeted therapies, and to create new strategies to treat and prevent tumor resurgence.

The Proteasome Pathway is Integral to the Maintenance of Cellular Homeostasis

Cellular homeostasis requires strict regulation of intracellular regulatory molecules, as altered protein levels of these molecules are common in pathological conditions. The ubiquitin-mediated proteolytic pathway is one mechanism of strict control of cellular short-lived regulatory molecules and is important to many cellular functions critical to normal cell physiology, including DNA repair, protein quality regulation, cell morphogenesis, signaling pathway modulation, and cell cycle control⁷². Central to the mechanism of this proteolytic pathway is the target molecule specificity that is mediated by E3 ubiquitin ligases. The E3 ligases facilitate substrate specificity by binding to the molecule targeted for destruction with consequent complex formation with an ubiquitin bound E2 ligase, leading to the subsequent ubiquitination and degradation of the target molecule. There are three types of E3 ubiquitin ligases: RING, HECT, and RBR domain ligases⁷³. Of these three classes, the RING family is the most abundant type of E3 and is unique due to its ability of directly catalyze the transfer of ubiquitin from the E2 to the substrate^{73, 74}. The other members require a two-step reaction, whereby the ubiquitin is transferred from the E2 to the HECT/RBR domain E3 ligase and then subsequently onto the targeted molecule⁷³. Ubiquitination typically occurs on the canonical lysine (Lys-48) residue of the target molecule, leading to proteosomal degradation. However, non-canonical ubiquitination on the Lys-63 residue can also occur which can be important in responses to stress, DNA repair, membrane trafficking, kinase activation, correct ribosomal functioning, and chromatin dynamics^{72, 75}.

The structures of these E3 ligases can exhibit great variance between the many different types of E3s, even within the same E3 class, and these multiple supplementary domains that exist in addition to the RING domain can help to recruit the target substrate protein. Moreover, E3 ligases can form homodimers or heterodimers, thus increasing the diversity of substrates that can be targeted by changing their associated binding partners⁷². One prominent family of RING E3 ligases, the TRIM family, is characterized by the addition of one or two B-box domains, a coiled coil domain, and alternative C-terminal domains which are important for their diverse substrate binding specificity and the ability of TRIM proteins to either homodimerize or heterodimerize to other TRIM family proteins. More than 70 TRIM proteins are now known to exist in humans and mice and are implicated in many diverse biological processes, including innate immunity, transcriptional regulation, cell death, and development⁷². Alteration of these TRIM family members, at the genetic, mRNA, or protein level, can lead to various pathological conditions; for example, developmental disorders, modified viral virulence, neurodegenerative diseases, and cancer can develop from modulation of these genes⁷².

The Role of the TRIM Family Proteins as Tumor Suppressors in Cancer

The role of the TRIM family in cancer has been shown to be complex, with their involvement being highly context dependent. Most of the activities currently described for the TRIM family genes are oncogenic. Functions for these TRIM family genes include, but are not limited to, mediating enhanced androgen receptor transcriptional activity (TRIM24), regulation of chromatin modification (TRIM24, TRIM28, PML-RAR α fusion, and TRIM33), enhancing cell proliferation (TRIM25), promotion of p53 degradation (TRIM24 and TRIM28) and cellular transformation (TRIM24-FGFR1 fusion)⁷². Despite the relatively reduced degree to which the tumor suppressive activities of this family have been described compared to their oncogenic counterparts, their activities in suppressing tumorigenesis are nonetheless just as highly potent. Many of the same cellular functions that TRIM proteins have been shown to use to promote tumorigenic potential, like regulation of proliferation, immune response, and migration, have also been reported as areas in which other TRIM proteins play large roles in suppressing in order to prevent tumorigenesis. Notable TRIM family members associated with tumor suppressive activities

include PML (TRIM19) and TRIM24. PML acts as a tumor suppressor through its recruitment of large protein complexes to form PML nuclear bodies, which are known to play critical roles in DNA repair, transcription, apoptosis, and stem cell self-renewal⁷⁶. Loss of PML expression is found in multiple tumor types including central nervous system tumors, colorectal cancer, lung tumors, prostate adenocarcinoma, breast cancer, germ cell tumors, and non-Hodgkin's Lymphoma⁷⁷. Its fellow TRIM family member, TRIM24, has also been described to have tumor suppressive functions. However, in contrast to PML, TRIM24's tumor suppressive role is highly tissue-type dependent. TRIM24 has been well characterized as a tumor suppressor in hepatocellular carcinoma (HCC), with the formation of spontaneous liver tumors in *Trim24* null mice⁷⁸. TRIM family members play large roles in suppressing tumorigenesis, in manners both independent and dependent of their E3 ubiquitin ligase activity, through their integral activities in the maintenance of genomic integrity, cell cycle control, immune modulation, inhibition of migration, and promotion of differentiation and apoptosis^{72,79}.

The TRIM Family Play Central Roles in Cell Cycle Regulation and Maintenance of Genomic Integrity

Maintenance of genomic integrity and cell cycle control is integral to the inhibition of tumorigenesis. Loss of a cell's ability to sustain correct control of the cell cycle during replicative checkpoints as well as preserving critical pathways important for DNA repair are not only themselves hallmarks of cancer, but also can enable the acquisition of other hallmarks of cancer vital to the formation and progression of tumors⁵². One TRIM family protein integral to genomic integrity is also one of the most notable tumor suppressors of this family. PML (TRIM19) is vital for genomic integrity through its promotion of subnuclear structures called PML nuclear bodies (PML-NBs). Within these PML-NBs, PML has been found to sequester more than 100 proteins, including transcription cofactors and chromatin modifiers like DAXX and SP100⁷⁶. It is thought that these PML-NBs are critical for DNA repair as the foci of the damaged DNA marker, γ H2AX, have been found to co-localize with the nuclear bodies⁸⁰. Moreover, gamma irradiation has also been shown to increase the number of PML-NBs⁸¹. It is thought that its role in DNA damage through these PML-NBs is very important for its tumor suppressive activity as evidenced by the absence of PML-NBs in solid tumor cells⁷². Further, the downregulation of

PML has also associated with tumor progression⁷². Besides genomic integrity, PML also plays important roles in stem cell self-renewal capacity, inhibition of AKT signaling, and in innate viral immunity^{72,82}. However, as stated previously, the roles of the TRIM family proteins are complex in cancer and in accordance with this, *PML* has been found to undergo translocation (t(15;17)) to the retinoic acid receptor alpha (*RARα*) to form a fusion protein that is thought to have a dominant negative effect on multiple proteins, including PML, *RARα*, and RXR. The PML-*RARα* fusion impedes correct localization of PML and inhibits promyelocytic maturation⁸³. In leukemia, PML has been shown to be important in the maintenance of leukemic stem cells and as such, Chronic Myeloid Leukemia (CML) patients who exhibited lower PML expression showed higher rates of complete response than patients with high PML expression⁸⁴. Further, PML has been shown to be overexpressed in basal and triple negative breast cancer and this overexpression in breast cancer was shown to be correlated with mutant p53 status, reduced disease free survival, and poor prognosis^{85,86}.

Another manner in which TRIM members can suppress tumorigenesis is through their direct interaction with cell cycle regulators to inhibit cell proliferation. For example, *Trim3* suppresses tumor growth via sequestration of p21 from cyclin-D1/cdk-4, thus reducing cell proliferation^{87,88}. Heterozygous loss of chromosomal locus 11p15.5, encompassing *TRIM3*, is observed in about 24% of gliomas^{87,89}. *TRIM3* has also been reported to be homozygously deleted in gliomas and Glioblastoma Multiforme (GBM)^{87,89}. Mice harboring intracranial *Trim3* expressing GBM cells showed significantly longer survival than mice injected with GBM cells lacking *Trim3*⁹⁰. *Trim3* can suppress Pdgf-induced GBM development in mice⁸⁷. Besides regulating proliferation, overexpression of *TRIM3* has also been found to reduce anchorage independent colony formation⁹⁰. Further, expression of *TRIM3* can induce differentiation by its promotion of asymmetrical stem cell division via suppression of the Musashi-Numb-Notch signaling axis⁹⁰. *TRIM3* is significantly downregulated in hepatocellular carcinomas (HCC) and GBM^{90,91}. Low expression of *TRIM3* has been clinically associated with tumor size, histological grade, and TNM stage⁹¹. Moreover, *TRIM3* is an independent prognostic factor for overall survival⁹¹.

Another TRIM protein able to directly interact with cell cycle regulators is TRIM8. TRIM8 participates in a feed forward loop with p53, in which p53 induction of TRIM8 can potentiate p53 expression through TRIM8-mediated stabilization, activating the growth arrest arm of the pathway only⁹². Moreover, TRIM8 has been shown to be important in p53-mediated anti-proliferative action after Nutlin or Cisplatin treatment. TRIM8 is downregulated in clear cell Renal Cell Carcinoma (ccRCC) and its low expression correlates with larynx squamous cell carcinoma nodal metastasis and is a verified growth suppressor in this tissue^{93,94}.

Regulation of Apoptosis by TRIM Family Members Can Promote Tumor Suppression

Cellular homeostasis requires a balance between proliferation and cell death. Inability to properly control these two functions are known hallmarks of cancer⁵². The TRIM family's role in regulating proliferation and genomic integrity has already been described. However, their roles in the promotion of programmed cell death are just as important. TRIM proteins' regulation of major survival pathways can play a critical role in their ability to regulate apoptosis. For example, TRIM31 is known to negatively regulate c-Src's promotion of anchorage-independent survival via its binding to p52^{shc95}. TRIM31 has been found to reduce clonogenicity of the colon cancer cell line, HCT-116⁹⁶. Further, TRIM31 can inhibit proliferation and invasion of lung cancer cells and is associated with advanced TNM ((T)umor N(Lymph Node) (M)etastasis Classification of Malignant Tumors) stage as well as positive lymph node status in Non-Small Cell Lung Cancer⁹⁷. Another TRIM protein, TRIM13 (RFP2), can act as an E3 ubiquitin ligase for pro-survival genes MDM2 and AKT⁹⁸. TRIM13's proteosomal regulation of MDM2 stabilizes p53, activating the growth arrest/cell death arm of the pathway⁹⁹. Its negative regulation of MDM2 and AKT can help to augment apoptosis rates following ionizing radiation⁹⁸. TRIM13 has been shown to reduce clonogenicity levels through its negative regulation of NFκB's pro-survival arm¹⁰⁰. TRIM proteins, like TRIM13, have been shown to regulate not only vital survival pathways often implicated in cancer but also the factors directly associated with the apoptotic pathway. For example, TRIM13 is a potent inducer of cell death upon endoplasmic reticulum stress through its transport of direct apoptosis activator, casapase-8, to the autophagolysosomes, leading to its activation

and subsequent initiation of the pro-cell death caspase cascade¹⁰¹. Its regulation of autophagic cell death has been shown to reduce clonogenicity of breast cancer cell line MCF7¹⁰². In accordance with its promotion of apoptosis, *TRIM13* is downregulated and deleted in B-cell chronic lymphocytic leukemia¹⁰³.

TRIM17 (TERF) is another family member known to regulate direct cell death factors. TRIM17 is necessary for the induction of the mitochondrial associated intrinsic cell death pathway in cerebellar granule neurons (CGN)¹⁰⁴. It regulates neural apoptosis through its ubiquitination and subsequent degradation of major anti-apoptotic pathway member, Myeloid Cell Leukemia-1 (Mcl-1)¹⁰⁵. TRIM17 can also promote apoptosis in CGN cells through its participation in a feed-forward loop with NFATc3, via its sumoylation activity¹⁰⁶. Moreover, TRIM17 can negatively regulate breast cancer cell proliferation through its regulation of the mitotic spindle checkpoint, via its promotion of the degradation of the kinetochore complex member, *ZWINT*¹⁰⁷. Lending to its tumor suppressive activities, focal deletion of *TRIM17* has also been found in bladder and kidney cancer cell lines¹⁰⁸.

Other TRIM proteins shown to be important in the regulation of apoptosis factors include TRIM32 and TRIM39. TRIM32 has been shown to sensitize cells to TNF- α induced apoptosis via its E3 activity against anti-apoptotic factor, X-linked Inhibitor of Apoptosis Protein (XIAP)¹⁰⁹. Alternatively, TRIM proteins can also stabilize apoptosis promoter genes as exemplified by TRIM39's ability to stabilize Modulator of Apoptosis-1 (MOAP-1) via competitive binding with MOAP-1's inhibitor, APC/C^{Cdh-1}¹¹⁰. TRIM39's facilitation of increased levels of MOAP-1 enhances etoposide-induced, BAX mediated apoptosis. The TRIM family's pro-apoptotic regulation has been shown to be important in maintaining homeostatic levels of cellular growth and in doing so, play a vital role in the prevention of tumorigenesis.

TRIM Proteins' Modulate Pro-Tumorigenic Inflammatory Response Driven by NF κ B

Long standing evidence has already implicated the importance of TRIM proteins in the immune system and in viral response¹¹¹. Recent evidence suggests that TRIM family proteins can

regulate the immune system via regulation of Nuclear Factor κ -B (NF κ B) signaling, in roles that both stimulate and suppress this integral pathway. NF κ B is an important immune modulator as well as a regulator of both proliferation and apoptosis through its effects on apoptotic factors, adhesion molecules, and cytokines^{79,112}. Negative regulation of this pathway by TRIM family proteins has been noted to occur in both the cytoplasm and in the nucleus through the regulation of upstream cytoplasmic pathway members and direct interaction with nuclear NF κ B. Two examples of TRIM family members that regulate the NF κ B pathway via proteosomal degradation are TRIM13 and TRIM45. TRIM13's poly-ubiquitination of NEMO, a member of the IKK complex that regulates NF κ B's release to the nucleus, can induce proteosomal degradation of NEMO, suppressing NF κ B pathway activation¹⁰⁰. TRIM45 has also been found to suppress TNF induced NF κ B mediated proliferation via its RING domain¹¹³. Inability to restrain the activation of the NF κ B pathway in tumor associated immune cells within the tumor microenvironment can lead to paracrine activation of proliferation and anti-apoptotic pathways in nearby tumor cells, thereby promoting tumorigenesis¹¹².

Besides regulation of the NF κ B pathway by the TRIM family proteins through ubiquitin mediated proteosomal degradation, ubiquitination of activating NF κ B pathway members can also lead to their degradation in manners independent of the proteasome. For example, mono-ubiquitination of critical pathway activator IKK β by TRIM21 is known to induce autophagy mediated degradation¹¹⁴. Further, TRIM30a mediates endocytic lysosomal degradation of the TAB2-TAB3-TAK1 complex which suppresses Toll-like Receptor (TLR) induced NF κ B activation¹¹⁵. Regulation of the localization of NF κ B is also a major mechanism of fine-tuning pathway activation as NF κ B subunits p65 and p50 require nuclear localization in order to enact their transcriptional regulation. PML suppresses NF κ B mediated transcription by sequestering of NF κ B in PML nuclear bodies¹¹⁶. Moreover, TRIM9 has been shown to stabilize I κ B α through its competitive binding for the NF κ B inhibitor with β -TrCP E3 ligase, thereby effectively restricting NF κ B to the cytoplasm and inhibiting inflammatory cytokine production¹¹⁷. TRIM40 is also important in I κ B α stabilization, however TRIM40 maintains the pathway inhibitor's protein expression through its post-translational neddylation activity¹¹⁸. Moreover, an inverse

correlation between TRIM40 and NF κ B expression has been found in gastric tumors and their associated inflammation sites¹¹⁸. By playing large roles in the suppression of the NF κ B pathway, TRIM family proteins restrict pro-tumorigenic inflammatory signals from stimulating important pro-survival and proliferative pathways.

TRIM Proteins' Regulation of Differentiation and Migration are Integral to Tumor Suppression

Two important factors associated with increased tumor aggressiveness are the loss of differentiation and the gain of enhanced migratory capacity. Through multiple mechanisms, many TRIM family proteins are important for regulating differentiation and migration, which is vital to preventing tumor progression. TRIM24 (Tif1 α) has been shown to promote differentiation by its interaction with RAR α in a ligand dependent manner, leading to constrained RAR α associated transcription and subsequent tumor suppression in hepatocellular carcinoma^{72,119}. *Trim24*-knockout in mice can induce formation and stepwise progression of pre-neoplastic lesions (clear-cell foci of altered hepatocytes or FCA) to hepatocellular adenomas and carcinomas within one year at high incidence rates^{78, 120}. Additionally, double mutation of *Trim24* and its known major binding partner, *Trim33* (Tif1 γ), can synergize to potently enhance HCC formation in mice more than either single mutant alone⁷⁸. Besides its regulation of differentiation, molecular analysis has shown that ectopic expression of TRIM24 can stall cells in the G0/G1 phase of the cell cycle, effectively reducing cell growth, as well as inhibiting anchorage-independent colony formation¹¹⁹. Moreover, human hepatic cellular carcinoma (HCC) have been found to exhibit chromosomal loss of the 7q32 locus harboring *TRIM24*⁷².

Besides its synergism with TRIM24, TRIM family protein TRIM33, also known as Transcriptional Intermediary Factor 1 γ (TIF-1 γ), has shown in its own right to have tumor suppressive abilities through the regulation of differentiation factors. TRIM33 is important for normal hematopoiesis via its regulation of hematopoietic progenitor differentiation in response to TGF β and in its interaction with SMAD4¹²¹. Targeted deletion of *Trim33* in mice hematopoietic tissue alters the hematopoietic differentiation cascade by selectively expanding the granulo-monocytic progenitors, leading to

progressive hyperleukocytosis and severe hepatosplenomegaly in knock-out mice after or by 6 months of age¹²². In human Chronic Myelo Monocyticleukemia (CMML) and pancreatic ductal adenocarcinoma, both TRIM33 mRNA and protein levels were found to be significantly decreased^{122,123}. Moreover, conditional knockout of *Trim33* in the pancreas was found to cooperate with mutant *Kras* to induce pancreatic cystic tumors at 100% incident rate¹²³.

Increased migratory and invasive capacities are vital steps in the Epithelial-Mesenchymal Transition (EMT), which can promote metastasis to distant sites. Loss of some of the TRIM family members are known to singularly effect migration (TRIM15 and TRIM16) whereas others are known to be involved in multiple steps of EMT (DEAR1 and TRIM29). Cellular movements are coordinated through actin dynamics and the regulation of the turnover of these related proteins are important to the control of migration of cells. In accordance with this, TRIM15 has been shown to be able to regulate migration through its interacts with adaptors or scaffold proteins involved in actin dynamics¹²⁴. Notably, TRIM15 has been found to prevent the phosphorylation and subsequent activation of CTTN, a regulator of actin polymerization that has been shown to be important for cellular migration. TRIM15 is downregulated in colorectal and gastric cancer^{124,125}. Its expression in colon cancer cell lines decreases clonogenicity and tumor burden in mice inoculated with colon cancer cell lines expressing TRIM15, compared to tumors formed by the same cell lines lacking TRIM15 expression¹²⁴. Another TRIM protein, TRIM16, can act to suppress tumorigenesis by regulating not only by migration but differentiation activities as well. TRIM16 has been shown to both bind and modulate the expression of EMT marker vimentin in neuroblastoma, squamous cell carcinoma, and lung cancer, thereby leading to reduced migration capabilities^{126,127}. TRIM16 has also been found to effect migration, along with proliferation rates, through Interferon Beta 1 (IFN β 1) in melanoma¹²⁸. Consequently, its expression is downregulated during the transition from normal skin to squamous cell carcinoma¹²⁶. TRIM16 has also been found to be differentially expressed in neuroblastoma, in concordance with the degree of histological differentiation, i.e. higher expression correlated to more differentiated regions¹²⁷. In accordance, TRIM16 has been found to be able to induce differentiation within neuroblastoma and

keratinocyte cells^{127,129}. Moreover, TRIM16 was found to have reduced expression in melanoma cell lines and was correlated with lymph node metastasis and poor prognosis¹²⁸. In melanoma, TRIM16 was also able to serve as one mechanism for BRAF inhibitor Vemurafenib's induced growth arrest and subsequently, could act as a predictive marker for this clinical inhibitor¹²⁸. Besides migration and proliferation, TRIM16 also suppress proliferation through its regulation of cell cycle progression and in accordance, was shown to be able to reduce tumor growth in vivo^{126,127 130}.

Two TRIM proteins, DEAR1 (Ductal Epithelium Associated RING Chromosome 1) and TRIM29, have been identified uniquely as regulators of both early and late steps of EMT, with the loss of these two genes being associated with aberrant polarity as well as increased migration and invasion capacity. Knockdown of DEAR1 in Human Mammary Epithelial Cells (HMECs) in 3D culture results in mislocalization of polarity marker alpha-6-integrin and diffuse cell death leading to luminal filling of the breast acini, a hallmark of early breast cancer, Ductal Carcinoma *in Situ*¹³¹. Moreover, DEAR1 is a dominant regulator of TGFβ driven EMT through its activity as an E3 ubiquitin ligase for downstream TGFβ effector SMAD3². Loss of DEAR1 in the presence of TGFβ drives cellular migration and invasion in human mammary cells². A more detailed discussion of DEAR1's tumor suppressive activities are further described in the next section. Like DEAR1, TRIM29 (ATDC) also regulates both migration and invasion; however, its regulation of EMT is through its participation in a negative feedback loop regulating the prominent EMT promoter, TWIST1^{132, 133}. Further, knockdown of TRIM29 identified this TRIM family member as a dominant regulator of acinar morphogenesis through its potent regulation of polarity and proliferation via inhibition of the Estrogen Receptor (ER) and Mitogen Activated Protein Kinase (MAPK) pathways¹³². TRIM29 is significantly downregulated in prostate and breast cancer and its lower expression levels are associated with larger breast tumors as well as lymph node spread^{133,134,135}. Additionally, in early stage ER+ breast cancer without adjuvant therapy, TRIM29 expression was able to predict longer relapse free survival in women younger than 55¹³².

Tumor Suppressor and TRIM Family Protein DEAR1 (TRIM62) is a Vital Regulator of Polarity and EMT, Whose Loss is Correlated with Important Clinical Parameters

As previously discussed, DEAR1 (TRIM62) is a TRIM protein that has been shown to be important in regulating both early and late steps of EMT, an integral process in tumor progression and metastasis. DEAR1 and TRIM29, despite being structurally different, are currently the only TRIM family proteins known to regulate EMT processes of polarity deregulation and migration/invasion, albeit through different mechanisms (see previous section for detailed discussion of TRIM29's regulation of EMT). Knockdown of DEAR1 within Human Mammary Epithelial Cells (HMECs) in 3D culture resulted in loss of apical-basal polarity as well as diffuse and decreased cell death activity, leading to luminal filling of the acini¹³¹. Luminal filling within the breast ducts have been noted as a hallmark of Ductal Carcinoma *in Situ* (DCIS), one of the earliest forms of breast cancer¹³⁶. Moreover, the importance of DEAR1's regulation of acinar morphogenesis was shown by the reversion of the aberrant acinar morphology of metastatic breast cancer cell line 21MT to a smaller, more normal like morphology. Wildtype DEAR1 expression was shown to be able to restore apical-basal polarity and luminal cell death after genetic complementation of an inherent single nucleotide missense mutation in *DEAR1* at codon 187¹³¹. DEAR1 has also been reported as an E3 ubiquitin ligase for downstream TGF β effector, SMAD3². Knockdown of DEAR1 has been found not only to allow higher total and phosphorylated SMAD3 expression levels but higher nuclear expression of activated SMAD3 as well. Further, TGF β -SMAD3 dependent signaling was shown to be potentiated with DEAR1 loss of expression, including increased expression of downstream target genes and EMT effectors SNAIL1 and SNAIL2². In HMECs and the normal mammary cell line MCF10A, the loss of DEAR1 in the presence of TGF β prevented acini formation in 3D culture, upregulation of EMT markers, promotion of anoikis resistance, and drove migration and invasion of the cell lines². DEAR1, however, is not the only TRIM family member known to be an E3 ligase for the TGF β pathway. TRIM33 is also known to bind to TGF β pathway members SMAD2/3 and SMAD4¹³⁷. TRIM33's binding to SMAD2/3 and SMAD4, unlike DEAR1's binding to SMAD3, does not lead to degradation. TRIM33 acts in a dual manner in its interaction with SMAD proteins through its inhibitory mono-ubiquitination of SMAD4 and its co-transduction of TGF β signaling by complex formation with SMAD2/3¹³⁷. These interactions have been found to be important in the differentiation axis of the TGF β pathway, with no effect on TGF β 's regulation of proliferation¹³⁷. Like

TRIM33, DEAR1 does not effect the proliferation arm of the TGF β pathway but rather is known only to modulate TGF β induced migration and invasion¹³¹. Further, *Dear1* loss can also synergize with mutant *Kras* to drive lung cancer cell migration and invasion through enhanced promotion of EMT, increasing metastatic potential in mice¹³⁸. Previously reported results, as discussed here, have indicated that DEAR1 is a dominant regulator of acinar morphogenesis and potent inhibitor of TGF β -induced EMT, which are cellular functions that can be important in suppression of tumor progression.

Loss of heterozygosity (LOH) has been found to be common amongst the TRIM family members who exhibit tumor suppressive activities and is a common mechanism of tumor suppressor inactivation^{2, 89, 103, 131, 139-141}. *DEAR1* is localized to chromosome 1p35.1, a region which often undergoes chromosomal deletion in multiple epithelial cancers, including breast, lung, and colon cancer^{142, 143}. As such, *DEAR1* has been shown to be frequently heterozygously deleted as well as rarely homozygously deleted in many cancer types, including brain, breast, colorectal, lung, and endometrial carcinoma^{2, 131}. Loss of *Dear1* in mice resulted in late onset formation of multiple epithelial cancers comprising, but not limited, to lymphomas, sarcomas, lung adenocarcinomas, gastrointestinal carcinomas and breast adenocarcinoma². Further, a few primary tumors were found to be associated with metastatic lesions. *Dear1*^{-/-} and *Dear1*^{-/+} mice formed tumors at similar rates, with evidence for *Dear1* acting as a haploinsufficient tumor suppressor in lymphomas and a classical tumor suppressor, i.e. requiring two hits for inactivation, in epithelial tissues. The spectrum of human tumor types undergoing LOH at the *DEAR1* locus was able to be recapitulated by heterozygous and homozygous loss of *Dear1* in a mouse model². Moreover, *DEAR1* undergoes rare mutation in multiple cancers in a spectrum similar to those tumor types which experienced LOH, including lung squamous, endometrial, colorectal, and breast cancer^{2,131}. Currently, the highest frequency of mutation within *DEAR1* is 13% (n=55) in invasive breast cancer, as identified by our lab¹³¹. Invasive breast cancer sequencing efforts by TCGA only report 1 mutation (frameshift) in 962 cases; however, sequencing coverage achieved by TCGA is typically low with a median of 20x coverage which has low sensitivity to find rare subclonal variants¹⁴⁴⁻¹⁴⁶. For example, TCGA failed to identify *ESR1* mutations in invasive breast cancer, yet multiple investigators later

reported finding common mutation of the gene¹⁴⁷. Pancreatic, stomach and colorectal cancer experience the next highest frequencies of mutation in *DEAR1* at 3.6% (n=55), 1.8-3.3% (n=30,220,287), and 1.4-1.9% (n=72,212,220) respectively^{2, 145, 146}. Most of the mutations previously identified in *DEAR1* are missense mutations; however, a small number of frameshift and nonsense mutations have been reported as well^{2, 145, 146}. The *DEAR1* missense mutation R187W that was identified within both an invasive breast cancer patient and in the metastatic breast cancer cell line 21MT, to our knowledge, represents the first identified loss of function missense mutation in TRIM family members in cancer, as confirmed by molecular analysis¹³¹. Many TRIM family genes have been found to be mutated in cancer, including the instances of nonsense and frameshift mutations, as shown by mutation catalogues COSMIC and cBIO; however, no TRIM family member, except for *DEAR1*, has been described by literature to have loss of function missense mutations in cancer^{145, 146, 148}. However, previously reported TRIM family loss of function missense mutations are associated with multiple developmental disorders including Limb girdle muscular dystrophy type 2I (TRIM32) and X-linked Opitz G/BBB syndrome (MID1)^{149, 150}.

Besides alteration at the genetic level, *DEAR1* also is downregulated at the protein level. *DEAR1* exhibits reduced expression in multiple epithelial cancers, including pancreatic cancer, acute myeloid leukemia (AML), lung cancer and breast carcinoma^{2, 138, 151, 131}. In pancreatic cancer, 62% of tumors showed downregulation of *DEAR1*, as assayed by immunohistochemistry². In AML patients, *DEAR1* showed reduced expression levels when compared to CD34+ cells from healthy controls¹⁵¹. *DEAR1* was also found to show complete loss of expression in 64-87% of Non-Small Cell Lung Cancer (NSCLC), with a stepwise decrease in expression associated with progression of the disease from normal bronchial epithelium to NSCLC¹³⁸. Similarly, in the transition from normal breast ductal epithelium to one of the earliest forms of breast cancer, DCIS, *DEAR1* was shown to be downregulated in 56% of early onset breast cancer lesions (ages 25-49 years) in premenopausal women¹³¹. Further, the frequent loss of expression of *DEAR1* has been shown to be clinically important. *DEAR1* has been reported as an independent marker of adverse prognosis in AML with low expression of *DEAR1* associated with shorter complete remission and reduced overall survival¹⁵¹. Further, in early stage NSCLC, *DEAR1* loss

of expression correlated with significantly shorter time to relapse¹³⁸. DEAR1 also correlated with triple negative breast cancers with a strong family history and poor prognosis¹³¹. Moreover, DEAR1 was associated with reduced local recurrence free survival as well in early onset breast cancer¹³¹.

Multiple classes of tumor suppressors have been proposed throughout the years which regulate different classes of cellular pathways critical to maintaining homeostasis, as previously discussed. One novel class of tumor suppressors that was proposed by Petersen and colleagues were described for their unique abilities to sense the microenvironment and thus regulate the organization of tissues in an appropriate manner¹⁵². Alterations leading to loss of function of these tumor suppressors would cause deregulation of spatial recognition through impeding proper cellular signaling and loss of spatially restricted localization of adhesion and polarity pathway members. Loss of polarity and weakening of cellular adhesion are early steps in EMT, and as such, tumor suppressors who sense the microenvironment and in accordance, regulate epithelial plasticity through correct spatial control would be important in suppressing tumor progression through EMT¹⁵³. Examples of this type of tumor suppressor include p53, which has been reported to have gain of function mutations that can disrupt polarity, acinar morphogenesis and EMT, as well as polarity regulators LKB1 and SCRIBBLE which are noted to be downregulated in breast cancer and whose loss of function induces luminal filling, which is reminiscent of the luminal filling that occurs in DCIS¹⁵⁴⁻¹⁵⁹. Similarly, DEAR1 has also been shown to be important in the regulation of acinar morphogenesis, polarity and EMT induced by extracellular signaling molecule, TGF β . As such, it is plausible for DEAR1, as well as its fellow TRIM family member TRIM29, to be identified as the first two TRIM family proteins to fit within this novel class of tumor suppressors which sense factors from the microenvironment and regulate pathways involved in the control of cellular adhesions and polarity. Loss of DEAR1, through genetic aberrations or loss of expression in cancer, leading to the loss of a cell's ability to sense the microenvironment and thus as a result, failure of the cell to regulate spatial control, can play a vital role in the progression of cancer and subsequently in determine patients' prognosis.

Investigation of Genomic Alterations in *DEARI* Using Pan-Cancer Analysis and Ultra-Deep Targeted Sequencing in *Ductal Carcinoma In Situ* (DCIS)

Given previous data supporting the importance of the loss of expression and function of *DEARI* in promoting tumorigenesis, the degree and potential functionality of genetic alterations within this gene were characterized. I hypothesized that *DEARI* is mutated and chromosomally lost in multiple epithelial cancer types, consistent with a tumor spectrum associated with chromosome 1p loss, and that these alterations are genetic drivers in tumorigenesis. Further, it is important to specifically determine if *DEARI* is mutated in the earliest form of pre-invasive breast cancer, DCIS, as *DEARI* has been previously reported to be mutated in invasive breast cancer (IDC)¹³¹. Therefore, I hypothesized that mutations in the EMT regulator, *DEARI*, not only exists but are functional and can act as genetic drivers in the progression from DCIS to IDC. To determine the validity of our hypothesis, the following aims were carried out:

1. Characterization of *DEARI* as a 1p35 tumor suppressor
 1. Determine if *DEARI* undergoes copy number losses and mutation in cancers associated with chromosome 1p35 LOH.
 2. Determine if *DEARI* genetic alteration could be useful as a prognostic biomarker in breast cancer.
2. Create a custom targeted Ampliseq panel for *DEARI* and characterize its analytical performance
 1. Design a *DEARI* ampliseq Panel and characterize its sequencing capacity
 2. Test *DEARI* ampliseq panel accuracy by novel spike-in assay
 3. Determine the sensitivity and specificity of *DEARI* targeted Ampliseq panel
3. Determine if *DEARI* is mutated in DCIS and in DCIS associated with an Invasive Component

1. Complete ultra-deep targeted sequencing of *DEARI* on 17 Pure DCIS samples by next generation sequencing
2. Microdissect 19 DCIS lesions and adjacent invasive component and complete ultra-deep targeted sequencing of *DEARI* on these components independently by next generation sequencing
3. Analyze *DEARI* mutation spectrum in pure DCIS and compare to DCIS associated with Invasive components (DCIS/INV).
4. Validate and functionally characterize variants found in *DEARI* in DCIS

The results of the research herein, following these specific aims, describe a vast tumor spectrum associated with mutation or loss of the critical polarity and EMT regulator *DEARI* and how these alterations can act as functional drivers in tumorigenesis. The relevancy of the *Dear1* knock-out mouse model has been validated for its ability to recapitulate the tumor types developed in humans that are associated with chromosome 1p loss. Heterozygous loss of *DEARI* has also shown the ability to synergize with amplification of EMT promoter *SNAI2* to predict overall survival in IDC. Further, a highly sensitive, custom Ampliseq panel has been developed and used to complete ultra-deep targeted sequencing of *DEARI* in DCIS. Results indicated the high degree of frequency that *DEARI* is altered in these early lesions of breast cancer and evidence also supported a parallel yet independent model for tumor evolution for DCIS and IDC lesions. Future work is still needed to complete our understanding of the full role of *DEARI* in cancer, but our results hint at the importance of these alterations in driving tumorigenesis.

Identification of DEAR1 Alterations across Cancers Using Pan-Cancer Database Analysis

This chapter is based upon Nanyue Chen, Seetharaman Balasenthil, Jacquelyn Reuther, Aileen Frayna, Ying Wang, Dawn S. Chandler, Lynne V. Abruzzo, Asif Rashid, Jaime Rodriguez, Guillermina Lozano, Yu Cao, Erica Lokken, Jinyun Chen, Marsha L. Frazier, Aysegul A. Sahin, Ignacio I. Wistuba, Subrata Sen, Steven T. Lott and Ann McNeill Killary. “DEAR1 Is a Chromosome 1p35 Tumor Suppressor and Master Regulator of TGF- β -Driven Epithelial–Mesenchymal Transition” *Cancer Discovery* 2013; 3: 1172-1189. No permission is required for the reprinting of the article as Cancer Discovery states:

“Authors of articles published in AACR journals are permitted to use their article or parts of their article in the following ways without requesting permission from the AACR. All such uses must include appropriate attribution to the original AACR publication. Authors may do the following as applicable: Submit a copy of the article to a doctoral candidate's university in support of a doctoral thesis or dissertation”.

Introduction

Tumor initiation and progression are processes dominated by the culmination of aberrant changes at the gene, transcript, and protein level. By understanding the aberrant landscape of the cancer at each of these levels, the possibilities for better assessment of cancer risk, improved earlier detection, and identification of new therapeutic targets are possible. For solid tumors, elucidating a clear targetable pattern of genomic alterations has been difficult and only few success stories exist. Two examples of the relatively few targeted therapies in solid tumors that exist are Herceptin and Vemurafenib. A targeted therapy that has revolutionized treatment of breast cancer is Herceptin, a monoclonal antibody against HER-2 (ERBB2), a gene that is commonly amplified in breast cancer and marks an aggressive subtype of this disease. Herceptin has a 34% response rate as a single agent therapy in breast cancer patients experiencing HER2 amplification¹⁶⁰. Moreover, Vemurafenib is a therapy that has been developed in recent years for metastatic melanoma as a selective mutant B-Raf inhibitor targeting cells exhibiting

mutations within codon 600 of the gene, with highest binding efficiency against mutant B-Raf V600E¹⁶¹. Vemurafenib has shown to be highly effective with greater than 50% of melanoma patients with mutant B-Raf responding to treatment with this targeted therapy¹⁶¹. Further, both Herceptin and Vemurafenib have shown the ability to synergize with other therapeutic treatments for even greater response rates^{162, 163}. The relative failure to find more targets which can be utilized for the treatment of solid tumor types is in part due to the complexity that exists at the genomic level in solid tumors that is unlike that which is found in sarcomas, leukemia, and lymphomas³³. Whereas sarcomas and liquid tumors have highly recurrent chromosomal arrangements and few mutations, solid tumors often have relatively low levels of recurrent chromosomal rearrangements and moderate to high frequencies of mutational alterations, with a mutational pattern that has been defined as “long tailed”, meaning that relatively few genes are altered at high frequencies in these cancer and rather the population is dominated by many genes mutated at low frequencies in these lesions^{33,164,56}. Due to the complexity of genomic alterations observed in solid tumors, the relatively few recurrent chromosomal alterations and mutations which occur in these cancers signal important areas which may feature genes vital to tumor initiation and progression. For example, a fusion of the genes *TMPRSS2* (21q22.3) and *ERG* (21q22.2) represent one of the most common chromosome rearrangements in human cancer and is associated with invasion in prostate cancer¹⁶⁵. Moreover, another example includes the loss of the 1p chromosome arm, which is noted as a common feature of many epithelial cancers. Deletion resulting in loss of heterozygosity (LOH) has been known to occur at this chromosomal arm at a relatively high frequency in multiple epithelial tumor types (29-72%), including stomach, breast, lung, kidney, and colorectal cancer, especially within regions 1p31 and 1p34-35 which experience the highest frequency of chromosomal loss^{143,142}. Moreover, LOH at the 1p locus has been associated with ER+ breast tumors and was linked to, in a multivariate analysis in breast cancer, with a 2.7 fold increase in relative risk of death^{142,166}. Often large scale deletions affecting the entire chromosome arm occur in cancer development and progression, suggesting a mechanism for the loss of multiple tumor suppressors at one time that reside within the genomic interval. For example, the chromosome 1p arm harbors *p73* and *CHD5* within 1p36, *DEAR1* (*TRIM62*) at 1p35, and *MUTYH* at 1p34. All of these genes exhibit important tumor suppressive activities. *p73* and *CDH5* are important

regulators of cell cycle control, apoptosis, and senescence¹⁶⁷⁻¹⁶⁹. *DEAR1* is a dominant regulator of acinar morphogenesis and polarity, as well as an inhibitor of TGF β induced Epithelial to Mesenchymal Transition (EMT) and cellular invasion^{2, 131, 138}. Lastly, *MUTYH* is a base excision repair gene important in DNA damage repair whose biallelic mutation is common in the inherited colorectal cancer syndrome MutYH Associated Polyposis¹⁷⁰. The combined loss of important regulators of cell cycle control, apoptosis, polarity, invasion, and DNA repair can potentially greatly promote tumorigenesis via its allowance of uncontrolled cellular growth, increased mutation rate, and promotion of migration and invasion. A better understanding is needed to fully comprehend the functional effect concerning the combinatorial loss of multiple important tumor suppressors in driving tumor progression.

An indication of the importance of *DEAR1* as a “bona fide” tumor suppressor as well as the critical nature of copy number alteration in *DEAR1*’s function as a tumor suppressor was discovered by targeted disruption of the *Dear1* locus in the mouse^{2, 131}. TRIM family proteins have been noted to play significant roles in cancer and previous to this, our lab had shown that *DEAR1* was mutated and homozygously deleted in breast cancer, indicating its potential importance in oncogenesis. Targeted disruption of the *Dear1* locus in the mouse indicated that *Dear1*^{-/-} and *Dear1*^{+/-} mice developed late onset tumors with a frequency of 12.9% (8/62) and 17.7% (17/96), respectively². The tumors that arose in the mice were diverse in spectrum and encompassed multiple epithelial tumor types; for example, they included hepatocellular, mammary, lung, and pancreatic tumors, as well as sarcomas and lymphomas². Interestingly, *Dear1*^{+/-} mice developed tumors with similar frequencies as the *Dear1*^{-/-} mice. Moreover, many of the *Dear1*^{+/-} mice tumors exhibited further deletion of the wildtype allele, except tumors of the lymphatic system, which suggested that *Dear1* is a classic tumor suppressor with the potential ability to act as a tissue dependent haploinsufficient tumor suppressor. Therefore, loss of a single allele of *DEAR1* can potentially have significant effects upon gene dosage and function. Understanding the frequency of *DEAR1*’s loss in solid tumors and other types of alterations, including mutations and expression changes, is important for further elucidating the significance of the loss of the p arm of chromosome 1 in cancer.

Methods

Databases including cBio (MSKCC), CONAN (Wellcome-Sanger), Catalogue of Somatic Mutations in Cancer (COSMIC) (Wellcome-Sanger), and OncoPrint were investigated for information concerning genomic alterations and transcript expression in human tissue samples, The Cancer Genome Atlas (TCGA) cohorts, human cell lines, and other reported papers^{145, 146, 148, 171}. Mutation functionality assessment was completed via PolyPhen2 software (V2.2.2) and SIFT^{172, 173}. Survival curves were generated by cBio, using Kaplan-Meier analysis through querying complete tumor sets in the BRCA cohort for *DEAR1* heterozygous loss and *SNAI2* amplification^{145, 146}.

Results

DEAR1 Displays Chromosomal Loss and mRNA Downregulation in Human Tumors Associated with Chromosome 1p Loss, in a Tumor Spectrum Similar to the Tumors Derived from the DEAR1 Knockout Mouse Model

LOH of an important tumor suppressive chromosomal locus can greatly affect gene expression, especially if those genes involved are haploinsufficient, in which the loss of a single allele is enough to cause inactivation of the tumor suppressor and can thereby greatly potentiate tumor initiation and progression. Since previous data had hinted at the frequent recurrent loss of chromosome 1p and further, our lab had shown that loss of a single allele of *Dear1* in the mouse model can potentiate tumor formation, it was determined if the tumor spectrum that was found in the *Dear1* knockout mouse model recapitulated the human tumor spectrum associated with *DEAR1* locus chromosomal loss. Large genomic characterization efforts from TCGA and the Broad Institute have led to an updated view of genomic alterations in many cancers. To determine if the *Dear1* mouse model tumor spectrum was reflective of the spectrum found in human tumors associated with chromosome 1p loss, database analysis was conducted to ascertain the human tumor spectrum associated with *DEAR1* chromosomal loss and gene mutation. *DEAR1* was found to exhibit LOH in a similar tumor spectrum as the *Dear1* knockout mice by analysis of human cell lines and tissues from the CONAN and cBIO databases, sharing an

association with intestinal, breast, lung, hepatocellular, pancreas, hematopoietic, and sarcoma tumors (Fig. 1). In the CONAN human cell line database, *DEARI* chromosomal loss was found in tissues including, soft tissue sarcomas (37%), pancreatic (31%), thyroid (50%), lung (32%), hepatocellular (33%), renal (33%), breast (18%), and gastrointestinal tumors (28%) (Fig. 1)^{2, 148}. Further, the same tumor spectrum featuring LOH of the *DEARI* locus was found by analyzing data from human tumor samples in TCGA, accessed by cBio (Table 1)^{2, 145, 146}. Colorectal and hepatocellular cancer, as well as invasive breast cancer were found to have the highest frequency of LOH (Table 1; unpublished data)². Further, a putative homozygous deletion of *DEARI* was found in two samples of hepatocellular carcinoma and single samples of glioblastoma multiforme, low grade glioma, ovarian serous cystadenocarcinoma, and breast carcinoma from the TCGA project, as accessed by cBio (unpublished data)^{2, 145, 146}. Previously, our lab had shown that the existence of a homozygous deletion in an invasive breast cancer sample, as well¹³¹. Moreover, it was found that the heterozygous chromosomal loss of the *DEARI* locus, which occurs in about 32% of invasive breast cancer (IDC) patients, along with the genetic alteration of the Epithelial to Mesenchymal Transition (EMT) promoter *SNAI2* which occurs in about 6% of IDC cases, significantly correlated with worse overall survival (p=0.016) (Fig. 2)^{145, 146}. This data shows that chromosomal alterations of driver genes, even of a single allele, can potentially be important drivers in the progression of cancer.

The concordance of downstream gene expression changes effected by chromosomal alterations is important to validate as this indicates the effectiveness of these alterations. Our group has demonstrated that, in multiple tissue types, *DEARI* chromosomal loss does correlate with mRNA expression via transcriptome sequencing in colorectal, lung adenocarcinoma, head and neck squamous cell carcinoma, hepatocellular, ovarian, serous cystadenocarcinoma, prostate, stomach, and lung squamous cancer (unpublished data)² (Table 2). Besides these tumor types, *DEARI* also showed a general mRNA downregulation in glioblastoma and lymphoma, as well as protein downregulation in the transition of normal breast to DCIS, and DCIS to invasive breast cancer as our lab has previously shown. All of these tumor types known to exhibit mRNA alteration in accordance with chromosomal alteration

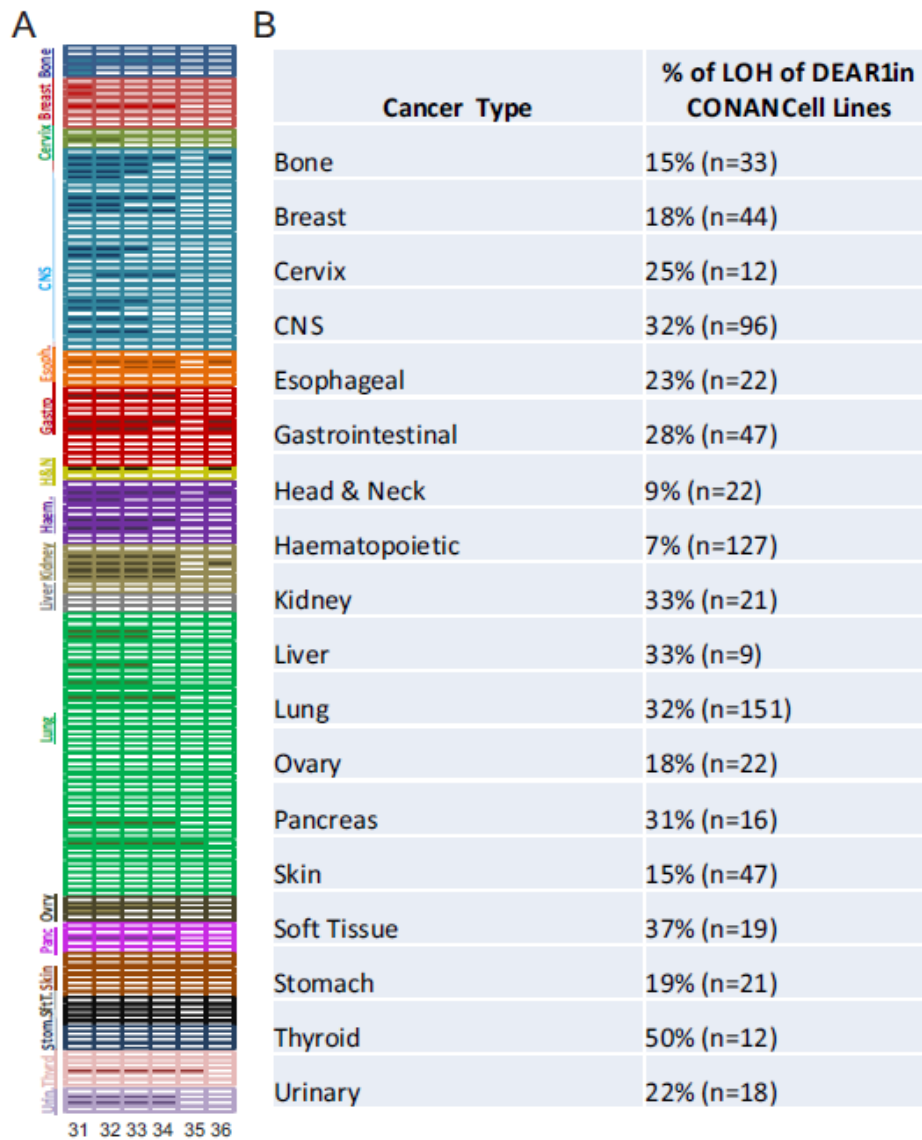


Figure 1. *DEAR1* LOH in Various Cancer Cell Lines as Shown By CONAN-Copy Number Analysis (Sanger Institute) software. (A) LOH within chromosome 1p31 to 1p36 involving *DEAR1* in CONAN cell lines with empty boxes corresponding to loss of alleles and colored boxes indicative of retention of alleles. The x axis indicates the genomic interval within the p arm that is deleted and the y axis indicates the tissue type. Each line represents an individual cell line. (B) Table summarizing data visualized in (A) and describing the percentage of LOH of *DEAR1* in multiple cancer types in the CONAN database.

Cancer Type	Heterozygous Loss in TCGA
Breast Cancer	33.5% (n=866)
Colorectal Cancer	30.3% (n=575)
Ovarian Serous Cystadenocarcinoma	24.5% (n=559)
Lung Squamous Cancer	22.9% (n=179)
Kidney Renal Clear Cell Carcinoma	13.1% (n=436)
Glioblastoma Multifome	8.5% (n=497)
Uterine Corpus Endometriod Carcinoma	4.7% (n=363)

Table 1- DEAR1 Exhibits Heterozygous Loss in Multiple Human Tumor Types. DEAR1 was found to undergo heterozygous chromosomal loss in many epithelial tissues and moderate to low frequencies. Often, loss of DEAR1 locus was encompassed in the loss of the entire p arm of chromosome 1, an area containing multiple tumor suppressors.

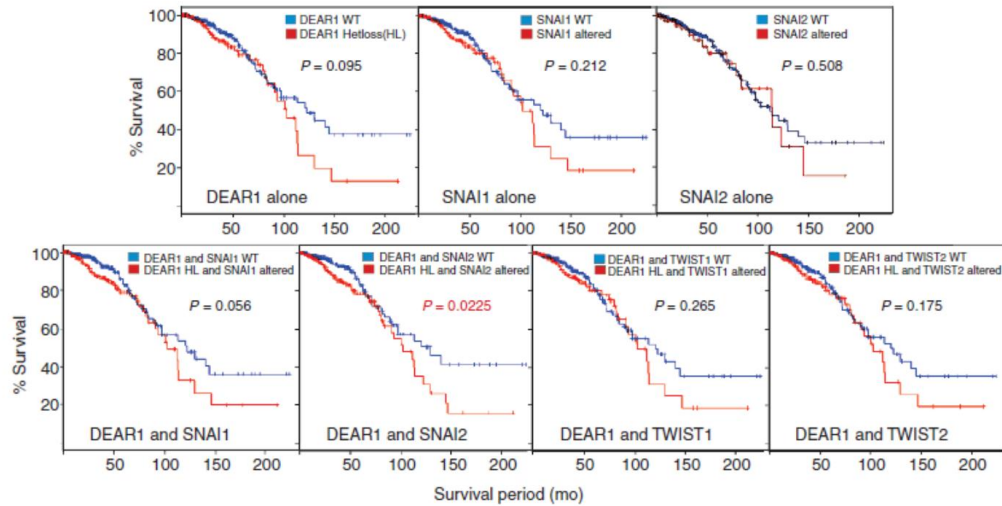


Figure 2- *DEAR1* Heterozygous Loss and *SNAI2* Alterations Can Predict Overall Survival. This figure shows the effect of *DEAR1* heterozygous loss and *SNAI1/2* gene upregulation on survival of patients with invasive breast cancer. Survival curves were generated by cBio, using Kaplan–Meier analysis through querying complete tumor sets in the BRCA cohort for *DEAR1* heterozygous loss, *SNAI1*, *SNAI2*, *TWIST1*, and *TWIST2*. Alteration of *SNAI1/2* and *TWIST1/2* includes amplification, upregulation of mRNA/protein expression (if applicable) greater than two SDs from the mean. Figure is taken from Chen et al. 2013 (Figure 7c)².

Cancers that exhibit DEAR1 mRNA downregulation		Median log ₂ mRNA (Range) in Normal Samples	Median log ₂ mRNA (Range) in Tumor Samples	P-Value
COLORECTAL	<i>Colon Adenocarcinoma</i> (n=101)	.02 (.319 - .22)	-.433 (-.047 - .834)	3.09 E-9*
	<i>Rectal Adenocarcinoma</i> (n=60)	-.263 (.051 - .593)	-.504 (-.047 - .942)	7.48 E-10*
	<i>Cecum Adenocarcinoma</i> (n=22)	.02 (.319 - .22)	-.115 (.285 - .784)	2.98 E-5 ^Δ
BRAIN	<i>Glioblastoma</i> (n=515; n=81)	.305 (.718 - .03); .855 (1.48 - .609)	-.29 (.281 - .768); .52 (1.287 - .279)	5.32 E-5 ^Δ 1.01 E-5 ^Δ
LYMPHOMA	<i>Peripheral T-Cell Lymphoma</i> (n=28)	1.847 (2.511 - 1.08)	.965 (1.448 - .059)	9.58 E-10*

Table 2- DEAR1's Expression is Downregulated in Multiple Tumor Types. Using data from the TCGA project (colorectal and brain), Sun et al. 2006 (brain n=81), and Piccaluga *et al.* 2007 (lymphoma), along with analysis from the Oncomine database (oncomine.org), DEAR1 is found to have significant downregulation in multiple tissue types. Significance was assessed using a Bonferroni corrected p-value of 2.45 E-6 for colorectal cancer, 3.96 E-6 (2.55 E-6 for Sun 2006) for brain cancer, and 2.55 E-6 for lymphoma due to multiple comparisons. It is important to note that the data from colorectal cancer and lymphoma also reached a significance level associated with genome wide significance (5 E-8)●. Other data was shown to be approach significance in cecum adenocarcinoma and brain cancer Δ. Figure is taken from Chen et al. 2013²

are also known to undergo chromosomal loss of the 1p35 region^{2, 131}. Thus, *DEARI* alterations at the genetic and transcript level in human tumors occur in multiple tumor types, with a tumor spectrum similar to what is observed in the *Dear1* mouse model, suggestive that these alterations can potentially drive tumor progression in cancer.

DEARI Exhibits Rare Mutation in Multiple Epithelial Cancers

In addition to chromosomal alterations, genetic mutations play a large role in altering gene function. By analysis of data from the TCGA project and other sequencing projects, as accessed by cBIO, and COSMIC as well as by sequencing data produced by our own lab, *DEARI* is mutated at rare frequencies in multiple tumor types (Table 3)^{2, 131}. *DEARI* has been found to be mutated in Lung Squamous (2.4%), Uterine (1.8%), Stomach (1.8%), Colorectal (1.8%), Melanoma (1%), Renal Cell Carcinoma (0.5%), Breast (13% MD Anderson cohort/ 0.1% TCGA cohort), and Bladder cancer (0.8%). Moreover, the *DEARI* mutation tumor spectrum of human tissue samples also reflected the tumor spectrum of the *Dear1* knockout mouse, including breast, lung, pancreatic, and colorectal cancer, similar to those tissues that developed LOH at the *DEARI* locus². To date, most of the mutations reported in *DEARI* are within the coding region and predicted by tools like PolyPhen2, MutationTaster, and SIFT to be potentially deleterious^{2, 131, 174}. Further, sequencing by the Broad Institute and the TCGA project has discovered the presence of multiple variants that are highly abundant within the tumor samples, including a V40M exonic variant found in a Head and Neck patient residing in 50% of the tumor as well as a E138K exonic variant found in a melanoma patient in which their normal allele was lost leading to 100% allele frequency of the variant^{145, 146, 175}. These sequencing efforts have also, and more importantly, found the presence of ultra-rare frameshift and nonsense mutations which presumably knockout *DEARI*'s expression. Moreover, since chromosome 1p35 is known to undergo frequent chromosomal loss as previously discussed, a mutation in the retained copy or even a copy neutral event (deletion followed by a duplication event) could induce a homozygous mutant condition, thereby potentially increasing the functional effect of the mutant. On this note, analysis of the TCGA data has also shown the rare co-occurrence in patients with mutation and copy number alteration of *DEARI*. For example,

Cancer Type	Missense Frequency	Source
Invasive Breast Carcinoma	13% (7/55) 0.1% (1/962)	Lott et al 2009 cBIO (TCGA)
Pancreatic Carcinoma	3% (2/65)	Chen et al 2013
Colorectal Carcinoma	2% (4/221)	cBIO (TCGA)
Stomach Adenocarcinoma	2% (4/219)	cBIO (TCGA)
Kidney Renal Clear Cell Carcin.	0.5% (2/418)	cBIO (TCGA)
Lung Adenocarcinoma	0.5% (1/182)	cBIO (Broad)
Lung Squamous Cell Adenocar.	0.6% (1/178)	cBIO (TCGA)
Skin Cutaneous Melanoma	0.4% (1/228)	cBIO (TCGA)
Glioblastoma Multiforme	0.5% (1/236)	cBIO (TCGA)
Uterine Corpus Endomet. Carcin.	0.8% (2/240)	cBIO (TCGA)
Lower Grade Glioma	0.5% (1/215)	cBIO (TCGA)

Table 3- *DEARI* Undergoes Rare Mutation in Multiple Tumor Types. Mutation frequencies of *DEARI* as detailed by Lott et al. 2009, Chen et al. 2013, and Memorial Sloan Kettering Cancer Center's (MSKCC) cBio Portal.

in the TCGA cohort, the variant D421G in colorectal cancer is associated with a chromosomal gain^{145, 146}. Moreover, a Y439H variant in a renal clear cell carcinoma patient and a P270* frameshift variant in an invasive breast cancer patient, also discovered through TCGA sequencing, was found to exhibit concomitant heterozygous loss of the other allele^{145, 146}. The co-occurrence of chromosomal gains or losses can alter mutation allele frequencies, thus promoting the deleterious effects of the driver mutations. The vast number of mutations in *DEARI* that are reported are missense mutations. Missense alterations can potentially play large roles in effecting the function of tumor suppressors if they are localized in areas known to be important for binding to other cell signaling regulators, thereby possibly effecting downstream signaling and resulting in large phenotypical changes. For example, one of *DEARI*'s exonic mutations, R187W, that was found in both an 87 year old breast cancer patient and the 21MT breast cancer cell line series, has been shown to be functional, as the genetic complementation with wildtype *DEARI* into the mutant 21MT cell line restored acinar morphogenesis in 3D culture¹³¹. There is also potential for mutations to have the ability to stratify for recurrence, progression to invasive disease, or even therapeutic sensitivity. For instance, though only a small number of samples (n=10) were available with relevant clinical information, 70% of these mutations in *DEARI* were associated with lymph node involvement or metastasis (Table 4)².

Discussion

Our lab has previously shown that *DEAR1* is a pivotal tumor suppressor in breast cancer through its ability to regulate acinar morphogenesis and to suppress EMT, an important step for the dissemination of tumor cells to distant sites^{2, 131}. *DEAR1* loss of expression or loss of normal function leads to aberrant acinar polarity, luminal filling, and enhanced activation of downstream SMAD3 targets in the presence of the TGF β ligand^{2, 131}. Heterozygous or homozygous loss of the *DEARI* allele has been shown in the mouse model to result in late onset tumor formation, at similar rates respectively, in multiple epithelial tumors including intestinal adenocarcinoma and lymphoma as well as other epithelial cancers like mammary, lung, and pancreas². Results of database analysis indicate that the spectrum of tumors formed in the mice upon *Dear1* allelic loss is similar to the tumor spectrum associated with

Mutation	Cancer	Stage	Disease Status	Invasive/Metastatic
D106V	Breast Carcinoma	N/A	N/A	Yes; Lymph node involvement
E138K	Melanoma	N/A	Malignant	metastatic
R187W	21MT cell line; metastatic breast cancer	IV	Cell line was derived from metastatic lesion of patient who had expired from Stage IV breast cancer	Yes; Metastatic
R190H	Colon Adenocarcinoma	IIA	Recurred/Progressed	Invasive
R223C	Colon Adenocarcinoma	IIB	Recurred/Progressed	Invasive
R307C	Colon Mucinous	I	Recurred/Progressed	Invasive
R333C	Endometrial Carcinoma	I; Grade 3	Disease free; Grade 3 indicates poor differentiation/aggressiveness	No
V350I	Breast Carcinoma	III	N/A	Invasive
E370K	Endometrial Carcinoma	I; Grade 1	Disease free	No
D421G	Rectal Carcinoma	IIIC	Recurred/Progressed	Yes; Lymph node involvement

Table 4- Correlation between *DEARI* Mutation and Clinical Outcome. Clinical information included records detailed to us by Dr. Kelly Hunt and Dr. Aysegul Sahin as well as clinical information provided by MSKCC's cBio Portal. *Cases with Accessible Clinical Information.

chromosome 1p LOH in humans, including lung, colorectal, lymphoma, and breast cancer, indicating the relevancy of the mouse model as well as the importance of *Dear1* to tumor development. Moreover, human mutation in *DEARI*, despite occurring at a relatively low incident rate, was also found to be associated with the spectrum of tumors formed in the *Dear1* knockout mouse model including colorectal, stomach, and renal cell carcinoma^{145, 146}. However, current methods to sequence *DEARI* within large consortium sequencing projects and even previously by our lab involved low sequencing coverage techniques, meaning that these technologies may not have been able to detect *DEARI* mutations if the variant allele frequencies occurred at levels below the sequencing platforms' threshold of detection. It is possible that the incident rate of *DEARI* mutations in human cancer reported herein may be underestimated. Therefore, ultra-deep sequencing may be required to detect *DEARI* mutations that can occur at lower variant allele frequencies within small subpopulations of the tumors. In all, the recapitulation of the human tumor spectrum by the *Dear1* mouse model indicates the significance of the loss of the *Dear1* locus as well as how LOH and homozygous deletion of *DEARI* in human tumors can act as a potential driver in cancer.

Recently it has been shown that germline variants and somatic genomic alterations explain about only 39% of all expression changes¹⁷⁶. Moreover, copy number alterations do not always correlate with mRNA expression, with only about 20% of cis-acting copy number alterations having been shown to effect mRNA expression¹⁷⁶. This discrepancy between copy number loss and expression, therefore, shows not only the ability of cells to compensate for gene expression after the loss of a chromosomal locus, but also the importance of particular genes whose chromosomal loss does correlate with mRNA downregulation, as these genes may have important functions within the tumor. In credence with this theory, our lab has shown that *DEARI* chromosomal loss is associated with reductions in gene expression in multiple tumor types, many of which are reflective of the tumor types that developed within the *Dear1* knockout mouse model. This is important as *DEARI* was shown to exhibit a relatively high frequency of LOH in multiple epithelial human tumor tissues, with largest frequencies in colorectal, breast, and hepatocellular carcinoma. In fact, heterozygous loss of *DEARI* was found to be the most

common form of genetic alteration of the gene in cancer. Moreover, through database analysis of TCGA via cBIO, *DEARI* has also been found to be homozygously deleted in multiple tumor types including hepatocellular and breast cancer. Knudson's description of classical tumor suppressor inactivation requires two separate events to affect gene function³⁰. It is possible that one of these events may be inherited through the germline, making one more susceptible to cancer development³⁰. However, in most cases, both inactivating events occur after birth and are typically a combination of mutation and/or deletions. As such, *DEARI* has been described as both a classical tumor suppressor as well as a haploinsufficient tumor suppressor in a tissue dependent manner. In support of *DEARI* acting as a classic tumor suppressor in cancer, our lab had previously described secondary loss of the wildtype allele in *Dear1* heterozygous mice². Herein, additional evidence for *DEARI* acting as a classical tumor suppressor in human cancer is described with the report of tumors exhibiting combinatorial heterozygous loss of the *DEARI* locus as well as the presence of *DEARI* mutation. These alterations are important in cancer due to their ability to fully inactivate the gene and allow for singular expression of the mutant allele. Therefore, synergism of copy number changes of the *DEARI* locus and mutations may lead to alterations in gene dosage and function that can potentiate tumor progression.

It is well known that malignancy results from the accumulation of a diverse array of genetic alterations including mutations followed by the subsequent expansion of particular cellular clones that possess the greatest advantage at that particular time in tumor progression. The temporal nature of clonal expansion can hint at the importance of a particular set of mutations at a particular time in the timeline for tumor progression¹⁷⁷. For example, some mutations may be highly prevalent in the primary tumor, indicating that these mutations may be within founding clones and therefore important in tumor initiation. Other mutations may be more prevalent in the invasive leading edge of the primary or in the metastasis, indicating their importance in invasion or re-colonization of the metastatic lesion¹⁷⁸. Evidence has shown that 56% (n= 10/18; range= 22-100%) of *DEARI* variants with known allele frequencies within TCGA cohorts including glioblastoma multiforme, melanoma, stomach adenocarcinoma, and uterine corpus endometrioid carcinoma have variant allele frequencies greater than 20% (unpublished

data)^{145, 146}. The majority of these variants were predicted to be deleterious as shown by algorithms from MutationAssessor and PolyPhen2^{172, 179, 180}. These variants, if functional as many of them are predicted to be, represent a large proportion of altered cells that may have been evolutionarily selected for and could be important in the progression of the tumor.

Genetic alterations including copy number changes and mutations have a long history of being used to stratify for prognosis and therapy due to their relative ease of determination in clinical samples. One of the first to be developed was Imatinib which targets the *BCR-ABL* gene fusion caused by the t(9;22)(q34;q11) translocation¹⁸¹. Imatinib has been shown to be highly effective in Chronic Myeloid Leukemia (CML) with an overall 5 year survival rate at 89%¹⁸¹. Other *BCR-ABL* inhibitors which have been developed are Dasatinib and Nilotinib⁷¹. Crizotinib is also a targeted therapy which was developed to treat *Anaplastic Lymphoma Kinase* (ALK) translocations which are common in inflammatory myofibroblastic and anaplastic large cell lymphoma as well as less frequently in breast, colon, and lung¹⁸². Moreover, Trastuzumab and Lapatinib were developed for patients exhibiting *ErbB2* overexpression or amplification^{71, 183}. Vemurafinib has been created as a *B-RAF* mutant specific therapy for melanoma patients⁷¹. Other therapies known to be effective against activating mutations are Erlotinib and Gefitinib which target mutated *Epidermal Growth Factor Receptor* (EGFR) in Non-Small Cell Lung Cancer (NSCLC)⁷¹. Further, chromosomal loss of 9p21 and gains of 20q11 and 1q21 have been shown to stratify survival outcomes in patients with renal cell carcinoma¹⁸⁴. These are just a few examples of the genetic alterations that have been used to stratify patients for prognosis and treatment. Since *DEARI* was associated with a relatively frequency of chromosomal loss in invasive breast cancer and is a known suppressor of TGF β induced EMT, it was ascertained if *DEARI* heterozygous loss could cooperate with alteration of a downstream effector of the TGF β pathway. *DEARI* heterozygous loss and *SNAI2* amplification was shown to be able to act synergistically to predict overall worse survival in a TCGA invasive breast cancer cohort of 889 patients². Recent additions of patient samples into the TCGA invasive breast cancer cohort (n=959), has shown that not only is the combined heterozygous loss of *DEARI* and amplification of *SNAI2* still significant (p=0.016), but heterozygous loss of *DEARI* alone is

now able to predict overall survival ($p=0.042$) (unpublished data)^{145, 146}. This data suggests that *DEARI* and *SNAI2* can potentially act as a clinical marker panel to stratify patients for overall survival. The development of a FISH assay to determine if clinical samples harbor *DEARI* heterozygous loss and *SNAI2* amplification could assist in the determination of prognosis for invasive breast cancer patients. Clinical validation of these tests as well as clinical trials are needed to verify the ability of these markers to stratify patients. Altogether, the evidence presented has shown that *DEARI* is a driver in cancer that is inactivated via mutation and copy number alterations. Thus, it is possible in the future that we can capitalize on these alterations in order to better stratify patients that harbor these alterations for clinical prognosis.

Development and Performance Evaluation of a *DEARI* Ultra Deep Targeted Sequencing Assay for Ion Torrent Next Generation Sequencing Platform

Introduction

Next Generation Sequencing and Precision Medicine in Oncology Care

The advent of next generation sequencing (NGS) technology has revolutionized genomic science and molecular diagnostics, brought on by the highly sensitive, quantifiable, and high throughput nature of NGS. This technology allows for simultaneous detection of single nucleotide variants, copy number alterations, large structural changes like translocations and inversions, as well as the presence of small and large insertions/deletions (INDELS). Current sequencing applications that have been used on this platform include DNaseq via whole genome, exome, and targeted resequencing, as well as RNA-seq and miRNA-seq. It is believed that NGS will be a major enabler for the dawn of global “precision medicine”, which involves the selection of therapeutic strategies based on the individual molecular and genetic information of each patient¹⁸⁵. Currently it is recognized that targeted panel resequencing is the most clinic ready form of NGS technologies¹⁸⁶. Targeted NGS is more applicable for the clinic currently than other versions of DNaseq for multiple reasons including lower cost, faster output, better quality and coverage achievable via the locus enrichment strategy, and reduction in ethical issues connected with unsolicited findings of unanticipated germline inherited disease¹⁸⁷. Moreover, targeted resequencing panels have already been shown to be able to detect clinically relevant variants in multiple diseases and assist in the treatment of patients (see Rehm et al. 2013 for a comprehensive review¹⁸⁸). For example, Sehn and colleagues showed that targeted NGS panels can help in the clarification of diagnosis of oncology cases with ambiguous histology and may assist in redefining therapeutic approaches¹⁸⁹. Moreover, custom targeted panels have shown that 20% of castrate resistant prostate cancer harbor BRCA2 gene deletions and ATM point mutations which can be clinically important, as germline BRCA alterations have been shown to indicate sensitivity to PARP inhibitors in prostate, breast and ovarian

cancer¹⁹⁰. A 384 gene panel using Illumina based NGS found about half of the 44 gastric cancer samples sequenced had variants in TRAPP, which can affect the efficacy of HDAC and CHK1 inhibitors¹⁹¹. These studies have shown the usefulness of targeted panel based NGS and how this application can have a real impact on making “precision medicine” a reality.

Sensitivity of NGS Platforms is Essential for the Identification of Clinically Important Rare Variants

One major challenge to “precision medicine” that has been revealed by NGS is the great extent of intra-tumor heterogeneity. Gerlinger et al. showed in a seminal paper the existence of extensive regional heterogeneity in tumors, using renal cell carcinoma as a model, with only 31-37% of mutations being shared throughout the primary and the metastatic lesions⁵⁸. Similar evidence has been found for other tumors such as high grade serous ovarian cancer¹⁹². This regional heterogeneity can have large implications for clinical treatments. A study reported that relapsed Non-Small Cell Lung Cancer (NSCLC) lesions resistant to Epidermal Growth Factor Receptor (EGFR) tyrosine kinase inhibitors (TKI) due to *MET* amplifications were evolutionarily selected from a low frequency subclone (<1%) that existed prior to therapy¹⁹³. Moreover, Su and colleagues have shown that an *EFGR* variant T790M present at low frequency (<5% of tumor variant frequency) in a pre-treatment primary biopsy of NSCLC was found to be enriched in a post-treatment biopsy after treatment with an EGFR TKI¹⁹⁴. Rare variants, despite their low frequency within a tumor, can have deleterious functional effects on a gene product, and thus, given a particular time in the tumor’s developmental history, can play a potential role in tumor progression¹⁹⁵.

Next generation sequencing is unique in its ability to truly assess the extent of intra-tumor heterogeneity due to its high degree of sensitivity compared to other technologies and its quantifiable nature. For example, ABI’s SOLiD NGS platform was found to be able to achieve about 93% sensitivity and 100% specificity¹⁹⁶. Sikkema-Raddatz and colleagues found that targeted NGS by Sure Select capture and sequencing on Illumina’s MiSeq achieved 100% sensitivity (95% confidence interval: 97.76-100%) and a non-concordance rate of 0.00315% when compared to Sanger sequencing¹⁸⁷. Further, they

suggested high quality variants with adequate coverage (>30) may not need secondary validation by Sanger sequencing, as NGS technology seems to be superior to the previous gold standard. The sensitivity of NGS targeted resequencing technology has also been compared to real-time PCR's ability to detect variants. A high concordance between 96.3-100% was found to exist between the two technologies, with real-time PCR detecting an additional four known variants and NGS detecting eight novel variants not detected by real-time PCR in FFPE NSCLC tumor tissue¹⁹⁷. Moreover, Peter Campbell's group has shown the ability to detect down to a 0.02% minor allele frequency using a nested PCR approach followed by 454 sequencing¹⁹⁸. These assays have not only shown the high degree of sensitivity of NGS assays but have also shown their clinical utility. To that degree, the FDA has recently granted marketing authorization for Illumin'a MiSeqDx and Ion Torrent's Ion PGM Dx for clinical use based on its accuracy across the genome, along with its precision and reproducibility^{199, 200}. Moreover, in response, the College of American Pathologists has recently released its standards for the use of next generation sequencing in clinical tests. Further, there are over 80 clinical trials currently ongoing across the nation using NGS for oncology studies to assess and confirm the utility of this technology in cancer care (www.clinicaltrials.gov).

Methods

Design of Custom DEARI Targeted Ampliseq Panel

Ion Ampliseq Designer version 1.2.8 (<https://www.ampliseq.com>; Life Technologies) was used to design an Ampliseq panel to target the genomic region of chr1: 33,610,351-33,681,308 (hg19). The custom *DEARI* Ampliseq panel contained primers for a highly multiplexed amplification reaction to amplify 150bp fragments tiled across the genomic region of interest.

Creation of Ampliseq Spike-In for determination of Accuracy

To understand our ability to accurately detect allele frequencies of a known population, a spike-in assay was completed using variants with known frequencies to determine the capability of our custom *DEARI* ampliseq panel to detect accurate variant frequencies. Dr. Steven Lott designed the novel spike-

in assay used in this study. Four amplicons were selected for this validation test that resided in the linear phase of coverage during test sequencing of Human Mammary Epithelial Cell line-76NE6 (HMEC-76NE6). Coverage of amplicons from the sequencing of HMEC-76NE6 was plotted from lowest to highest and amplicons residing within the median range of coverage were chosen for use in the plasmid spike-in. An oligo was created featuring the full sequence of these four amplicons with barcode adapters at each terminal. Within each amplicon, marked changes to the reference sequence were made to make the amplicons contain various different types of variants that are known to frequently occur in sequencing data (Fig. 3; Appendix I). These variants included single and multiple nucleotide variants, as well as INDELS. The pMA-RQ vector plasmid containing the full sequence of the four amplicons with the manufactured variants and barcode adapters were generated using Life Technology's GeneArt gene synthesis.



Figure 3-Illustration of the Spike-in Plasmid. The plasmid contains the sequence of four amplicons with distinct artificial nucleotide changes including nucleotide substitutions and INDELS. The nucleotide changes occurred in both high complexity and low complexity (homopolymer) regions.

The artificial plasmid was then spiked into control CEPH (sample with Northern and Western European ancestry residing in Utah) DNA NA12878 (Coriell-NA12878) based on genome equivalents. Plasmid to cell line DNA ratios were determined as shown below:

A single human diploid cell has about 6.5 pg of DNA; therefore in a 200ng solution, there would be about 30,769 genomes (cells).

The quantity of total genomes took into account both plasmid and cell line DNA. Therefore, the quantity of total genomes for a 10% spike-in of the plasmid genomes also accounts for 90% of cell line genomes. Thus in this situation, 90% of the control NA12878 DNA accounts for 34,188 genomes and

10% of the artificial plasmid spike-in equates to 3,419 genomes for a total of 37,607 total genomes. However, the plasmid exists in a haploid condition and therefore needs to be doubled to reflect diploid conditions, as is represented in the control cell line DNA. Therefore, according to this, 6,838 genomes of the artificial plasmid is needed to equal 10% of the overall total genomes.

Mass of the plasmid to add was calculated as follows:

$$\frac{1894989.6g}{mol} * \frac{1 mol}{6.022^{23}} * 6838 plasmids = 2.15^{-14}g \text{ or } \frac{21.5fg}{100ul}$$

Because the overall mass of the spike-in plasmid was miniscule, the mass of the plasmid in the solution was not figured into the dilution of the control DNA.

Dilutions of artificial plasmid DNA into NA12878 control DNA were made for 10%, 5%, and 1%. Additional dilutions were also made for 50%, 20%, and 0.5%. All dilutions were checked by PicoGreen Kit (Life Technology P7589) and TaqMan RNase P Detection Reagents Kit (Life Technology 4316831) for validation of concentrations and then sequenced on either the Ion Torrent PGM or Proton sequencer. The PicoGreen assay was completed exactly as protocol stated (<https://tools.lifetechnologies.com/content/sfs/manuals/mp07581.pdf>). TaqMan RNase P kit was preformed according to the protocol (“Measuring Template Efficiency” <https://tools.lifetechnologies.com/content/sfs/brochures/cms042785.pdf>). Briefly, the standard Human genome DNA provided in the kit was serially diluted from 10ng/μl to 0.5ng/μl and each serial dilution was used to create a standard curve. For the 96 well PCR plate format, to each well was added: 3ul of FFPE DNA or diluted standards, 12.5ul 2 × Universal Master Mix (No AmpErase UNG) (Life Technologies 4364343), 1.25ul 20x RNase P Probe/Primer Mix (from kit 4316831 Life Technologies), and 8.25ul RNase-free water. qPCR cycling was performed as follows: 50° Celsius (C) 2 min, 95°C 10 min, and 40 cycles of 95°C 15 sec and 60°C 1 min. Using a standard curve amplification, quantities were determined for samples. Correlation of logarithmic best fit line of CT values of standards (Y axis) vs Concentration (X axis) was used to determine accuracy of quantities, with R² typically being over

0.9900. Amplicon sequences were manually pulled out of bam files using the Extract Barcodes R script (Appendix II).

Ampliseq Library Construction and Sequencing

In order to amplify *DEARI* for sequencing, Ampliseq Designer version 1.2.8 (Life Technologies <https://www.ampliseq.com>) was used to design 150bp amplicons, spanning across a 48kb region on chromosome 1: 33,610,600-33,658,985 (hg19). The 184 amplicons, split into two pools, were used to complete library construction according to the Ion Torrent library construction protocol (Life Technologies MAN0006775 Revision 4.0 <http://ioncommunity.lifetechnologies.com/docs/DOC-3254>). Briefly, the Ampliseq panel was used to amplify *DEARI* in a highly multiplexed manner using 10ng of DNA for each primer pool as well as the Ion Ampliseq Library Kit 2.0 (Life Technologies 4475345). Specifically the FFPE DNA was added with 10ul of the 400nM 2x Ion Custom Ampliseq Primer Pool, and 4ul of 5x of the Ion Ampliseq HiFi Master Mix in each reaction. The following thermocycling protocol was used: 1 cycle at 99°C for two minutes followed by 19 cycles of 99°C for 15 seconds and 60°C for four minutes. The primer sequences were then partially digested by the addition of 2ul of FuPa reagent to each samples that were then exposed to the following conditions: 50°C for 10 minutes, 55°C for 10 minutes, and then 60°C for 20 minutes. After the primer sequences were digested, the Ion Express Barcode Adapters were added (Life Technologies 4471250), in addition to Switch solution and DNA ligase, to the amplified library and subjected to the following temperature conditions: 22°C for 30 minutes followed by 72°C for 10 minutes. The completed library was then purified with the Agencourt AMPureXP magnetic beads done according to Beckman Coulter protocol, with scaled volumes to fit the input volume (Beckman Coulter protocol # B37419AA).

The purified libraries ran on the Ion Torrent PGM were amplified using a Platinum PCR SuperMix High Fidelity polymerase and a library amplification primer mix, using the following conditions: 98°C for 2 minutes followed by 5 cycles of two steps including 98°C for 15 seconds and 60°C for 1 minute. The amplified library was then re-purified with Agencourt AMPureXP magnetic

beads using the same conditions during the previous library purification as before (see above). For the samples run on the Ion Torrent Proton, libraries were not amplified after purification. Once final purified libraries were created, the two libraries associated with the 2 different Ampliseq primer pools were added together in equal concentrations for each patient sample. Templates were then created with the 9x diluted pooled libraries by following the Ion PI and PGM Template OT2 200 Kit V2 protocol (Life Technologies 4480974 and 4485146; MAN0007624 <https://tools.lifetechnologies.com/content/sfs/manuals/MAN0007624.pdf>). Briefly, the following reagents were added, at the specified amounts, to 65µl of the diluted library: Ion PI Reagent mix TL (750µl), Ion PI PCR Reagent B (450µl), Ion PI Enzyme Mix TL (75µl), Ion PI PCR Reagent X (60µl), and the Ion PI Ion Sphere Particles (100µl). Samples were then run on the Ion OneTouch 2 instrument. Template positive Ion Sphere Particles (ISPs) were recovered by centrifugation and then ~1.4mL of supernatant was removed. Beads were then washed twice with ~ 1.1mL of water and the pellet was resuspended in Ion PI ISP Resuspension Solution. Quality was assessed by Qubit 2.0 fluorometer. The template-positive ISPs were then enriched on the Ion OneTouch ES by the attachment of ISPs to washed streptavidin C1 beads. The enrichment was completed by following the protocol as described in the Ion PI Template OT2 200 kit v2 manual: after the attachment of the ISPs to beads, subsequent washes as well as a melt-off reaction occurred on the OneTouch ES. Melt-off solution contained 40µl of 1M NaOH, 3 µl of 10% Tween 20 and 277µl of water.

Enriched ISPs were then prepared for sequencing using the Ion PI or PGM Sequencing 200 Kit V2 (Life Technologies 4485149 and 4482006) as described by the Ion PI and PGM Sequencing 200 kit v2 protocol (MAN0007961 and MAN0007273 <http://ioncommunity.lifetechnologies.com/docs/DOC-7459> and <http://ioncommunity.lifetechnologies.com/docs/DOC-6775>). Briefly, the enriched ISPs were washed with water, followed by the addition of Ion PI Annealing buffer (15µl) and Ion PI Sequencing Primer (20µl) to the enriched ISPs and then amplified on a thermocycler with the following steps: 95°C for 2 minutes and 37°C for 2 minutes. After amplifying the ISPs, 10µl of the Ion PI Loading buffer was added and the final mixture was loaded onto a washed and calibrated Ion PGM 316 or Proton chip. The

chips were then washed with a foam created from 45µl of 50% Annealing buffer and 5µl of 2% TritonX-100, followed by the loading of 55µl of 50% Annealing buffer. This was followed by the chips being centrifuged and flushed with Flushing solution and 100µl of 50% Annealing buffer. Ion PI Sequencing Polymerase was then flushed across the chip in a solution with 50% Annealing buffer. Upon completion, samples were then sequenced on Proton and PGM316 chips using the Ion PI and PGM 316 Chip Kit (Life Technologies 4482321 and 4483324). For samples barcoded on the Proton chips, Ion Torrent Suite was used to separate barcoded sequences and exported as individual BAM files.

Sequence Analysis

Sequencing bam files were analyzed by Torrent Variant Caller v.4.2 via customized *DEAR1* parameters, followed by its output as a Variant Calling Format (VCF) file²⁰¹. The customized DEAR1 parameters consisted of recommended Ion Torrent variant calling parameters with the following changes: Splice Site Size-20bp; Maximum Coverage-100,000; Data Quality Stringency- 8.5; Downsample to Coverage- 100,000; Snp Minimum Coverage Each Strand-7; Snp Minimum Allele Frequency-0.01; Snp Minimum Coverage-20; SNP strand bias-0.90; Mnp Minimum Coverage Each Strand-7; Mnp Minimum Variant Score-400; Mnp Min Allele Frequency- 0.01, Mnp Minimum Coverage-20; Mnp Strand bias- 0.90, Kmer length-19, Minimum Frequency of Variant Reported-0.15; Indel Minimum Variant Score-10; Indel Minimum Coverage Each Strand-3; Indel Minimum Allele Frequency- 0.05; Indel Minimum Coverage-20; and Indel Strand Bias-0.85. Sample VCF files were annotated by SnpEff/SnpSift using the tool's inherent annotation as well as with Phase 1 Version 3 data from the 1000 Genomes project²⁰²⁻²⁰⁴. SnpSift was also used to complete variant filtering in accordance with the following parameter requirements: quality score was required to have a greater than 30 phred score, as well as at least 1,000x overall locus coverage, 60x or greater coverage of the variant allele, and a minimum of 30x coverage on both forward and reverse strands. Stringency filters allowed for 13-66% reduction in incidental variants as well as the retention of noteworthy variants (Appendix III). Variants identified in both the tumor and normal samples were filtered out. However, a small number of potentially germline variants (allele frequencies around 50% or 100%) were found to exist in samples to which the matched normal sample

was unable to be sequenced. These unconfirmed potential germline variants were not filtered out of our analysis as there was no definitive proof these were in fact germline; therefore, our total variant count is slightly inflated by these variants. Genome Analysis ToolKit was used to determine overall coverage²⁰⁵. The R program for statistical computing packages' ggplot2 were used to create figures detailing sequencing data, including coverage plots (Appendix IV)^{202, 203, 206, 207}.

Digital PCR

Digital PCR was completed on either the QX200 Droplet Digital PCR System (Bio-Rad 186-4001) or QuantStudio 3D (Life Technologies 4489084) platform. Custom genotyping TaqMan (4331349) was ordered from Life Technologies using a dual reporter system with VIC fluorescence tagging the reference sequence and FAM tagging the mutant sequence. PCR reaction mix was created using dPCR master mix (Life Technologies 4482710), target probes, and diluted DNA (10-36ng) per manufacturer's suggestion. TaqMan genotyping probes were first optimized for largest cycle separation between the VIC and FAM probes with at least 1.5x cycle difference using qPCR with the following conditions: 95°C for 10min with 40-50 cycles of 92°C for 15sec and 60-64°C for 1 min. For Bio-Rad, samples underwent droplet generation (Bio-Rad 186-4002) whereas the samples undergoing digital PCR on the Life Technologies platform were spread across a microchip featuring 20,000 wells using an autoloader (Life Technologies 4485507 and 4482592). PCR was performed according to manufacturer's protocol with specified cycle number and annealing temperatures as determined in preliminary qPCR optimization. Data was then analyzed by QX200 droplet reader (Bio-Rad 186-4003) or by QuantStudio 3D (Life Technologies 4489084). Each variant was tested with 2-3 replicates. The R program for statistical computing packages' ggplot2 were used to create figures detailing Digital PCR data (Appendix V)^{202, 203, 206, 207}.

Read Count Accuracy

To determine the read count accuracy, the approximate number of cells present with a particular variant given a predicted allele frequency was determined. The model is based on assuming

diploid cell condition and extrapolated to 1,500 cells in the 10ng needed for Ion Torrent Ampliseq targeted sequencing based on the following calculations:

3 million bp (human genome) * 2 (diploid) * 660g/mol (Avg MW dNTP) * 1.67*10⁻¹² (weight of 1 dalton)= 6.6pg or 0.0066ng

10ng input into Ampliseq amplification/0.006ng = 1,515 cells

The percentage of cells with a variant was determined by the percentage of heterozygous cells needed to produce a given variant allele frequency based on the presence of 1,515 cells within the 10ng of input DNA. Further, the number of reads expected to be positive for a variant under a specified coverage was determined. For comparison to our actual sequencing data, examples of read counts for the particular allele frequency were given along with the experimental coverage received. Read coverage accuracy was calculated by the following equation:

$$\frac{(\text{Predicted read count for a given allele freq} \times \text{Experimental coverage for given allele freq})}{(\text{Experimental read count for given allele freq} \times \text{Predicted coverage for a given allele freq})} \times 100$$

Read counts and allele frequencies used to calculate read coverage accuracy originate from VCF files of the FFPE DCIS samples sequenced. VCF files were created through the use of the Torrent Variant Caller 4.2 and the specific parameters mentioned in the previous method section detailing the sequence analysis. It is important to note that the read count accuracy was completed by manual determination and was not used as an empirical method to quantify the accuracy of the variant caller or sequencing platform.

Sensitivity and Specificity

To determine the sensitivity and specificity of the custom Ampliseq *DEARI* panel, VCF files of the lymphocyte DNA of normal Caucasian female NA12878 from the Coriell (Coriell NA12878) repository created using the *DEARI* Ampliseq panel on the Ion Torrent Proton and PGM platforms was

compared to the National Institute of Standards and Technology's (NIST) highly confident integrated genotype calls of the same sample²⁰⁸. VCF files created from the Ion Torrent Proton and PGM platforms from our sequencing of NA12878 were generated via use of the automated Torrent Variant Caller according to parameters previously discussed in the sequence analysis section of the methods within this chapter. Moreover, the VCF files underwent additional variant filtration completed according to the protocols discussed within the sequence analysis section of the methods. Samples were noted for the absence or presence of variants (not including common single nucleotide polymorphisms) known to occur in the NIST's NA12878 genotype calls as well as for variants in randomly chosen and randomly distributed nucleotides across the *DEARI* locus within areas covered by the Ampliseq. To determine if the degree of sensitivity and specificity was altered in homopolymers, samples were also noted for the absence or presence of variants within randomly chosen homopolymer areas across *DEARI*. Using two by two contingency tables and the package "epiR" within the statistical software R, sensitivity and specificity was calculated as well as the 95% confidence interval for each factor (Appendix VI)^{209, 210}.

Results

Determination of Appropriate NGS Platform for Experimental Procedure

Multiple platforms exist for NGS applications, including Illumina's MiSeq and HiSeq, Roche's 454 GSJ system, ABI's SOLiD platform, and the Ion Torrent system. The vast majority of these systems use labeled di-deoxy terminating nucleotides and imaging software to capture the incorporation of the labeled nucleotide. The Ion Torrent platform is unique in the way it detects nucleotide incorporation. This platform takes advantage of the natural chemistry that occurs upon the bonding of two nucleotides by detecting the release of proton molecules after the flow of specific nucleotides. This method bypasses some of the issues associated with image analysis many of the other NGS platforms have, like "dead" fluorophores and overlapping signals²¹¹. Each platform has its own benefits and negatives, and each situation must be considered to determine which platform is best for the experimental design. Illumina platforms have higher throughput capability and thus lower sequencing

cost per gigabyte, but greater instrument costs and longer run times from 27hrs to 11 days on the MiSeq and HiSeq, respectively²¹². Their competitor, the Ion Torrent PGM sequencer platform is itself lower in cost and has a much lower run time, only 2 hours²¹². Further, each platform is associated with its own INDEL and substitution error rates^{213, 214}. The 454's GSJ and the Ion Torrent PGM (using the 200bp sequencing kit) exhibit primarily INDEL errors and have a similar INDEL error rates per 100 basepairs (bp) at 0.40. The MiSeq platform has the lowest INDEL error rate per 100 bp with a rate of only 0.0009. The MiSeq is known to exhibit mostly substitution errors with an error rate of 0.09 per 100 bp. On the other hand, the 454's GSJ had the second lowest substitution rate at 0.05 and the Ion Torrent PGM (using 200bp sequencing kit) had the lowest rate at 0.03. Depending on the type of variant that is most likely expected in the experiment, the variant specific error rate corresponding to each type of platform can be very important in choosing which platform is best for the experiment planned.

Our lab was interested in finding novel variants and determining the variant frequency in an early stage of breast cancer. Previous analysis of genetic alterations of *DEARI*, described in chapter 2, has shown that the majority of reported variants harbored within the gene are substitutions. Deletions involving the gene in cancer tend to encompass not only the entire gene itself but large regions of the 1p chromosome arm as well. As these large deletions can be found easily by other less expensive techniques than NGS and has been extensively described in cancer, our lab decided to focus on characterizing the substitution variants within *DEARI* in early stage breast cancer using NGS technology. Further, current sequencing efforts of breast cancer have experienced reduced sensitivity due to intra-tumor heterogeneity, normal tissue contamination within the samples, and lower median coverage rates, all reducing the degree of sensitivity to find rare alleles. By completing ultra-deep sequencing of *DEARI*, we can potentially achieve greater sensitivity to detect those rare variants for which previous lower coverage sequencing may not have been able to discover. Ultra deep sequencing has previously been shown to be effective in detecting low prevalence somatic variants, mostly through amplicon based sequencing^{215, 216}. The Ion Torrent platform was determined to be the appropriate fit for our experiment

because of its low substitution error rate, low cost, quick processing time and the proven history of amplicon based sequencing being able to detect rare somatic variants.

Design of Ion Torrent Custom DEAR1 Targeted Ampliseq Panel

The Ion Torrent System as described above is centered on an amplicon based strategy in which a large number of primers are designed to amplify small 150-200bp regions across an area of interest, thereby amplifying a large region in a tiled based manner. Due to the complexity of designing primers for large multiplexing reactions as required for the Ampliseq panel, it is best to use computer based algorithms that take into account GC-content, melting temperatures, primer-primer interactions, etc. ThermoFisher Life Technologies offers a free web-based Ampliseq design service that uses as input either a specified genomic range or a genes list. Using Ampliseq Designer version 1.2.8, a custom Ampliseq panel was designed to amplify a 71kb genomic region that included the reading frame of *DEAR1* as well as a 33kb region upstream of the canonical promoter of *DEAR1* to include variants that might influence the expression of the gene (Fig. 4). However, due to low complexity of the locus, a final 184 amplicon panel was designed to cover a 48kb region which includes the *DEAR1* locus. In general, there is a high concordance between the areas that were unable to be included in the coverage by the Ampliseq panel and the areas that are known to be genetically repetitive. In fact, the vast majority of the 33kb region upstream, as well as much of the intronic regions of *DEAR1*, were unable to be amplified due to their highly repetitive nature (Fig. 4b). When focusing on the *DEAR1* locus, including intronic regions and a 11kb region upstream of *DEAR1*, 47% coverage longitudinal coverage was achieved. Of the non-repetitive areas, 100% coverage of the canonical promoter of *DEAR1* as well as the exonic regions and 75% coverage of the 3'UTR was achieved. Areas able to be amplified by the *DEAR1* custom Ampliseq panel include regions known to harbor variants previously identified in various cancers and catalogued in the COSMIC depository and cBio^{145, 146, 148}. These variants include the D421G mutation found in a TCGA stage IIIC rectal cancer patient who is listed by the project as having progressed and a frameshift at P270fs in a stage 2 breast cancer patient with positive lymph node involvement^{145, 146}.

Control Artificial Spike-in

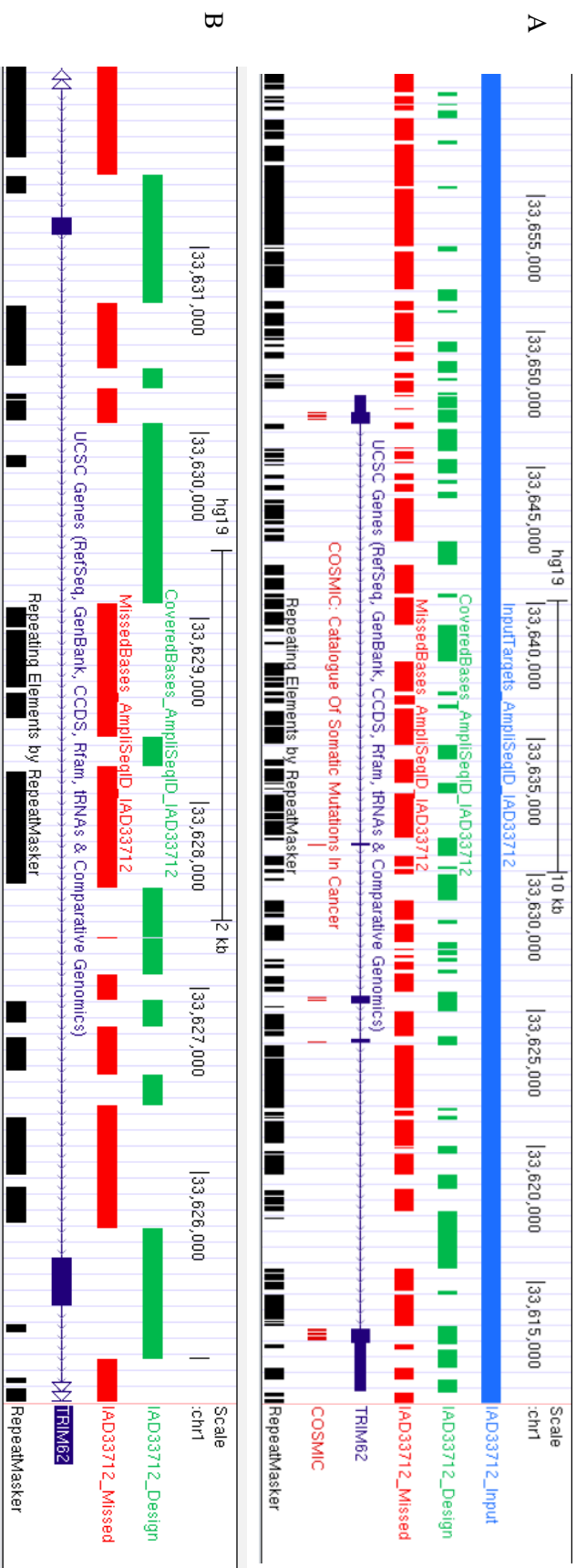


Figure 4- Illustration of Custom *DEAR1* Ampliseg Design. A genomic region of chr1:33,610,580-33,659,680 was submitted to Life

Technologies' Ion Ampliseg Designer (as shown by the blue IAD33712_Input UCSC Genome Browser track). A) The genomic regions in which Ampliseg oligos were able to be designed are indicated by the green IAD33712_Design UCSC Genome Browser track. Regions of the submitted genomic locus that were unable to be covered by the Ampliseg oligos are indicated by the red IAD33712_Missed UCSC Genome Browser track, which mostly corresponds to the low complexity regions as shown by the RepeatMasker UCSC Genome Browser track. Areas that were able to be covered included variants previously described by the COSMIC database, shown by the red COSMIC UCSC Genome Browser track. B) A zoomed in capture of the Ampliseg designs highlights correspondence between areas unable to be covered by the design and repetitive elements.

Interpretation of NGS data relies on the idea that a variant frequency is accurately represented in the NGS reads to the same extent the variant is represented in the sample itself. Understanding this degree of accuracy and having empirical data to bolster this idea is very important for one to have confidence in the variant data that the Ampliseq panel accrues. For this reason, a novel spike-in assay was designed to determine the precision of our variant frequency read outs. This was completed by the designing of an artificial plasmid containing 4 amplicons from our *DEARI* custom Ampliseq panel featuring inserted synthetic variants within these amplicons reflecting nucleotide variants as well as INDELS (Fig. 3). This plasmid was then spiked into control normal DNA NA12878 at various frequencies and sequenced. The results showed that for most of the amplicons, the sequencing data was able to accurately recapitulate the variant frequencies spiked into the control DNA (Fig. 5). Some amplicons were able to outperform others. For example, control C amplicon had 100% accuracy for the 1% and 5% spike-in and less than 0.5% deviation for the 10% spike-in. Control A and control B amplicons had approximately 100% accuracy for the 1% spike in, whereas control D amplicon exhibited about a 1% deviation for the 1% spike-in. Control A and control D amplicons showed slight deviation from expected allele frequencies, within 1.5% frequency, for the 5% spike in. Moreover, control B amplicon showed significant deviations from expected variant frequencies in the 5% (2.5-4%) and 10% spike-ins (5-8%), despite no relatively reduced coverage of the amplicon in relation to the other amplicons. Control A and control D amplicons showed a 2-2.5% deviation from expected frequency for the 10% spike in. In support of our findings, library construction processing of DNA samples used in NGS sequencing have been shown to have a slight effect on variant frequency read-outs in sequencing data (personal communication from BioRad). Overall, these data reflect the variability of the accuracy of sequencing data and how this may affect the precision regarding variant allele frequencies.

Determination of Read Count Accuracy

An important innovation of NGS technology is its ability to be quantitative regarding allele frequencies of a variant within a sample. This is accomplished by determining the ratio of the read count of the variant to the total coverage of the locus. The use of this innovative approach has recently revealed

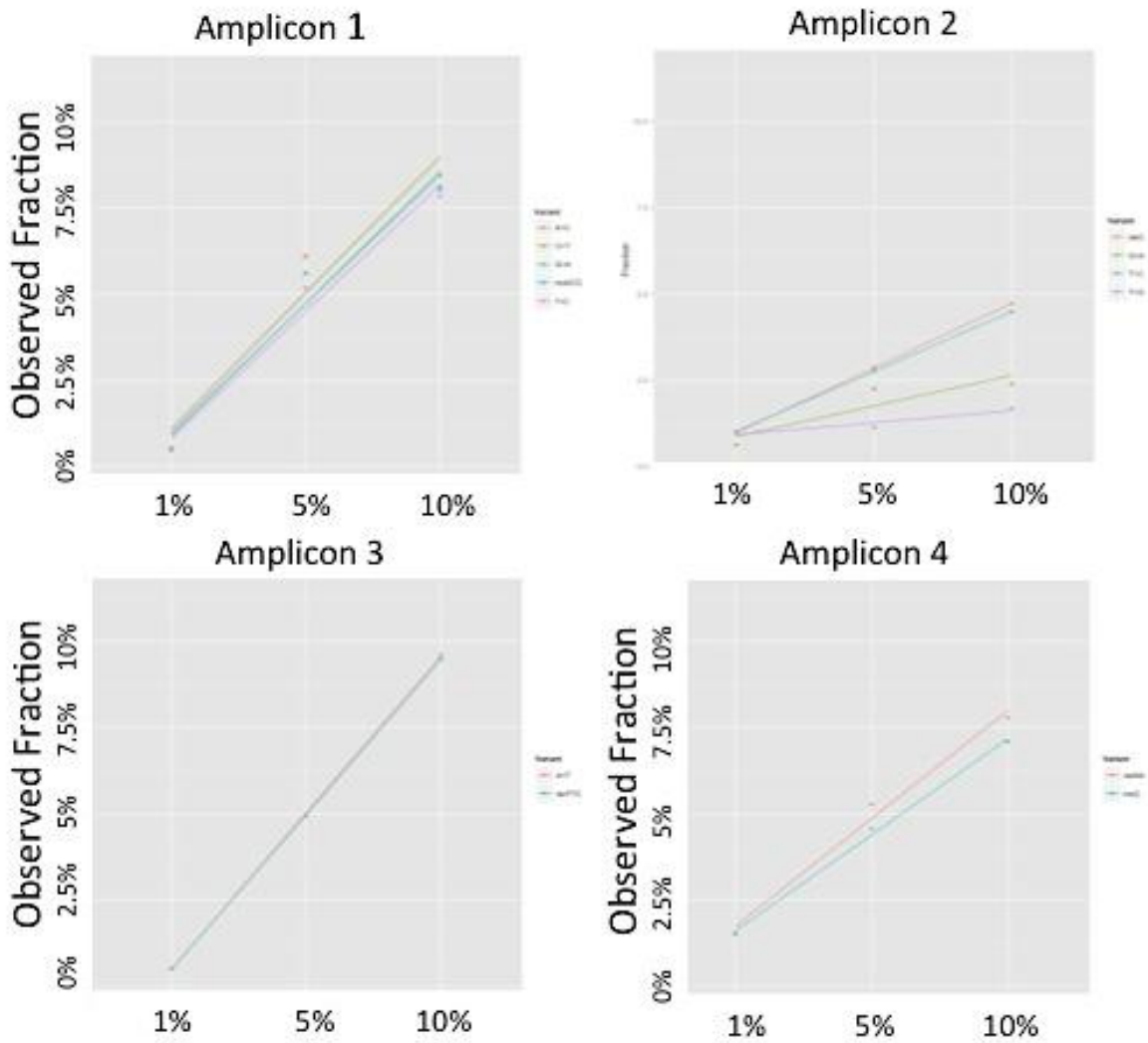


Figure 5- Comparison of the Artificial Spike-In Frequency to Observed Allele Frequency in Next Generation Sequencing. The artificial plasmid detailed in Figure 2 was spiked in at 1%, 5%, and 10% (x-axis). Figure details the correlation of the spiked-in frequency to the observed fraction of sequence reads. Figure was created with help from Dr. Steven Lott, Ph.D.

the large degree of intra-tumor heterogeneity within cancer samples, which can have major implications for therapeutic approaches²¹⁷. Heterogeneity is often identified by the presence of variants within small sub-clonal populations. The sub-clonal populations can be quantified by the determination of variant allele frequencies of specific variants marking these populations. Allele frequencies of these variants are determined by the fraction of supporting read counts carrying the variant in relation to the total sequencing read counts of the locus. Therefore it is important to determine the accuracy of the read counts achieved by the targeted sequencing panel per given variant allele frequencies. Thus, in order to determine the read count accuracy of the sequencing Ampliseq panel, achieved sequencing read counts were compared to the number of read counts that would be predicted based on a given allele frequency in a diploid model (Table 5). Further, determining the number of cells carrying a variant based on NGS allele frequency can be very important clinically if clinicians base therapeutic strategies on the potential sensitivity of a patients' tumor to a therapy against that particular genetic alteration, and important to this discussion, the variant frequency within the tumor based on the read out by NGS technology. As an example to show how allele frequencies can relate to the number of cells positive for a variant for a given sampled population, the frequency of cell populations harboring a variant based on different sequencing allele frequencies was modeled (Table 5).

As shown by Table 5, a heterozygous variant detected by NGS at an allele frequency of 25%, is actually harbored within 50% of the population sampled. Further, extrapolating that about 1,515 cells are contained within the 10ng sampled by the Ampliseq procedure, approximately 758 cells within the sampled population are carrying the variant. In the previously given example, the allele frequency given from the NGS read count determined that the variant was present in about a quarter of the sampled population of cells. However, with knowledge that the variant existed in a heterozygous state and at a loci that is diploid, it can be extrapolated that the variant actually is harbored in half the cells, a point that can be important clinically as this can affect the percentage of the tumor that may be affected if an oncologist were treating with a targeted therapy against this particular variant. Therefore, zygosity of the variant as well as copy number of the loci carrying the variant are very important determinants of the

Predicted Allele Frequency	Cell Count %		# of Cells with Variant	Predicted Variant Positive Read Count				Example of Actual Read Counts In Sequence Data
	(A/A)	(A/B)		Assume ~200x coverage	Assume ~1000x coverage	Assume ~9000x coverage	Assume ~25000x coverage	
50%	0%	100 %	1500	100 reads	500 reads	4500 reads	12500 reads	7315 reads (14626x coverage) 99.97% accuracy
40%	20%	80%	1200	80 reads	400 reads	3600 reads	10000 reads	3741 reads (9423x coverage) 99.25% accuracy
30%	40%	60%	900	60 reads	300 reads	2700 reads	7500 reads	2206 reads (6575x coverage) 89.42% accuracy
25%	50%	50%	750	50 reads	250 reads	2250 reads	6250 reads	1293 reads (5151x coverage) 99.59% accuracy
20%	60%	40%	600	40 reads	200 reads	1800 reads	5000 reads	111 reads (551x coverage) 99.28% accuracy
10%	80%	20%	300	20 reads	100 reads	900 reads	2500 reads	291 reads (2942x coverage) 98.91% accuracy
5%	90%	10%	150	10 reads	50 reads	450 reads	1250 reads	120 reads (2434x coverage) 98.60% accuracy
1%	98%	2%	30	2 reads	10 reads	90 reads	250 reads	48 reads (1.5% variant frequency) (3322x coverage) 96.33% accuracy

Table 5- Accuracy of Sequencing Data as Determined by Comparison to Predicted Read Counts. Predicted read counts were modeled based on differential coverage and allele frequencies, using a diploid cell model. Positive variant cell populations were also determine based on allele frequencies as well as the cell number within the 10ng used for Ion Torrent Ampliseq sequencing. The accuracy of our experimental sequencing read counts were determined as well in comparison to the predicted read counts.

abundance of a particular variant in terms of determining the number of cells positive for a variant in a given sampled population.

Allele frequencies, as previously mentioned, are a determinant of the ratio of the number of unique reads of the alternative allele to the total number of unique reads of the locus (coverage). The coverage of the locus helps to determine the sensitivity of the sequencing in detecting rare alleles as these rare alleles will only be represented in a very small proportion of reads. Using our model, for instance, sequencing of a rare 1% variant in a NGS sequencing reaction that obtained about 200x coverage of the locus, will only attain 2 supporting reads for the variant. A 5% variant, also sequenced at 200x coverage, will be detected in 10 reads. In contrast, a 1% variant sequenced under 9,000x coverage will be represented in 90 reads, giving large support for this rare allele. Variants receiving low supportive reads may be potential errors as library creation and next generation sequencing have inherent error rates which can induce false positive nucleotide changes that can be identified in sequencing data at various frequencies, including low frequencies which may be misinterpreted as a rare alleles. Thus, strict filtering methods need to be developed in order to identify potential errors which can be based on multiple factors including the sequencing platform error rate, sequencing depth, coverage of alternative allele, and quality score.

To determine the precision of our read counts using our custom Ampliseq panel, the accuracy of the read counts in accordance with the predicted read counts for a given allele frequency was determined. The mean sample coverage across the Ampliseq panel was 29,054x. Results indicate that the sequencing coverage was detected at frequencies higher than above clinical 1000x standards and all variants were supported by at least 60 reads (Fig. 6). Across the *DEARI* locus, consistency of overall coverage varied, with amplicons in exon 3 exhibiting the lowest degree of variation and amplicons in exon 1 and the promoter exhibiting the highest degree of variation (Fig. 6). However, variant read counts within our sequencing data, when compared to the expected reads for a given allele frequency and locus coverage, reflected high degrees of accuracy with a median accuracy of 99.08% (range 96.33%-

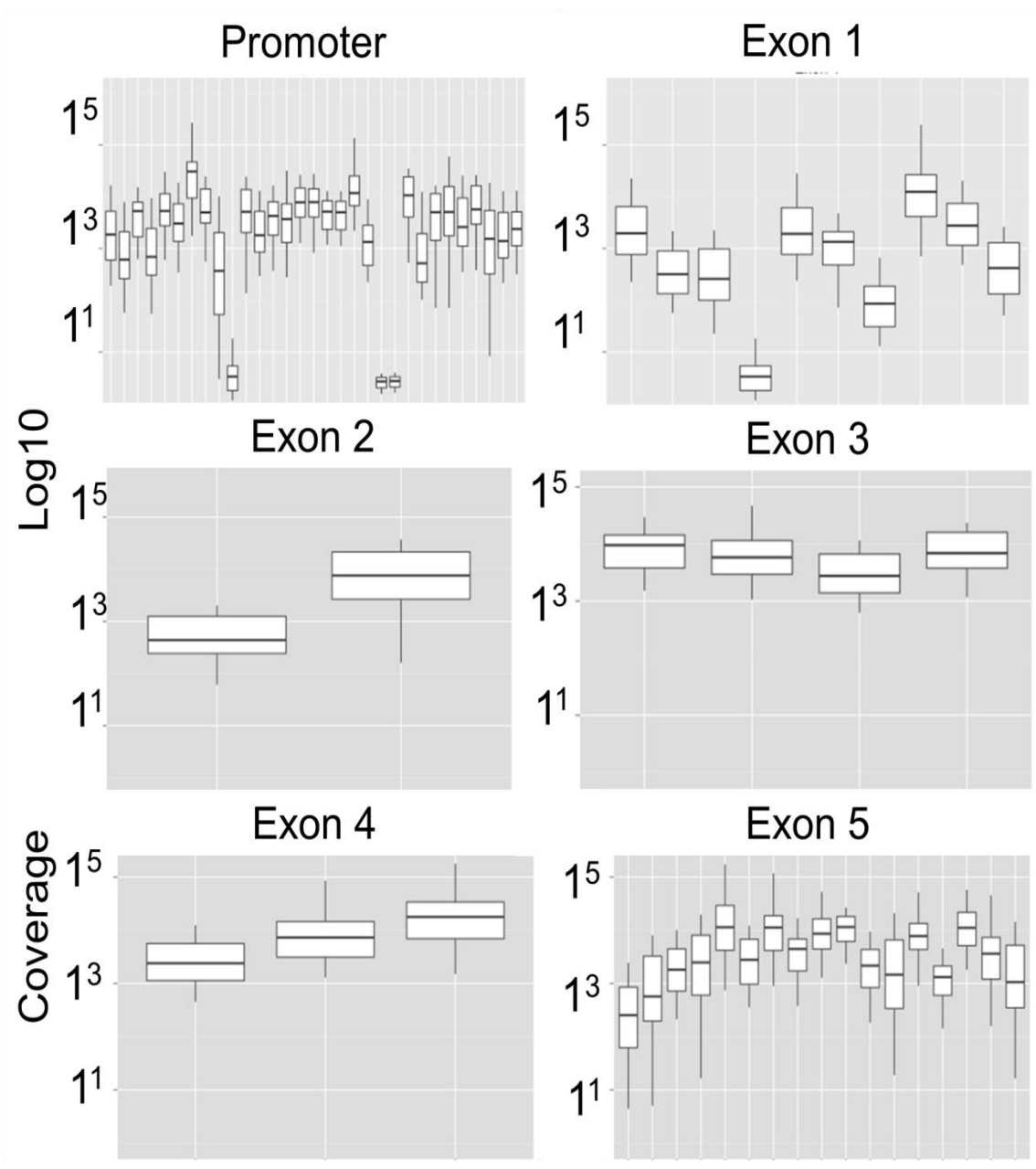


Figure 6- Coverage of the *DEARI* Ampliseq Panel Amplicons Within the *DEARI* Promoter and Exonic Regions. Figure represents the coverage obtained by the *DEARI* ampliseq panel within the 5'UTR and exonic regions of the gene. Each bar represents a single amplicon.

Figure was created with help from Dr. Steven Lott, Ph.D.

99.97) (Table 5). Strong accuracy was even shown for variants reflective of 1% of the population, indicating the efficacy of our Ampliseq panel to detect rare variants.

Sensitivity and Specificity of the Custom DEARIAmpliseq Panel

In order to fully understand the capacity of what is captured within one's sequencing data, it is important to understand the sensitivity and specificity of the targeted sequencing panel. This helps to understand the degree of false-positive variants that may occur within the data as well as the number of true variants that may not be reflected in the sequencing data, also known as "false-negatives". To determine the sensitivity and specificity of the *DEARI* targeted Ampliseq panel, sequencing data of the normal control DNA NA12878 completed on our Ion Torrent PGM and Proton platforms was compared to the NIST's highly confident integrated genotype calls for the same sample (Table 6). Using the NIST's data set as "truth", sequencing on the PGM platform showed a high degree of overall sensitivity and specificity, with a sensitivity of 0.92 (CI:0.73-0.99) and specificity of 0.93 (CI:0.81-0.99). The Ion Proton platform performed similarly with an overall sensitivity of 0.96 (CI:0.71-1.00) and specificity of 0.86 (CI:0.71-0.95). Both platforms showed low sensitivity to detect INDELs, as is well known not only for the Ion Torrent platforms but for other NGS platforms as well [sensitivity 0.5 (CI: 0.01-0.99) and specificity 0.95 (CI: 0.74-1.00)]²¹². When focusing on the performance of the sequencing platforms in their detection of substitutions, the main type of variant expected to be harbored in *DEARI* based on previous sequencing data, the platforms showed high sensitivity and specificity to accurately detect these type of variants [PGM: sensitivity 0.95 (CI: 0.77-1.00) and specificity 0.91 (CI: 0.72-0.99); Proton: sensitivity 1.00 (CI: 0.78-1.00) and specificity 0.78 (CI: 0.56-0.93)]. Both platforms also showed a high degree of sensitivity in their abilities to detect substitutions in homopolymer regions [sensitivity 1.00 (CI: 0.28-1.00); PGM specificity 0.88 (CI: 0.62-0.98), Proton specificity 0.75 (CI: 0.56-0.93)]. The platforms both performed very well in FFPE tissue in a similar way to the cell line NA12878 DNA, indicative of the custom *DEARI* ampliseq panel's ability to have a high degree of precision in clinical samples [PGM: sensitivity 0.95 (CI: 0.92-0.98) and specificity 0.92 (CI: 0.89-0.94); Proton: sensitivity

Samples	“Truth”	“Test”	Variant Type	Sensitivity (95% CI)	Specificity (95% CI)
NA12878 v. NA12878	NIST	Ion Torrent Proton	All combined variants	0.96 (0.79, 1.00)	0.86 (0.71, 0.95)
			Indels	0.50 (0.01, 0.99)	0.95 (0.74, 1.00)
			Indels in homopolymers	0.50 (0.01, 0.99)	1.00 (0.64, 1.00)
			Substitutions	1.00 (0.78, 1.00)	0.78 (0.56, 0.93)
			Substitutions in homopolymers	1.00 (0.28, 1.00)	0.75 (0.48, 0.93)
NA12878 v. NA12878	NIST	Ion Torrent PGM	All combined variants	0.92 (0.73, 0.99)	0.93 (0.81, 0.99)
			Indels	0.50 (0.01, 0.99)	0.95 (0.74, 1.00)
			Indels in homopolymers	0.50 (0.01, 0.99)	1.00 (0.64, 1.00)
			Substitutions	0.95 (0.77, 1.00)	0.91 (0.72, 0.99)
			Substitutions in homopolymers	1.00 (0.28, 1.00)	0.88 (0.62, 0.98)
NA12878 v. NA12878	Ion Torrent Proton	Ion Torrent PGM	All combined variants	0.86 (0.68, 0.96)	1.00 (0.86, 1.00)
	Ion Torrent PGM	Ion Torrent Proton	All combined variants	1.00 (0.80, 1.00)	0.90 (0.77, 0.97)
	Ion Torrent Proton	Ion Torrent PGM	Indels*	1.00 (0.09, 1.00)	1.00 (0.75, 1.00)
	Ion Torrent Proton	Ion Torrent PGM	Indels in homopolymers*	1.00 (0.01, 1.00)	1.00 (0.66, 1.00)
	Ion Torrent Proton	Ion Torrent PGM	Substitutions	0.85 (0.66, 0.96)	1.00 (0.74, 1.00)
	Ion Torrent PGM	Ion Torrent Proton	Substitutions	1.00 (0.79, 1.00)	0.82 (0.60, 0.95)
	Ion Torrent Proton	Ion Torrent PGM	Substitutions in homopolymers	0.75 (0.35, 0.97)	1.00 (0.64, 1.00)
	Ion Torrent PGM	Ion Torrent Proton	Substitutions in homopolymers	1.00 (0.42, 1.00)	0.86 (0.57, 0.98)
FFPE Breast Tumors v. NA12878	NIST	Ion Torrent PGM	All combined variants	0.90 (0.87, 0.93)	0.92 (0.90, 0.94)

Samples	“Truth”	“Test”	Variant Type	Sensitivity (95% CI)	Specificity (95% CI)
FFPE Breast Tumors v. NA12878	NIST	Ion Torrent PGM	Indels	0.50 (0.32, 0.68)	0.95 (0.92, 0.97)
			Indels in homopolymers	0.50 (0.32, 0.68)	1.00 (0.97, 1.00)
FFPE Breast Tumors v. NA12878	NIST	Ion Torrent PGM	Substitutions	0.94 (0.91, 0.96)	0.90 (0.87, 0.93)
			Substitutions in homopolymers	0.91 (0.82, 0.97)	0.86 (0.81, 0.90)
FFPE Breast Tumors v. NA12878	NIST	Ion Torrent Proton	All combined variants	0.95 (0.92, 0.98)	0.92 (0.89, 0.94)
			Indels	0.50 (0.27, 0.73)	0.95 (0.91, 0.98)
			Indels in homopolymers	0.50 (0.27, 0.73)	1.00 (0.95, 1.00)
			Substitutions	1.00 (0.97, 1.00)	0.89 (0.84, 0.92)
			Substitutions in homopolymers	1.00 (0.87, 1.00)	0.86 (0.79, 0.91)

Table 6- Sensitivity and Specificity of Custom *DEARI* Ampliseq Panel. Sequencing data of CEPH control cell line NA12878 from the Ion Torrent Proton and PGM sequencing platforms were compared to the NIST’s highly confident integrated genotype calls in order to determine the sensitivity and specificity of the Ampliseq panel. The data from the PGM and Proton platforms were also compared to each other in order to determine differences in sensitivity and specificity of each platform. *Indicates the sensitivity and specificity remained the same independent of which Ion Torrent platform was used as “truth”. PGM data used the 1st generation of the 200bp Ion Torrent Sequencing Kit whereas Proton data used the 2nd generation version of the sequencing kit. To note, library construction protocols with the Proton sequencing platform were slightly changed upon noticing one primer pool acted inconsistently. The following changes were implemented: the reactions were completed in half reaction volumes and the amplification of the final library was removed from the protocol.

0.90(CI: 0.87-9.93) and specificity 0.92 (CI: 0.90-0.94)]. It is important to note that the variants identified in the sequencing of NA12878 on the Ion Torrent PGM and Proton platforms underwent stringent variant calling parameters and filtration methods. Many low quality variants as defined by low quality scores, those exhibiting strand bias, or those in low coverage areas were removed during these variant calling and filtration steps, thus possibly removing potential spurious variant calls and improving the sensitivity and specificity of the data. As these variant filtration methods were used to determine the presence of variants in our clinical samples, the achieved sensitivity and specificity of the sequencing panel in FFPE samples has made us highly confident in our ability to detect true variants harbored within our clinical samples.

Discussion

NGS technology has shown its potential to revolutionize clinical medicine and has helped the field to begin to understand the molecular mechanisms behind therapeutic resistance. In order for NGS to be truly effective, the technology must show a high degree of accuracy, sensitivity, and specificity, especially over previous gold standard methods. Many studies have shown NGS's ability to be as sensitive or more sensitive than Sanger sequencing or real time PCR, while also showing its ability to be more discovery based as well^{189, 196, 197, 213, 218}. Multiple platforms exist for completing NGS, each with their own benefits and pitfalls. One must determine which platform best fits the specific experimental conditions one is planning. We determined that the Ion Torrent platform was most applicable platform for our experiment, as the Ion Torrent sequencing platform is known for its quick turn-around time and low substitution error rates.

The Ion Torrent platform uses a highly multiplexed PCR reaction to target specific loci for sequencing. As reactions with hundreds of amplicons can have multiple off target and undesirable effects if the primers are not designed with great care, Life Technologies have created an Ampliseq designer to help in the process of designing the amplicons for a desired locus. The Ampliseq designer was used to design amplicons to amplify the *DEARI* locus, including the exonic and regulatory regions of the gene.

Moreover, in order to capture upstream regulatory regions, an 11 kb region upstream of the gene locus was included in the design. All non-repetitive areas were able to be captured in the design, as low genetic complexity was a major issue in amplicon primer design. Amplification by Ampliseq primer panels are affected by many of the traditional factors known to effect PCR efficiency, like GC content, primer-primer interactions, issues with low complex areas, and amplification bias. Due to some of these difficulties, one will see variances in the performance of amplicons across the loci. Despite these challenges inherent with PCR based systems, our Ampliseq panel was able to achieve sequencing depths at the *DEARI* locus that have never been previously described in breast cancer as well as cover areas known to harbor previously described variants by COSMIC and cBIO.

The inherent challenges associated with the PCR based Ampliseq reaction as mentioned previously can lead to variances in the quality of sequencing and its accuracy in detecting variant frequencies across the panel. As such, it's important to understand the limitations along with the sensitivity and specificity of one's system. This can be done in multiple ways. One popular method is population mixing, in which a population with a known variant is spiked-in at a particular percentage in relation to another population. However, one possible issue with this method is potential sub-clonal ploidy differences in the two sequenced populations that can change the variant frequencies, especially if the population harboring the variant is cancerous as tumor cells often exhibit ploidy differences. One way to subvert this is to use an artificial plasmid harboring an amplicon with the variant and to spike this plasmid into a cytogenetically normal DNA sample. This novel method was used to determine the accuracy of our sequencing to detect known variant frequencies. The data presented herein show that, for the most part, the amplicons tested were able to recapitulate the variant frequencies that were expected, giving confidence to the variant allele frequencies represented in our sequencing data. Some observed variances were present compared to the variant allele frequencies expected; however, this could be due to difficulties associated with PCR reactions, as discussed above, since Ampliseq sequencing is a PCR based method. However, it is important to note that variants represented in the spike-in assay required variant calling by manual extraction of the amplicons that contained the variant as many of the variants

within the spike-ins were not called by automatic variant callers or identified after variant call filtration. The exclusion of present variants by the automated variant callers warns against the sole reliance on these automated sequencing pipelines and exemplifies the importance of manually checking the sequencing data in viewers like Integrative Genome Viewer (IGV). For example, the variants present as part of the spike-in could be seen when viewing the data in IGV but were not represented in the resultant VCF files from the automated Torrent Variant Caller. Loss of the representation of particular variants may be due to issues with sequencing alignment to the reference sequence, low coverage of the particular locus, poor quality calculations, or the presence of the variant in genetically low complex areas like homopolymer regions. Therefore, a combination of automated variant callers with manual overview of the data in sequencing data viewers is important for ensuring complete and correct calling of the present variants. Overall, through the combined use of manual and automated variant calling, the novel spike-in assay created to determine accuracy regarding variant allele frequencies indicated that we can have confidence in our variant frequencies given by the NGS read out, even at the rare 1% variant frequency.

This confidence is further exemplified by the accuracy shown by the high concordance between our observed read counts and the predicted read counts within our model, given a specified variant frequency and overall locus coverage depth. Modeling of predicted read counts allows for a clear observation of the number of cell needed to carry a variant within a population in order to detect a specific variant allele frequency. This model can also help to demonstrate how certain variants may be filtered from the data due to lack of sufficient observations. Typical NGS practices require a minimum number of reads supporting the presence of a variant in order to reduce the chance that a particular variant could potentially be a sequencing error. Depending on the coverage achieved and stringency of variant filtering that is desired, rare variants may be filtered out due to not obtaining enough supporting reads to be determined a true variant rather than a sequencing error. However, as the *DEARI* Ampliseq panel achieved high coverage within our sequencing, we were able to attain large support for rare alleles. As determined by the comparison of the empirical read counts achieved in our sequencing to the predicted read count model, it was determined that the sequencing read counts had accuracies between

96-100%, giving us further confidence in our variant frequencies given by NGS read counts and increased assurances that the high coverage achieved for the variants were not representative of biases within the data. However, it is important to note that as higher coverage is obtained in order to increase the sensitivity to detect potential rare alleles within the samples, increased stringency of the variant filtering may be needed to reduce artifacts that occur during library construction or sequencing. This issue was specifically addressed through stringent variant quality filtering to reduce potential false positives associated with ultra-deep sequencing (see methods for detailed filtering procedures).

It is also important to note that there are limitations to our modeling of variant allele frequencies and expected read counts. Our model assumes very simple conditions, including diploid status, known zygosity of the variant, and no intra-sample heterogeneity. Clinical samples exhibit a high degree of complexity that is not recapitulated in our modeling. Tumor intra-heterogeneity has been shown to be highly common across cancers with very few variants being represented in the majority of the tumor population^{58, 217, 219}. Both variants and chromosomal alterations can exhibit intra-sample heterogeneity with mixes of both wildtype and variant allele populations, as well as the presence of diploid and haploid/2n+ populations. Chromosomal alterations are common in tumors, with amplifications, copy number gains, and deletions all potentially causing modulations in variant allele frequencies. For example, if a diploid tumor population existed harboring a heterozygous variant in 20% of the bulk tumor cells within a sample, NGS of the entire sampled tumor population would indicate that the variant allele frequency of the particular variant is 10%. Given the estimate that about 1,500 cells are sequenced in 10ng of DNA, an estimated 300 tumor cells at this stage harbor the variant. As the tumor evolves and progresses, a cell from the original 20% of tumor cells harboring the variant may undergo loss of heterozygosity of the wildtype allele. After clonal evolution, the tumor population may now exhibit the variant in 40% of the entire tumor mass sampled, with 30% still being diploid at the variant locus and 10% of the cells now demonstrating loss of heterozygosity. In this condition, though 40% of the tumor mass sequenced harbors the variant, the variant allele frequency as shown by NGS is actually 25%, with 600 of the 1,500 cells sequenced demonstrating the variant. This shows the importance of the

knowledge of zygosity status throughout the tumor population, specifically within the sub-clonal populations harboring variants of interests, and how intra-tumor heterogeneity of variants and copy number can play a large factor complicating the interpretation of NGS data. Furthermore, tumor purity can also affect variant allele frequency interpretation. Tumors, especially those of an invasive nature, are often marbled with multiple different types of normal cell populations, including stromal tissue and normal epithelium. The inclusion of these cell types within the sequenced sample can have profound effects on the ability to detect variants within the tumor cells, depending on the extent of normal cell contamination, simply due to reducing the number of cells of tumor origin being sequenced. For example, a 10ng tumor DNA sample sequenced which is noted for having 70% tumor purity is estimated to have around 1,050 tumor cells in the approximate 1,500 cells sequenced. Now, if a variant is present in only 30% of the tumor cells sequenced, it is predicted that only 315 cells of the 1,500 cells harbor the variant, or 21% of the cells sequenced. If the variant is a heterozygous variant carried in a diploid condition, and further the sequencing coverage received at this locus is approximately 1,000x, it can be expected to receive about 105 supporting sequencing reads for the variant. Variant filtration under the methods performed during our filtration would still have called the variant as we required at least 60 supporting reads for each variant. However, if tumor purity of the sample was only 35%, the variant would only be present in 157 cells within the sequenced population and, given about 1,000x sequencing coverage of the heterozygous variant, would only be present in 52 reads; thus, this variant would have been filtered out using our stringent variant filtration methods due to tumor purity alone. It is important to note that currently 1,000x coverage is typically only achievable via targeted sequencing methods. Therefore, using sequencing methods that are only capable of lower coverage presently, like whole genome sequencing, will further reduce that ability to detect these types of variants in heterogeneous populations. For example, given a sequencing coverage of 100x in the previous examples of 70% and 35% tumor purity, only about 10 and 5 reads supporting the variant would be present, respectively. Other factors that can affect final variant allele frequencies are biased-amplification or allelic dropout, which are known issues associated with library construction, as well as the natural structure of the genomic loci of interests like GC content and chromatin structure, which can as mentioned previously affect targeted

amplification and sequencing capabilities. All of these factors can have potentially large impacts on coverage uniformity and the accuracy of variant allele frequencies.

Due to these variances that can occur across the amplicons within the loci, it is imperative to quantify the degree to which one is capturing a true assessment of the data. This is often done by computation of the targeted sequencing panel's sensitivity and specificity to determine the capacity of the sequencing data to detect all variants, with limited false positives. Sensitivity and specificity was ascertained by comparing the sequence data of the CEPH normal control DNA NA12878 sequenced on our Ion Torrent platforms to the NIST's highly confident, integrated genotype calls for the same sample. The results showed a high degree of sensitivity and specificity for both Ion Torrent platforms, with the Ion Torrent Proton showing slightly greater sensitivity and the Ion Torrent PGM having slightly greater specificity. When specifically looking at the custom Ampliseq panel's abilities in FFPE clinical samples, the PGM platform performed similarly to the performance of intact cell line genomic DNA; however, the Proton platform showed specificity improvements within these samples. Both of the Ion Torrent platforms struggled with INDELS, a known weakness for the Ion Torrent Variant Caller^{212, 214}. It is due to this that all INDEL calls were disregarded in further contemplation of sequencing data. The degree of sensitivity and specificity able to be achieved on the platforms may have been slightly enhanced due to the use of pre-filtered genotype calls to compute these measures. Through stringent variant calling and filtration, low quality variants or variants within low coverage areas were removed before sensitivity and specificity was calculated. Inclusion of these poorer quality genotype calls which can represent sequencing or library construction errors would have led to a higher degree of false positive variant calls and reduced sensitivity performance of the *DEARI* ampliseq panel. The same variant calling and filtration methods used prior to sensitivity and specificity analysis performed herein were also used in the processing of the sequencing data of the clinical samples. Therefore, the sensitivity and specificity calculated reflects the performance of the complete processing of the final variant calls within the clinical samples, thus giving us assurance of the quality of final data produced. Overall, the Ion

Platforms showed great sensitivity and specificity in both the control NA12878 and FFPE data, giving us high confidence in the variants called by our Ampliseq panel.

As described herein, a highly sensitive and specific custom Ampliseq panel was generated to detect variants within the *DEAR1* locus, including upstream regulatory regions. *DEAR1* is a novel tumor suppressor, which plays important roles in acinar morphogenesis and TGF β driven EMT^{2, 131, 220}. Understanding the spectrum of genetic alterations encompassed in *DEAR1* within cancer can give further insight into the mechanism of *DEAR1*'s tumor suppressive actions as well as in understanding tumor progression in general. NGS technology is a major tool currently being used to describe the genomic landscape of cancer due to its highly sensitive and quantitative nature. However, it is important when using targeted sequencing panels with this technology, that the method for targeted capture allows for highly sensitive detection of variants. The custom *DEAR1* Ampliseq panel that was generated has been shown to be capable of the highly sensitive detection of rare variants as well as the ability to make highly accurate genotype and variant frequency calls.

Characterizing *DEARI's* (TRIM62) Genetic Role in Ductal Carcinoma *in Situ* Progression

Introduction

Ductal Carcinoma In Situ and Its Progression to Invasive Disease

Breast cancer is the number one diagnosed cancer in women²²¹. It is estimated by the American Cancer Society that about 1 in 8 women will be diagnosed with invasive breast cancer (IDC) within their lifetime²²¹. In 2014, 232,670 women were projected to be newly diagnosed with breast cancer, with 20-24% diagnosed with ductal carcinoma *in situ* (DCIS)^{221,222}. DCIS is defined as the presence of abnormal neoplastic cells filling the ductal lumen. As it is an *in situ* disease, the abnormal cells have not invaded through the basement membrane at this stage. DCIS accounted for 83% of *in situ* breast cancer, making it the most commonly diagnosed *in situ* breast lesion²²¹.

DCIS is the earliest form of breast cancer and is currently considered as a non-obligate precursor to its invasive form²²³. One of the single most important questions surrounding DCIS is why some lesions remain indolent and other lesions recur. Population-based studies have suggested that about 18-20% of DCIS lesions recur as a local recurrence in patients who received surgical resection^{224, 225}. Fifty percent of local recurrences are diagnosed as IDC²²⁶. The lack of biomarkers to stratify these patients with a higher risk of invasive recurrence have led to variance in how these lesions are treated in the clinic. Twenty-four percent of patients with DCIS elect to undergo unilateral mastectomy, a procedure that recent studies have indicated is linked to high rates of overtreatment as mastectomy has shown no significant difference in recurrence rates when compared to the less stringent treatment of lumpectomy plus radiation^{227, 228}. Other studies have also indicated that lumpectomy plus radiation is a superior treatment modality as it is associated with a 50% decrease in invasive recurrence and higher overall event free survival (81% versus 91%), when compared to lumpectomy alone^{225,229}. This has led to the American College of Radiation to release the Appropriateness Criteria on DCIS stating breast conservation therapy (breast conserving surgery (lumpectomy) with negative margins followed by whole

breast radiation) as an acceptable treatment and an alternative to mastectomy²³⁰. Due to this recommendation, there has been a large increase in the number of patients receiving breast conservation therapy, which now accounts for 43% of current chosen treatment modalities²²⁷. However, a portion of the women treated with the more stringent treatment modality will have harbored an intrinsically indolent lesion that would have never progressed, even without further radiation treatment²²⁵. This highlights the need for molecular and genetic factors that are able to stratify women for tailored treatment options according to their risk for invasive recurrence in order to reduce both over-treatment and under-treatment rates. The current recommended treatment option of lumpectomy plus radiation, though as mentioned previously is indicated to have a higher overall event free survival and reduced invasive recurrence rate, is also associated with a high degree of over-treatment as 14 DCIS patients would need to be treated with lumpectomy plus radiation in order to prevent one local recurrence²³¹. Further, under-treatment is also of clinical concern as more women are opting for lumpectomy alone²²⁷. For example, Van Leeuwen et al. 2011 has shown that patients undergoing surgery alone experienced a significant decrease in local recurrence free survival²³². Due to the uncertainty of the progression of DCIS lesions, surgical oncologists have difficulty in determining the best methods of treatment to reduce recurrence but limit the occurrence of over-treatment.

Many have attempted to determine the molecular mechanisms of DCIS progression in an effort to establish clinical biomarkers. The majority of studies have tried to understand the progression through determining distinctions between DCIS and IDC. Surprisingly, many of these comparison studies of DCIS and IDC have indicated that the lesions are very similar, molecularly. Hwang et al. 2004 showed chromosomal alterations tend to be synchronous between Pure DCIS and IDC, though the Pure DCIS lesions did show significant enrichment of particular chromosomal alteration differences associated with tumor grade²³³. Moreover, it has been shown that nuclear grade and the molecular subtypes associated with IDC (ER, PR, HER2, etc.) as well as cytokeratin 5/6 markers often corresponded between adjacent in situ and invasive lesions^{223, 234}. Similarities between the two lesions have also been seen at the epigenetic level with multiple studies indicating comparable rates of promoter hypermethylation during

the progression of DCIS to IDC²²³. The high degree of synchrony between the early, non-invasive lesion and the invasive counterpart have led to the proposed “branched” model of progression of DCIS to IDC²³⁵. The “branched” model describes a linear progression of DCIS to IDC, with different branches off the linear model describing the restriction of DCIS lesions progressing to IDC based on nuclear grade correspondence between the lesions, i.e. a low grade DCIS can only progress to a low grade IDC and visa-versa. This model indicates that differences exhibited during DCIS progression are not necessarily between the in situ and invasive tumors but are potentially associated more with differential grades of DCIS. Contrary to this model, another model, often called the “parallel” model has been described which proposes a common cell of origin but with subsequent independent yet synchronous progression of DCIS and IDC lesions²³⁵. Support for this model, also known as the “Sontag-Axelrod” model, is indicated by continued existence of progressed DCIS and IDC adjacent to one another as well as the possible limited chromosomal alterations that are confined to either the DCIS or IDC components reported by Johnson et al. 2012^{233, 235-237}. Due to the convoluted nature of the relationship between DCIS and IDC, much more work is still needed in order to better understand the molecular mechanism of DCIS progression and to help in the development of biomarkers for the stratification of treatment for this disease.

DEAR1 is an Important Regulator of Polarity and EMT

To better understand the progression of breast cancer, our lab has studied the novel tumor suppressor gene, *Ductal Epithelium Associated Ring chromosome 1 (DEAR1)*, also annotated as TRIM62, which is an E3 ubiquitin ligase often downregulated in breast cancer^{2, 131}. DEAR1 is part of the TRIM family of proteins, which are known to play important roles in immunity, differentiation, cell death, and proliferation. The important roles TRIM family members play in normal homeostasis, upon deregulation, can be integral to cancer progression^{72, 238}. DEAR1 has been shown to be a major regulator of polarity and acinar morphogenesis¹³¹. In breast cancer, DEAR1 is downregulated during the transition from normal epithelium to DCIS and the invasive lesion¹³¹. Knockdown of DEAR1 in human mammary epithelial cells (HMECs) grown in 3D culture was shown to be associated with luminal filling, a hallmark of DCIS^{131, 136}. Moreover, we have shown that DEAR1 is a master regulator of TGFβ mediated

Epithelial to Mesenchymal Transition (EMT) through its ubiquitination of SMAD3 in mammary cells^{2, 220}. *DEAR1*, located at chromosome 1p35.1, resides in an area often hemizygotously deleted in epithelial cancer and is mutated in a wide spectrum of tumors including breast, colon, and stomach cancer^{2, 131}. We have shown that a subset of these mutations can be functional, exemplified by the ability of wild-type *DEAR1* to correct, by genetic complementation, the aberrant acinar morphogenesis within the *DEAR1* mutated metastatic breast cancer cell line, 21MT¹³¹. *DEAR1* has also been found to undergo nonsense (N) and frameshift (FS) mutations in tissues including bladder (N), melanoma (N), breast (FS) and lung (FS) cancer^{145, 146}.

Since *DEAR1* has shown to be pivotal for the maintenance of correct acinar morphogenesis and prevention of TGF β induced EMT, and with the knowledge that *DEAR1* undergoes functional mutations that can promote aberrant acini formation in breast cancer, it was determined if *DEAR1* is mutated in the earliest form of breast cancer, DCIS, and if these mutations can inform us about DCIS progression to IDC. A custom next generation sequencing panel was developed to complete targeted ultra-deep sequencing of *DEAR1* that exhibited strong sensitivity and specificity. Using this *DEAR1* sequencing panel, Pure DCIS and DCIS adjacent to IDC were sequenced and found to contain *DEAR1* somatic variants in 71% of these lesions, including many predicted to be deleterious by PolyPhen2 and SIFT prediction tools as well as functional evidence for a subset of these variants. Excitingly, evidence for a potential germline variant in an early onset case of Pure DCIS was found. Lastly, data from the sequencing of DCIS lesions associated with IDC seem to support a parallel model of evolution for DCIS and IDC. Our work has described the ultra-deep targeted sequencing of an important tumor suppressor, *DEAR1*, which indicated the importance of variants within this gene in the earliest form of breast cancer, DCIS.

Methods

Human Specimen Collection

All human tissues were identified by Dr. Aysegul Sahin (MDACC) and obtained without

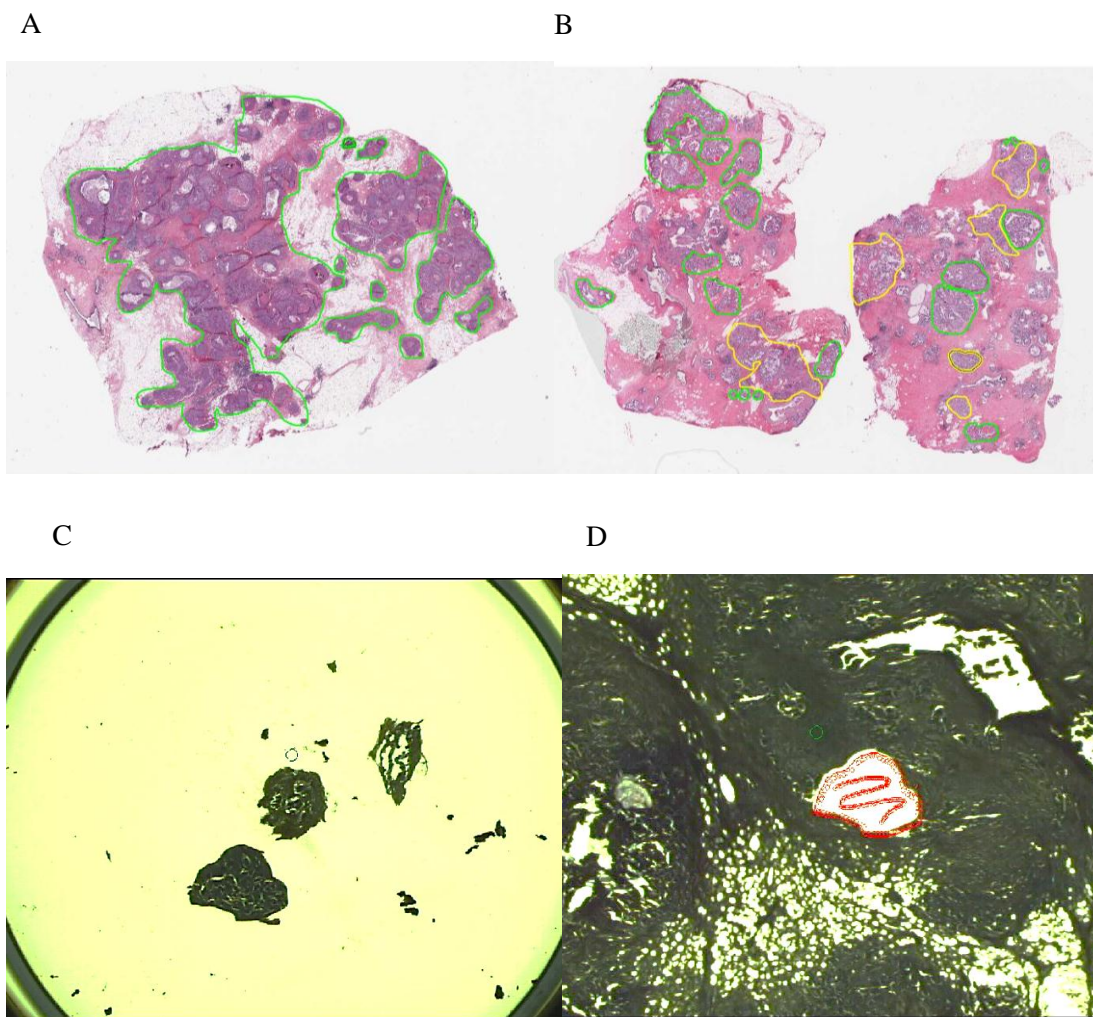


Figure 7- Histological Representation of DCIS FFPE Samples Used in Sequencing. Serial H&E sections were taken prior to sample sections, followed by digital annotation for both Pure DCIS (A) and DCIS with invasive components (B). The in situ lesions were marked digitally in a green outline where the invasive lesions were marked in a yellow outline. Laser capture microdissection (LCM) was completed on DCIS with invasive component samples. Representation of the capture of the tissues is shown in C, where the remaining tissue after LCM is shown in D.

identifiers from the MD Anderson Cancer Center tumor bank according to the approved IRB protocol. Samples were chosen based on identification of sufficient amounts of DCIS and/or invasive components to acquire at least 10 nanograms of DNA. Cases were also selected on presence of recurrence follow up data as well as age of disease onset as equal numbers of early and typical age of onset was requested. For a list of samples sequenced for each patient see Appendix VII.

DNA Extraction and Quantification

One to three 10um of DCIS and matched normal FFPE tissue were sectioned and mounted on either positively charged glass slides (Pure DCIS: VWR 48312-013) or PEN membrane slides (DCIS adjacent to IDC: Life Technologies LCM0521). Before tissues were sectioned onto the PEN membrane slides, the slides was exposed for 30 minutes to UV light in order to help with tissue adherence. A serial tissue section was taken prior to each sample collection for hematoxylin and eosin (H&E) staining. H&E slides were digitally scanned on the Aperio AT Turbo whole slide scanner and tumor area were digitally annotated on Aperio eSlideManager/Aperio ScanScope by pathologist, Dr. Fei Yang (MDACC). Tissue for DNA extraction was collected by either scraping whole slide sections (Pure DCIS/Normal) or by Laser Capture Microdissection (LCM) (DCIS adjacent to INV) on the Arcturus XT LCM system. For Pure DCIS and normal lymph node tissue undergoing extraction by scraping, samples were deparaffinized for two 5 minute 100% xylene washes followed by collection of tissue by scraping with a fresh scalpel. For DCIS with invasive component (DCIS/INV) cases collected by LCM, samples on PEN membrane slides underwent a 2 hour 65° Celsius (C) incubation to help with tissue adherence followed by a pre-staining procedure, completed as described by the Arcturus Paradise Plus staining protocol with a few modifications (Life Technologies KIT0312J; manual 1287200 <https://tools.lifetechnologies.com/content/sfs/manuals/1287200.pdf>). Briefly, samples were hydrated by exposure to 100% xylene followed by decreasing concentration of ethanol, from 100% to 75%, and water for the duration and manner as stated in manual. This was then followed by staining with the Paradise Plus stain (Life Technologies KIT0312J) for 7 seconds and then dehydration of the samples by incubation with increasing concentrations of ethanol, from 75% to 95% for two 30 second exposure,

followed by two 1 minute 100% ethanol incubation and 10 minute 100% xylene exposure. All tissues were then subject to a 5-10 drying time after final xylene incubation. Discrete lesion components were isolated by LCM and processed independently after staining. LCM collection was managed by pathologist, Dr. Fei Yang.

After tissue collection, DNA was extracted by PicoPure DNA Extraction kit (LifeTechnologies KIT0103), using the protocol for fixed tissue sections. Briefly, 150 μ l or 30 μ l of Proteinase K suspended in supplied reconstitution buffer was added to either scrapped tissue sections or LCM caps, respectively. The samples were then incubated at 65°C for 22-24 hours followed by an exposure to 95°C for 10 minutes to inactivate the Proteinase K. For the LCM samples, after proteinase K inactivation, extracted DNA was pooled from the LCM caps originating from a single lesion component of a single sample.

The extracted DNA samples were then purified by either Agencourt AmpPure XP (Beckman Coulter A63880) for the Pure DCIS samples or by a specialized ethanol purification technique for the DCIS/INV samples isolated by LCM. The Pure DCIS samples collected by scrapping were purified by AmpPureXP, according to the protocol. Briefly, the samples were incubated with 1.8x AmpPureXP beads for 5 minutes followed by magnetic separation. The beads were then washed twice with 600 μ l of freshly prepared 75% ethanol and then dried before suspending in 20 μ l of water. For the DCIS/INV samples isolated by LCM and scrapped normal samples, the DNA was purified by a modified ethanol purification protocol in which the samples were combined with 2 μ l of Pellet Paint (EMD Millipore 70748), 10% 3M Na Acetate pH 5.2, 1 μ l GenElute LPA (Sigma 56575-1ML), and 2.5x of 100% EtOH and then incubated for 20 minutes at room temp. Afterwards, samples were centrifuged at 14,000rpm for 5min and the pellet was then washed with freshly made 70% and then 100% ethanol with centrifugation followed by supernatant removal after each washing. The final pellet, collected after the two ethanol washes, was dried at 90°C for 5min. Pellet was then eluted with 10 μ l of water. Lastly, DNA was quantified by qPCR using TaqMan RNase P Detection kit (4316831 Life Technologies) using specified protocol (“Measuring Template Efficiency” <https://tools.lifetechnologies.com/content/sfs/brochures/cms042785.pdf>). Briefly, the standard human genome DNA provided in kit was serially diluted from

10ng/μl to 0.5ng/μl to form a standard curve. For the 96 well PCR plate format, to each well was added: 3 μl of FFPE DNA or diluted standards, 12.5 μl 2 × Universal Master Mix (No AmpErase UNG) (Life Technologies 4364343), 1.25 μl 20x RNase P Probe/Primer Mix (from kit 4316831 Life Technologies), and 8.25 μl RNase-free water. For the 384 well format, added to each well was 2.5 μl of each diluted standard (from kit 4316831 Life Technologies) or 0.2x diluted FFPE DNA (1.5ul of DNA into 6ul of dH₂O), 5 μl 2 × Universal Master Mix (No AmpErase UNG) (Life Technologies 4364343), 0.5 μl 20x RNase P Probe/Primer Mix (from kit 4316831 Life Technologies), and 2 μl RNase-free water. qPCR cycling was performed as follows: 50°C 2 min, 95°C 10 min, and 40 cycles of 95°C 15 sec and 60°C 1 min. Using standard curve amplification, quantities were determined for FFPE samples. Correlation of logarithmic best fit line of CT values of standards (Y axis) vs Concentration (X axis) was used to determine accuracy of quantities, with R² typically being over 0.9900. For samples quantified by 384 well format, which underwent sample dilution, quantities based on RNase P curve were then multiplied by dilution factor to receive final sample quantity.

Ampliseq Library Construction and Sequencing

In order to amplify *DEARI* for sequencing, Ampliseq Designer version 1.2.8 (Life Technologies <https://www.ampliseq.com>) was used to design 150bp amplicons, spanning across a 48kb region on chromosome 1: 33,610,600-33,658,985 (hg19). The 184 amplicons, split into two pools, were used to complete library construction according to the Ion Torrent library construction protocol (Life Technologies MAN0006775 Revision 4.0 <http://ioncommunity.lifetechnologies.com/docs/DOC-3254>). Briefly, the Ampliseq panel was used to amplify *DEARI* in a highly multiplexed manner using 10ng of DNA per primer pool with the Ion Ampliseq Library Kit 2.0 (Life Technologies 4475345). Specifically the FFPE DNA was added with 10ul of the 400nM 2x Ion Custom Ampliseq Primer Pool, and 4ul of 5x of the Ion Ampliseq HiFi Master Mix in each reaction. The following thermocycling protocol was used: 1 cycle at 99°C for two minutes followed by 19 cycles of 99°C for 15 seconds and 60°C for four minutes. The primer sequences were then partially digested by the addition of 2μl of FuPa reagent to each samples that were then exposed to the following conditions: 50°C for 10 minutes, 55°C for 10

minutes, and then 60°C for 20 minutes. After the primer sequences were digested, the Ion Express Barcode Adapters were added (Life Technologies 4471250), in addition to Switch solution and DNA ligase, to the amplified library and subjected to the following temperature conditions: 22°C for 30 minutes followed by 72°C for 10 minutes. The completed library was then purified with the Agencourt AMPureXP magnetic beads done according to Beckman Coulter protocol, with scaled volumes to fit the input volume (Beckman Coulter protocol # B37419AA).

The purified libraries ran on the Ion Torrent PGM were amplified using a Platinum PCR SuperMix High Fidelity polymerase and a library amplification primer mix, using the following conditions: 98°C for 2 minutes followed by 5 cycles of two steps including 98°C for 15 seconds and 60°C for 1 minute. The amplified library was then re-purified with Agencourt AMPureXP magnetic beads using the same conditions during the previous library purification as before (see above). For the samples ran on the Ion Torrent Proton, libraries were not amplified after purification to prevent unequal amplification of the two primer pools. Once final purified libraries were created, the libraries associated with the two different Ampliseq primer pools were added together by sample accordingly. Templates were then created with the 9x diluted pooled libraries by following the Ion PI and PGM Template OT2 200 Kit V2 protocol (Life Technologies 4480974 and 4485146; MAN0007624 <https://tools.lifetechnologies.com/content/sfs/manuals/MAN0007624.pdf>). Briefly, the following reagents were added, at the specified amounts, to 65µl of the diluted library: Ion PI Reagent mix TL (750 µl), Ion PI PCR Reagent B (450µl), Ion PI Enzyme Mix TL (75µl), Ion PI PCR Reagent X (60µl), and the Ion PI Ion Sphere Particles (100µl). Samples were then ran on the Ion OneTouch 2 instrument. Template positive Ion Sphere Particles (ISPs) were recovered from the One Touch 2 by centrifugation and then ~1.4mL of supernatant was removed. Beads were then washed twice with ~ 1.1mL of water and the pellet resuspended in Ion PI ISP Resuspension Solution. Quality was assessed by Qubit 2.0 flourometer. The template-positive ISPs were then enriched on the Ion OneTouch ES by the attachment of ISPs to washed streptavidin C1 beads. The enrichment was completed by following the protocol as described in the Ion PI Template OT2 200 kit v2 manual: after the attachment of the ISPs to beads,

subsequent washes as well as a melt-off reaction occurred on the OneTouch ES. Melt-off solution contained 40µl of 1M NaOH, 3 µl of 10% Tween 20 and 277µl of water.

Enriched ISPs were then prepared for sequencing using the Ion PI or PGM Sequencing 200 Kit V2 (Life Technologies 4485149 and 4482006) as described by the Ion PI and PGM Sequencing 200 kit v2 protocol (MAN0007961 and MAN0007273 <http://ioncommunity.lifetechnologies.com/docs/DOC-7459> and <http://ioncommunity.lifetechnologies.com/docs/DOC-6775>). Briefly, the enriched ISPs were washed with water, followed by the addition of Ion PI Annealing buffer (15µl) and Ion PI Sequencing Primer (20µl) to the enriched ISPs and then amplified on a thermocycler with the following steps: 95°C for 2 minutes and 37°C for 2 minutes. After amplifying the ISPs, 10µl of the Ion PI Loading buffer was added and the final mixture was loaded onto a washed and calibrated Ion PGM 316 or Proton chip. The chips were then washed with a foam created from 45µl of 50% Annealing buffer and 5µl of 2% TritonX-100, followed by the loading of 55µl of 50% Annealing buffer. This was followed by the chips being centrifuged and flushed with Flushing solution and 100µl of 50% Annealing buffer. Ion PI Sequencing Polymerase was then flushed across the chip in a solution with 50% Annealing buffer. Upon completion, samples were then sequenced on the Proton and PGM 316 chips using the Ion PI and PGM 316 Chip Kit (Life Technologies 4482321 and 4483324). Pure DCIS samples were sequenced each on one PGM 316 chip and DCIS microdissected samples as well as normal samples were barcoded and sequenced 10 samples per proton chip. For samples barcoded on the Proton chips, Ion Torrent Suite was used to separate barcoded sequences and exported as individual BAM files.

Sequence Analysis

Sequencing bam files were analyzed by Torrent Variant Caller v.4.2 via customized *DEAR1* parameters, followed by its output as a Variant Calling Format (VCF) file²⁰¹. The customized DEAR1 parameters consisted of recommended Ion Torrent variant calling parameters with the following changes: Splice Site Size-20bp; Maximum Coverage-100,000; Data Quality Stringency- 8.5; Downsample to Coverage- 100,000; Snp Minimum Coverage Each Strand-7; Snp Minimum Allele Frequency-0.01; Snp

Minimum Coverage-20; SNP strand bias-0.90; Mnp Minimum Coverage Each Strand-7; Mnp Minimum Variant Score-400; Mnp Min Allele Frequency- 0.01, Mnp Minimum Coverage-20; Mnp Strand bias-0.90, Kmer length-19, Minimum Frequency of Variant Reported-0.15; Indel Minimum Variant Score-10; Indel Minimum Coverage Each Strand-3; Indel Minimum Allele Frequency- 0.05; Indel Minimum Coverage-20; and Indel Strand Bias-0.85. Sample VCF files were annotated by SnpEff/SnpSift using the tool's inherent annotation as well as with Phase 1 Version 3 data from the 1000 Genomes project²⁰²⁻²⁰⁴. SnpSift was also used to complete variant filtering in accordance with the following parameter requirements: quality score was required to have a greater than 30 Phred score, as well as at least 1,000x overall locus coverage, 60x or greater coverage of the variant allele, and a minimum of 30x coverage on both forward and reverse strands. Stringency filters allowed for 13-66% reduction in incidental variants as well as the retention of noteworthy variants (Appendix III). Variants identified in both the tumor and normal samples were filtered out. However, a small number of potentially germline variants (allele frequencies around 50% or 100%) were found to exist in samples to which the matched normal sample was unable to be sequenced. These unconfirmed potential germline variants were not filtered out of our analysis as there was no definitive proof these were in fact germline; therefore, our total variant count is slightly inflated by these variants. Genome Analysis ToolKit was used to determine overall coverage²⁰⁵. The R program for statistical computing packages' ggplot2 were used to create figures detailing sequencing data, including coverage plots (Appendix IV)^{202, 203, 206, 207}. Mutation mapper was used to form figures detailing position of variants in DEAR1^{145, 146}.

Digital PCR

Digital PCR was completed on either the QX200 Droplet Digital PCR System (Bio-Rad 186-4001) or QuantStudio 3D (Life Technologies 4489084) platform. Custom genotyping TaqMan (4331349) was ordered from Life Technologies using a dual reporter system with VIC fluorescence tagging the reference sequence and FAM tagging the mutant sequence. PCR reaction mix was created using dPCR master mix (Life Technologies 4482710), target probes, and diluted DNA (10-36ng). TaqMan genotyping probes were first optimized for largest cycle separation between the VIC and FAM

probes with at least 1.5x cycle difference using qPCR with the following conditions: 95°C for 10min with 40-50 cycles of 92°C for 15sec and 60-64°C for 1 min. For Bio-Rad, samples underwent droplet generation (Bio-Rad 186-4002) whereas the samples undergoing digital PCR on the Life Technologies platform were spread across a microchip featuring 20,000 wells using an autoloader (Life Technologies 4485507 and 4482592). PCR was performed according to protocol with specified cycle number and annealing temperatures as determined in preliminary qPCR optimization. Data was then analyzed by QX200 droplet reader (Bio-Rad 186-4003) or by QuantStudio 3D (Life Technologies 4489084). Each variant was tested in 2-3 replicates. The R program for statistical computing packages' ggplot2 were used to create figures detailing Digital PCR data (Appendix V)^{202, 203, 206, 207}.

Variant Functional Studies

Luciferase reporter assays were completed in HEK293T by Dr. Nanyue Chen, as described previously². Mutation primers and constructs of *DEARI* in pcDNA vectors were created using Stratagene Quickchange Lightning Site Directed Mutagenesis according to manufacturer's protocol (Agilent 210518 Primer Design Program: <http://www.genomics.agilent.com/primerDesignProgram.jsp> ; Mutagenesis Protocol: <http://www.chem.agilent.com/library/usermanuals/Public/210518.pdf>). Wildtype and mutant *DEARI* with TGFβ Signal Transduction reporter luciferase plasmids were transfected into 293T cells with FuGENE HD according to manufacturer's protocol (Promega E2311; TM328 <https://www.promega.com/~media/files/resources/protcards/fugene%20hd%20transfection%20reagent%20quick%20protocol.pdf>). At 24 h after transfection, the cells were lysed in lysis buffer (Promega E1910; <https://www.promega.com/~media/files/resources/protcards/dual-luciferase%20reporter%20assay%20and%20dual-luciferase%20reporter%201000%20assay%20systems%20quick%20protocol.pdf>) according to manufacturer's protocol and luciferase activity was measured using a Monolight 2010 luminometer (Turner BioSystems). For the 3D culture assays, SKBR3 breast cancer cell lines were transfected with the same empty pcDNA or pcDNA plasmid containing wildtype *DEARI* or mutant *DEARI* (R187W or R254Q) as used for the luciferase assays. After transfection and expression of the plasmids were

established, the cells were selected by Puromycin exposure for stable expression of the plasmids. Pooled stable clones were verified by western and cDNA sequencing. Pooled clones were then grown on Matrigel Growth Factor Reduced (GFR) Basement Membrane Matrix (Corning 354230) according to Debnath et al. 2003 method²³⁹. Cultures were grown with 5% FBS in RPMI for 10 days. Acini cultures were then fixed and immunostained according to Debnath et al. 2003 protocol using an antibody against alpha 6 integrin (EMD Millipore CBL458), Alexa Fluor 488 Donkey Anti-Rat Secondary Antibody (Life Technologies A-21208) and DAPI stain (Life Technologies D1306)²³⁹. Immunostained cultures were then imaged using Nikon 80i Microscope System.

Results

Ultra-Deep Targeted Sequencing Reveals High Frequency of Alteration of DEAR1 in DCIS

Maintenance of polarity and constriction of migration are important in suppressing invasive capabilities^{153, 240}. Our previous data has indicated an important role for DEAR1 in maintaining polarity and restricting TGF β -mediated EMT^{2, 131, 220}. Moreover, genetic alterations of *DEAR1* has been shown to be able to alter DEAR1's ability to regulate these tumor suppressive activities^{2, 131, 220}. Therefore, *DEAR1* variants may play an important role in the progression from DCIS to IDC by their deregulation of DEAR1's regulation of polarity and migration. To better understand *DEAR1*'s role in DCIS progression, ultra-deep targeted sequencing of the *DEAR1* locus was completed in 17 Pure DCIS and 17 DCIS/INV using the ampliseq panel previously described in chapter 3. Patient characteristics used in this study are described in table 7. For the DCIS/INV, the individual lesion components were microdissected and sequenced separately in order to understand the concordance rate of genetic variants between the lesions, which can potentially provide insight to DCIS's progression to IDC. Sequencing of Pure DCIS and DCIS/INV indicated that 71% of both Pure DCIS as well as DCIS/IDC lesions exhibited variants within *DEAR1* following stringent filtering (please see methods for details and explanation of variant filtration). Sequencing of a 48kb locus encompassing *DEAR1* identified of a median of 8.25, 15.3, and 35.9 variants per FFPE sample in the Pure DCIS, DCIS component of DCIS/INV, and the invasive component of

Clinical Characteristics	Lesion Type	Characteristic Information
Median Age	Pure DCIS	53 years old (range 34-85)
	DCIS with IDC	55 years old (range 37-72)
ER+	Pure DCIS	71% (12/17)
	DCIS with IDC	94% (16/17)
PR+	Pure DCIS	59% (10/17)
	DCIS with IDC	94% (16/17)
ER-/PR-	Pure DCIS	29% (5/17)
	DCIS with IDC	0% (0/17)
Grade Status: Grade 1	Pure DCIS	0% (0/17)
	DCIS with IDC	11% (2/17)
Grade 2	Pure DCIS	35% (6/17)
	DCIS with IDC	65% (11/17)
Grade 3	Pure DCIS	65% (11/17)
	DCIS with IDC	24% (4/17)

Table 7- Patient Characteristics of Sampled Pure DCIS and DCIS with IDC Components. Table contains the information of age, hormone status, and nuclear grade for those samples sequenced with the targeted *DEARI* sequencing panel.

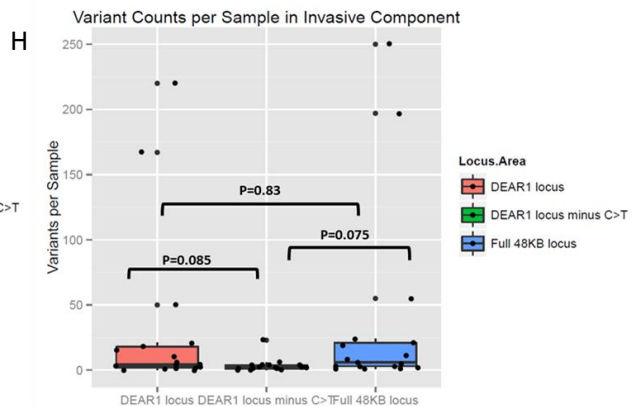
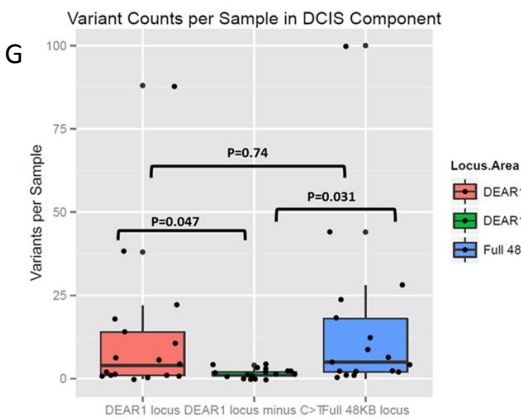
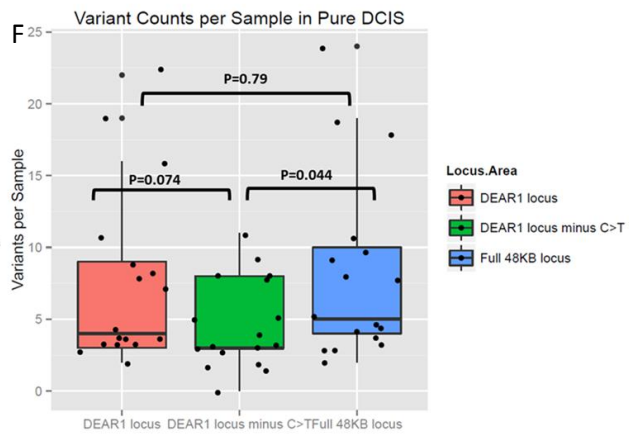
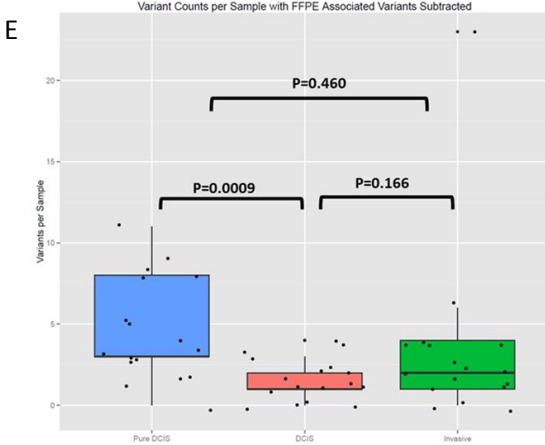
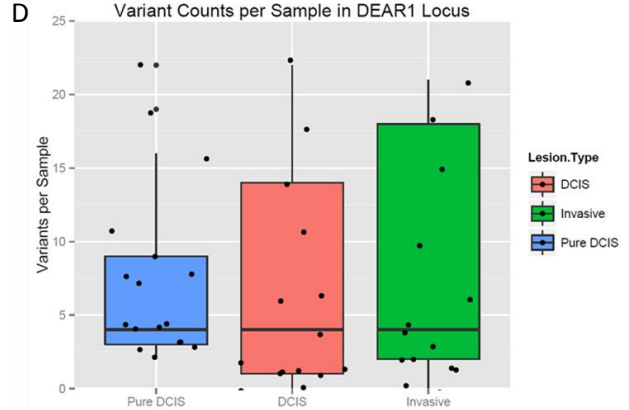
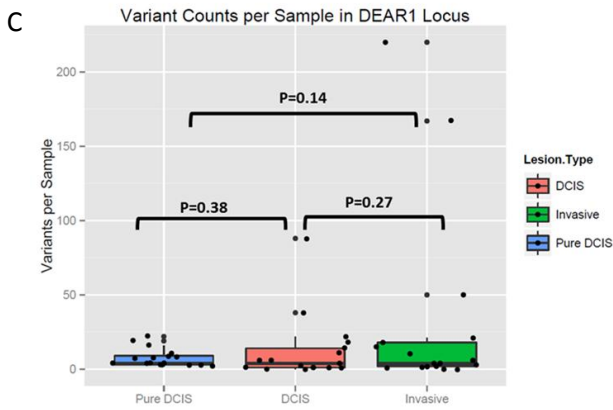
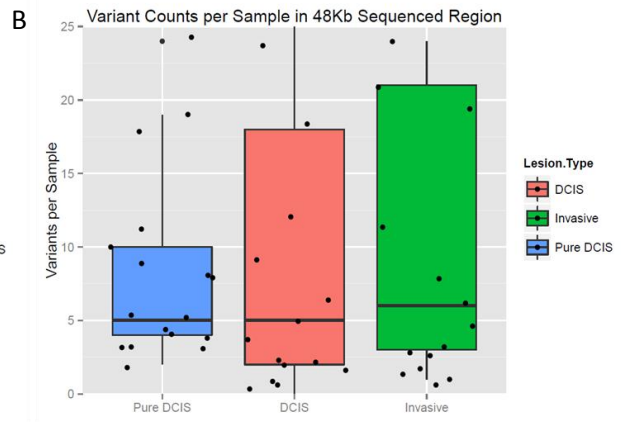
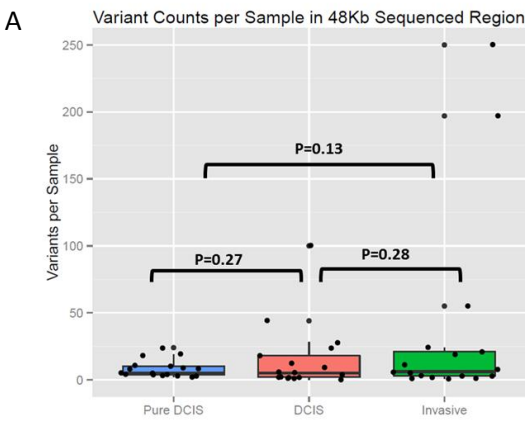


Figure 7- Variant Counts Per Sample in *DEARI* in DCIS. A-E) Figures indicates the range and median of variants within the (A) 48kb locus sequenced, (B) 48kb locus sequenced zoom, (C) 37kb locus of *DEARI*, (D) 37kb locus of *DEARI* zoom , and (E) when C>T and G>A variants, often associated with formalin fixation in FFPE samples, are removed. F-H) Figures indicates the range and median of variants within the variants within Pure DCIS (F), DCIS component of DCIS/INV samples (G), Invasive component of DCIS/INV (H) when comparing the 48kb locus sequenced, the 37kb locus of *DEARI*, and when C>T and G>A variants are removed. Statistical significance was established by a one-way Analysis of Variance (ANOVA).

DCIS/INV, respectively (no statistical difference between groups) (Fig. 7a and b). When focusing solely on the 37kb *DEAR1* gene through the exclusion of the 11kb intergenic region 5' to *DEAR1*, the median variant count per sample was 7.6, 12.6, and 30.8 in the Pure DCIS, DCIS component of DCIS/INV, and the invasive component of DCIS/INV, respectively (no statistical difference between groups) (Fig. 7c and d). However, a large degree of the variants in the FFPE samples were transitions consisting of C<T or G>A alterations (52 variants in Pure DCIS, 187 and 465 variants in the DCIS and invasive lesions, respectively, in DCIS/INV samples). C>T and G>A transitions are known to be enriched in samples undergoing formalin fixation as this process can cause cytosine deamination, inducing the transition²¹⁸. When C<T and G>A alterations were removed from the variant count, the median number of variants per sample was quite reduced. Specifically, the median variant count per sample in the 37kb *DEAR1* gene was 4.6, 1.2, and 3.5 in the Pure DCIS, DCIS component of DCIS/INV, and the invasive component of DCIS/INV, respectively (Fig. 7e). Statistically significant or near significant reductions in variant counts per sample after the removal of C>T and G>A were seen for all lesion types (Fig.7f-h). Interestingly, after removal of the C>T and G>A variants, there was a statistically significant difference in the variant count per sample between the pure DCIS and the DCIS component of the DCIS/INV samples (p=0.0009) (Fig.7e). Analysis by one-way ANOVA showed no statistical difference in variant counts per sample between the full 48kb locus sequenced and the 37kb region encompassing *DEAR1* for the pure DCIS as well as the DCIS and invasive components of samples with concurrent adjacent lesions (Fig.7f-h). *DEAR1* exhibited variants in all protein domains as well as the regulatory UTR regions (Fig. 8, Table 8 & 9; Appendix X). Some of these variants did occur at a moderate to high variant frequency (Table 10; see further discussion on pg 115). However, most of the variants across sample types were of intronic origin (80-88.5%) as expected. Digital PCR was used to validate a selection of exonic variants. Validation of 9 variants showed a 78% validation rate (7/9) and a 0.962 correlation rate between the variant allele frequencies observed from sequencing and those shown by digital PCR (Fig. 9).

Sequencing of Pure DCIS Indicated the Presence of DEAR1 Exonic and Regulatory Variants

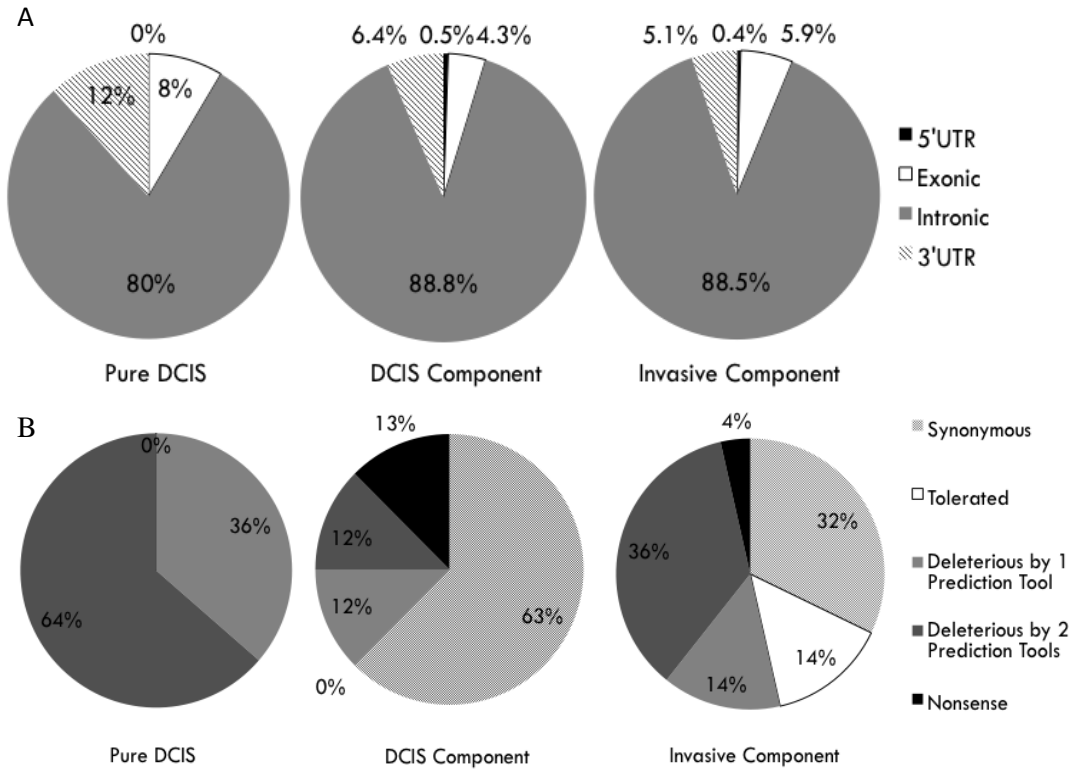


Figure 8- Distribution of Variants in *DEARI*. A) Figure shows the frequency of variants within each region of *DEARI*. B) Representation to functional classification of the variants in *DEARI*.

<i>Sample</i>	<i>Mutation</i>	<i>% Allele frequency</i>	<i>Validated</i>	<i>Area Within DEAR1</i>
PD03	D421G	4%		SPRY domain
PD04	D240N	7%		Coiled coil domain
PD05	D106V	4%	Yes	B-box domain
	D421G	7%	Yes	SPRY domain
PD06	D421G	6%		SPRY domain
PD07	D106V	4%	Yes	B-box domain
PD08	D106V	4%	Yes	B-box domain
PD12	R187W	13%	Yes	Coiled coil domain
	D240N	81%	Yes	Coiled coil domain

Table 8- Full List of *DEAR1* Variants in Pure DCIS FFPE Samples Within Exonic and Regulatory Regions. Table contains the list of variants in each Pure DCIS sample occurring in the exonic or regulatory regions of the gene, indicating the allele frequency, validation status, and region of *DEAR1* variant is contained in.

<i>Sample</i>	<i>Mutation</i>	<i>Lesion Type</i>		<i>% Allele frequency</i>	<i>Validated</i>	<i>Area Within DEAR1</i>
		<i>DCIS of DCIS/ INV</i>	<i>Invasive of DCIS/ INV</i>			
D01	G349G	X		4%		SPRY domain
	T>C Chr1:33648206		X	16%		Upstream; GATA3/KAP1 Binding site
D04	R384C		X	3%		SPRY domain
D05	T>C Chr1:33612419	X	X	3%DCIS 15%INV		Upstream; GATA3/KAP1 Binding site
	E183K		X	4%		Coiled coil domain
	L185L		X	3%		Coiled coil domain
	A202A		X	4%		Coiled coil domain
	Q233K		X	3%		Coiled coil domain
	L255F		X	4%		Exonic
	V356M		X	6%		SPRY domain
	T361S		X	9%		SPRY domain
	R374C		X	5%		SPRY domain
	G376D		X	4%		SPRY domain
	F386F		X	7%		SPRY domain
	W404*		X	3%		SPRY domain
	R410Q		X	3%		SPRY domain
	c.*676AGGG> TCCC		X	43%		3'UTR (mir10b)

<i>Sample</i>	<i>Mutation</i>	<i>Lesion Type</i>		<i>% Allele frequency</i>	<i>Validated</i>	<i>Area Within DEAR1</i>
		<i>DCIS of DCIS/ INV</i>	<i>Invasive of DCIS/ INV</i>			
D06	R184R	X		3%		Coiled coil domain
	E370K	X		4%		SPRY domain
	S23S		X	4%		RING domain
	G179D		X	4%		Coiled coil domain
	V364M		X	4%		SPRY domain
	G376S		X	3%		SPRY domain
	I378M		X	3%		SPRY domain
D08	G>A Chr1:33648250	X		15%		Upstream; GATA3/KAP1 Binding site
D10	c.*676AGGG> TCCC		X	22%		3'UTR (mir10b)
D11	E77E	X		3%		Exonic
	G257E	X		3%		Exonic
	N395N	X		3%		SPRY domain
	I128I		X	3%		Coiled coil domain
	D130D		X	16%		Coiled coil domain
	V336M		X	3%		PRY domain
	R374H		X	5%		SPRY domain
	G376D		X	4%		SPRY domain
	S383N		X	3%		SPRY domain

<i>Sample</i>	<i>Mutation</i>	<i>Lesion Type</i>		<i>% Allele frequency</i>	<i>Validated</i>	<i>Area Within DEAR1</i>
		<i>DCIS of DCIS/ INV</i>	<i>Invasive of DCIS/ INV</i>			
D11	R410W		X	3%		SPRY domain
D12	G>A Chr1:33648250	X	X	80%DCIS 6%INV		Upstream; GATA3/KAP1 Binding site
	Q19*	X		4%		RING domain
	L208L		X	5%		Coiled coil domain
	L420L		X	3%		SPRY domain
D13	T>C Chr1:33648206		X	29%		Upstream; GATA3/KAP1 Binding site
	D130D		X	6%		Coiled coil domain
	c.*676AGGG> TCCC		X	68%		3'UTR (mir10b)
D14	V364V	X		3%		SPRY domain

Table 9- Full List of *DEAR1* Variants in DCIS FFPE Samples with Adjacent Invasive Components in Exonic and Regulatory Regions. Table contains the list of variants in each sample, indicating the lesion type each variant was found in as well as the allele frequency, validation status, and functional region.

<i>Sample</i>	<i>Mutation</i>	<i>Lesion Type</i>			<i>% Allele frequency</i>	<i>Validated</i>	<i>Area of DEARI</i>
		<i>Pure DCIS</i>	<i>DCIS of DCIS/ INV</i>	<i>Invasive of DCIS/ INV</i>			
PD12	R187W	X			13%	Yes	Coiled coil domain
	D240N	X			81%	Yes	Coiled coil domain
D01	T>C Chr1:33648206			X	16%		Upstream; GATA3/KAP1 Binding site
D05	T>C Chr1:33612419		X	X	3% DCIS 15% INV		Upstream; GATA3/KAP1 Binding site
	c.*676AGGG> TCCC			X	43%		3'UTR (mir10b)
D08	G>A Chr1:33648250		X		15%		Upstream; GATA3/KAP1 Binding site
D10	c.*676AGGG> TCCC			X	22%		3'UTR (mir10b)
D11	D130D			X	16%		Coiled coil domain
D12	G>A Chr1:33648250		X	X	80% DCIS 6% INV		Upstream; GATA3/KAP1 Binding site

<i>Sample</i>	<i>Mutation</i>	<i>Lesion Type</i>			<i>% Allele frequency</i>	<i>Validated</i>	<i>Area Within DEARI</i>
		<i>Pure DCIS</i>	<i>DCIS of DCIS/ INV</i>	<i>Invasive of DCIS/ INV</i>			
D13	T>C Chr1:33648206			X	29%		Upstream; GATA3 Binding site
	c.*676AGGG> TCCC			X	68%		3'UTR (mir10b)

Table 10- List of High Variant Frequency (>10%) *DEARI* Variants in Functionally Important Areas. Table contains the list of variants in each sample that occurred at a higher allele frequency greater than 10%, indicating the lesion type each variant was found in as well as the validation status and functional region.

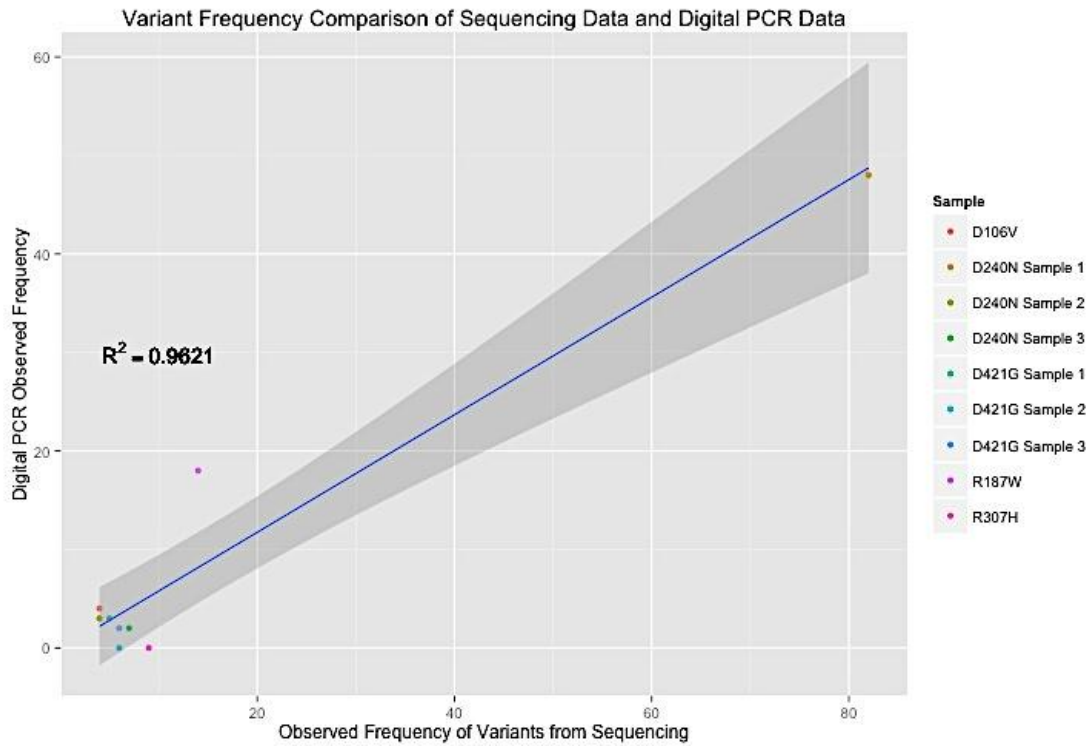


Figure 9- Validation of *DEARI* Variants Found Through Ultra-Deep Targeted Next Generation Sequencing by Digital PCR. A number of exonic variants found in sequencing were verified through the use of a TaqMan based digital PCR assay. Figure details the correlation of the observed frequency of the spike-in to the observed frequency detected by digital PCR.

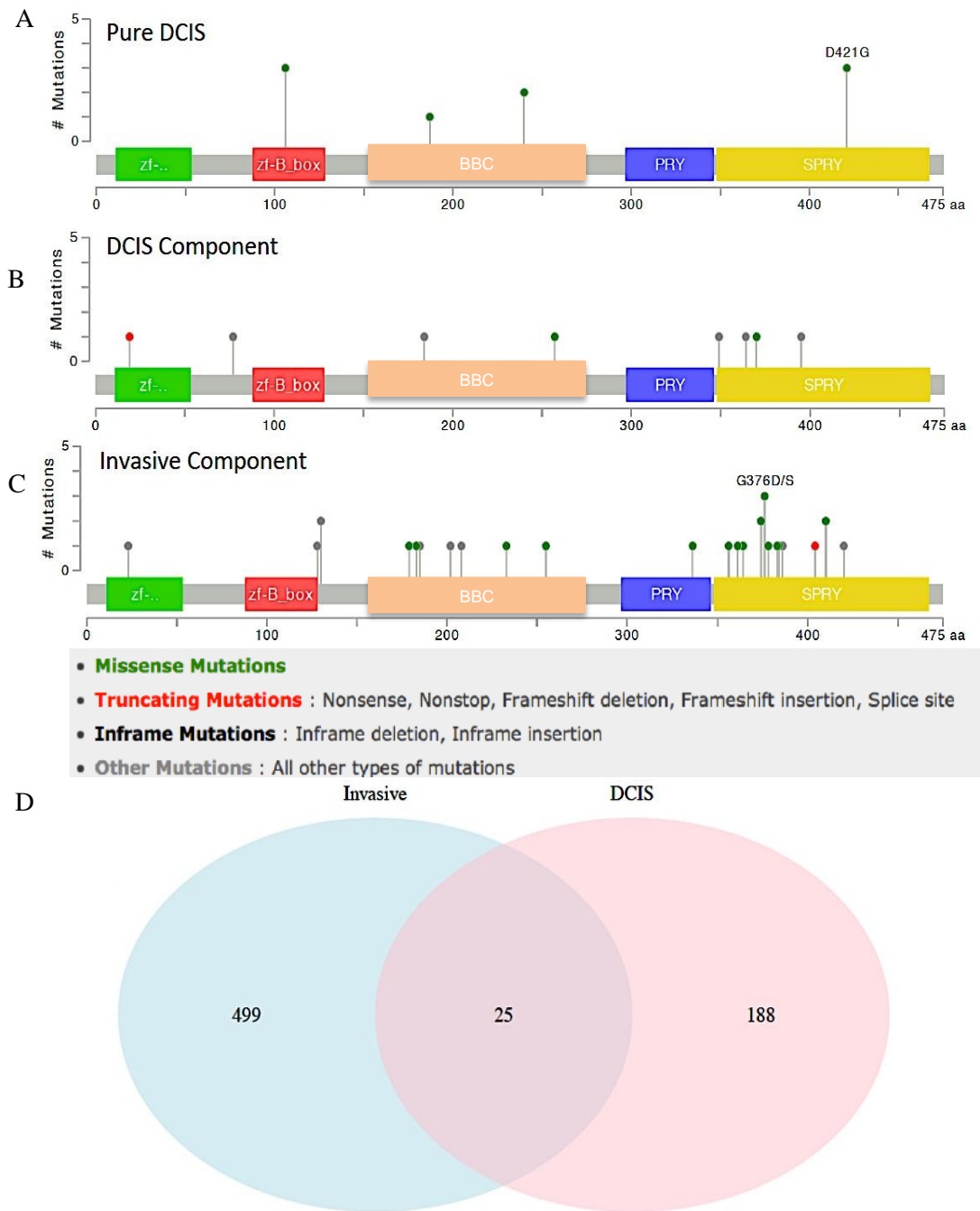


Figure 10- *DEAR1* Experiences Exonic Variants in DCIS of Which Few are Shared Within Lesion Components. Mutations mapper from cBio was used to visualize exonic variants from A) Pure DCIS and matched B) DCIS and C) Invasive components. D) Venn Diagram indicated the number of shared and private variants within samples.

For those variants not harbored in the intron, Pure DCIS samples exhibited a frequency of 12% of variants within the 3' UTR region and 8% of variants contained in the exons. The exonic variants within the pure DCIS samples tended to be repetitive in nature with multiple samples exhibiting missense variants at the same codon. Moreover, the missense variants were overwhelmingly predicted to be deleterious by PolyPhen2 and/or SIFT, with most (64%) showing a consensus by both functional prediction tools (Fig. 8b). Interestingly, no synonymous or benign variants were found in the pure DCIS samples. However, these samples contained notable exonic variants previously found by sequencing efforts completed by our lab and by The Cancer Genome Atlas (TCGA) project. These include missense variants D106V, R187W, D240N and D421G (Fig. 10a). Variants introducing the D106V missense mutation (c.317A>T), which is localized to the zinc finger domain in exon 1, were confirmed to be found in 3 Pure DCIS samples and was formally found in an invasive breast tumor from another cohort². The D106V variant existed sub-clonally manner within the samples as the variant allele frequency within the tumors was about 4%. Another variant, the D421G c.1262A>G variant within the exon 5 Spry domain, was discovered and confirmed as another sub-clonal variant (variant allele frequency of 4-7%) in 3 Pure DCIS samples. The D421G variant was previously reported by TCGA (as accessed by cBIO) to also have existed in a CIMP-high, stage IIIC rectal cancer patient who also exhibited a low level gain of *DEAR1*^{131, 145, 146, 241}. Further, a R187W (c.559C>T) confirmed somatic variant was discovered at a variant allele frequency of 13% in a luminal early onset (47 years old at time of diagnosis) patient from the Pure DCIS cohort. This same variant was also previously found in a metastatic breast cancer cell line (21MT) derived from an early onset patient as well as an invasive breast cancer tumor from another cohort^{2, 131}. The R187W mutation, harbored within the B-Boxed Coiled coil (BBC) domain of exon 3, was earlier reported by our lab to be important for abnormal acinar morphogenesis in the 21MT cell line¹³¹. Excitingly, the early onset patient who exhibited the R187W variant also contained a potential germline missense variant, c.718G>A D240N, which resided in the same domain as the R187W variant. Presence of the germline status of the variant was confirmed by droplet digital PCR from a matched normal adjacent breast tissue. Further, the D240N variant was also found to occur as a somatic variant in another pure DCIS patient who exhibited a more typical age of onset (60 years of age). The identification

of germline variants, especially exonic variants, with DCIS further supports the role of DEAR1 as a classic tumor suppressor that is important in breast cancer.

Spectrum of Variants Found in DCIS with Invasive Components is More Complex than Pure DCIS and Share Relatively Few Variants between the Adjacent in Situ and Invasive Lesions

Sequencing of DCIS lesions associated with IDC components revealed a much more complex spectrum of variants (Fig. 10b/c). Strikingly, the DCIS/INV samples showed a great reduction in the frequency of variants within repetitive codons (13%; 4/30 codons) in comparison to the pure DCIS samples (75%; 3/4 codons) (Fig. 10a/b). Interestingly, these repetitive codon variants happened to be private only to the invasive components of the DCIS/INV samples. Variants within the DCIS/IDC samples also had a greater diversity in the type of mutation and in their location within the gene. For example, the DCIS/INV samples contained variants in all domains of *DEAR1* and included missense, nonsense and synonymous variants, whereas the Pure DCIS samples were not found to contain variants in the RING and PRY domains of DEAR1, and only consisted of missense variants (Fig 8b). Further, no single variant was found to coexist in both Pure DCIS and in DCIS/INV samples.

When focusing only on the DCIS samples with an adjacent invasive lesion, these variants were for the most part, surprisingly, private to the individual components of the lesions. Only 25 of 712 variants found to be shared between the DCIS and invasive components within the DCIS/INV samples (Fig 10d; Appendix XI). The vast majority of the variants common between the two lesion components were variants deep within the intronic region of the gene (17/25). These intronic variants were not centered in any one area but rather spread throughout the gene's non-coding sequence. The other variants shared between the lesions (8/25) were in areas upstream of the promoter of *DEAR1*. Interestingly, no exonic or regulatory UTR variants were shared between the two components. In comparing the variants private to each lesion component, variants harbored within the DCIS component were largely synonymous (63%), whereas synonymous variants played a much smaller role in the variant distribution of the invasive component (32%) (Fig. 7d). Further, the invasive component of the DCIS

lesions exhibited a larger frequency of missense variants within the exonic region featuring the SPRY domain as well as in the region between the BBC and PRY domains compared to the DCIS component. Overall, variants predicted to be deleterious occurred at a much higher frequency with the invasive components (54%) when compared to the matched DCIS components (37%) (Fig. 8b). Sequencing of the invasive lesions indicated that 64% of the total variants resulted in missense changes, of which 78% of these missense variants (50% of total) were predicted to be deleterious by at least one of the deleterious prediction tools tested. In contrast, only 24% percent of the DCIS component variants were missense variants, yet a 100% of these missense variants were predicted to be deleterious. Therefore, though the adjacent invasive lesion showed a higher frequency of missense variants in *DEARI* in general, these variants displayed a varied ability to be tolerated, in contrast to missense variants in the DCIS component of the samples, of which all were predicted to be deleterious. Two codons found to be mutated in the sequencing of DCIS with invasive components were previously reported in Catalogue Of Somatic Mutations In Cancer (COSMIC). The mutation E370K, which was discovered as a private variant with the DCIS component, was also previously reported in endometrial cancer by the TCGA. Further, our sequencing indicated variance at codon 383 inducing a serine to asparagine codon change that was restricted to the invasive component of the particular sample. This codon is the same *DEARI* residue shown to undergo a missense mutation by Abaan OD et al. (S383I) in the large intestine cell line HCT-15²⁴². Further, the first reported *DEARI* nonsense variants described in breast cancer were discovered. These nonsense variants were restricted to DCIS/INV samples (Fig 10 b/c). The two nonsense variants were found in two different samples and occurred as small sub-clonal fractions (3-4% variant allele frequency), with both being private to either the DCIS or invasive component of the lesions. The nonsense variant harbored in the DCIS component occurred in an early onset patient (age 48 years old at diagnosis) at codon 19 in the RING finger domain of exon 1. The nonsense variant within the IDC component of another sample, occurred in an early onset patient (age 44 years old at diagnosis) at codon 404 in the SPRY domain of exon 5. The restriction of the nonsense variants to only early onset samples may hint at a higher propensity for these types of variants to occur in early onset patients. The sequencing of more early onset samples are needed in order to determine if there is a statistical

association between *DEARI* nonsense variants and age of onset. Moreover, sequencing data of the DCIS/INV samples showed four *DEARI* residues exhibiting alteration in more than one sample, with all four residues being restricted to the invasive portion of the tumors. Three of the four codons repetitively altered were missense variants shown to occur at 3-5% variant allele frequency. All three variants (R374C, G376D/S, R410Q) were predicted to be of deleterious nature^{172, 173}. Variants in the c.376 and c.410 amino acids were indicated as having the strongest deleterious scores of the codons exhibiting variants in multiple samples (PolyPhen2: 1.00 and SIFT: 0.00)^{172, 173}. Further, three samples showed the presence of a c.*676AGGG>TCCC variant in the 3'UTR of *DEARI*, in a putative binding site of miR10b. All c.*676AGGG>TCCC variants were restricted to the invasive component of their samples and occurred at a relatively high variant allele frequency (22-68%). Additionally, variants that occurred upstream of *DEARI*'s 5'UTR showed a slight enrichment in areas shown by the ENCODE project to be binding sites for the GATA3 (5/17 samples) and KAP1 (also annotated as TRIM28) (4/17 samples) transcription factors. These variants within the transcription factor binding sites existed at higher variant allele frequencies (8-29%). These variants may indicate the importance of GATA3 and KAP1 in regulating *DEARI*. GATA3 and KAP1 have both been associated with the regulation of migration and therefore, alteration of the transcriptional regulation of *DEARI* by these transcription factors may mediate GATA3 and KAP1's pro-migratory capacity seen in breast cancer^{243, 244}. The invasive components of the DCIS lesions were shown to exhibit multiple variants in both the coding and regulatory regions of *DEARI*, many of which were predicted to be functionally important as well as restricted to the invasive component of the DCIS/INV samples. Therefore, the presence of these deleterious variants restricted to invasive lesions may indicate the significance of these particular variants within *DEARI* to the invasive phenotype exhibited by these lesions.

DEARI Exhibits Functional Mutations in DCIS as Shown by Functional Assays

In order to understand the role that the variants found in *DEARI* play in DCIS and how they may functionally affect the cell, functional assays to assess the deleterious nature of particular mutations were completed. As mentioned previously, a missense variant in 3 Pure DCIS samples (c.317A>T),

which induced a codon change at codon 106 from Aspartic Acid to Valine (D106V) was identified. This variant was located within exon 1, proximal to the RING finger domain. RING finger domains are known to be often critical to the E3 ligase activity exhibited by TRIM family proteins. As such, the RING finger has been shown to be pivotal for the DEAR1 E3 ubiquitin ligase activity against TGF β pathway member SMAD3². Due to the proximity of the D106V variant to the RING finger (40 amino acids downstream from the 3' end of the RING finger), it was determined if this missense variant could affect DEAR1's ability to regulate TGF β signaling. To assess this, a D106V mutant *DEAR1* plasmid was produced through site directed mutagenesis and transiently transfected with a plasmid containing the luciferase response element reporter for TGF β signaling into 293T cells. In the presence of TGF β , the D106V variant was found to greatly reduce DEAR1's known ability to negatively regulate SMAD3 dependent TGF β signaling by increasing SMAD3 transcription factor activity over 4-fold, compared to wild-type *DEAR1* (p=.04) (Fig. 11)². The effect seen on TGF β signaling transduction by D106V was similar to the effect seen upon deletion of the entire exon 1 of *DEAR1*. In contrast, a pancreatic cancer variant, R254Q, harbored within the BBC domain of DEAR1 was not shown to have any effect upon DEAR1's regulation of TGF β signaling. Therefore, the D106V variant may be important in mediating increased TGF β induced signaling, which can be potentially important in promoting migration and invasion in DCIS in the presence of TGF β .

Further, sequencing efforts also revealed the presence of a confirmed somatic missense variant introducing a codon change from arginine to tryptophan at codon 187 (R187W c.559C>T), harbored within the BBC domain of exon 3 in a luminal early onset patient from the Pure DCIS cohort. As previously stated, this variant occurred at a relatively higher allele frequency and was also previously implicated in IDC^{2, 131}. The R187W mutation has been shown to be important for the deregulation of acinar morphogenesis in the 21MT cell line, through genetic complementation¹³¹. To further confirm R187W's ability to affect DEAR1's regulation of acinar morphogenesis, SKBR3 breast cancer (HER2 positive) cells stably expressing mutant *DEAR1* were used for further assessment in 3D culture. The R187W mutation construct was created by site directed mutagenesis. Pooled stably expressing SKBR3

cell lines were generated through transfection of the mutant plasmid and G418 selection. SKBR3 cell line was chosen due to the absence of detectable protein expression of DEAR1 as well as *DEAR1* mutation, as determined by Western and Sanger sequencing, respectively. Moreover, this cell line is known to exhibit altered polarity and abnormal acinar morphogenesis in 3D culture²⁴⁵. Upon culturing the SKBR3 mutant DEAR1 cell lines in 3D culture, the clone exhibiting the R187W variant failed to revert the abnormal acinar morphology of SKBR3 breast cancer cell lines in contrast to the ability of wild-type *DEAR1* to organize the SKBR3 acini (Fig. 12). The minimal degree of acinar organization resulting from the expression of R187W was similar to the degree seen in those cells expressing an empty vector plasmid. Further, in comparison, another variant construct reflecting the pancreatic cancer *DEAR1* mutation R254Q that occurs in the same *DEAR1* domain as R187W, was stably expressed in SKBR3 and showed comparable minimal abilities to revert abnormal acinar organization. The ability of two variants in the BBC domain to subvert DEAR1's regulation of acinar morphology may potentially indicate the importance of the BBC domain in maintaining normal acinar morphology. In all, results indicated that a subset of the variants found in *DEAR1* during the sequencing of DCIS samples are functional and can have profound effects on DEAR1's regulatory capacity of important signaling events and cellular phenotypes.

Discussion

Diagnosis of DCIS has greatly increased in recent years due to the upsurge of mammographic findings²²¹. The lack of biomarkers able to stratify DCIS patients at greater risk for progression has led to a large variance in the methods used to treat these patients, leading to the possibility of both over-treatment and under-treatment. One of the most important clinically relevant questions regarding breast cancer is deciphering the molecular relationship between DCIS and IDC as well as identifying drivers of DCIS progression which are able to predict invasive recurrence. Currently, no reliable and clinically validated biomarkers are available that predict an invasive recurrence after surgical treatment. Multiple histological markers have been proposed to be potential recurrence markers, including tumor size,

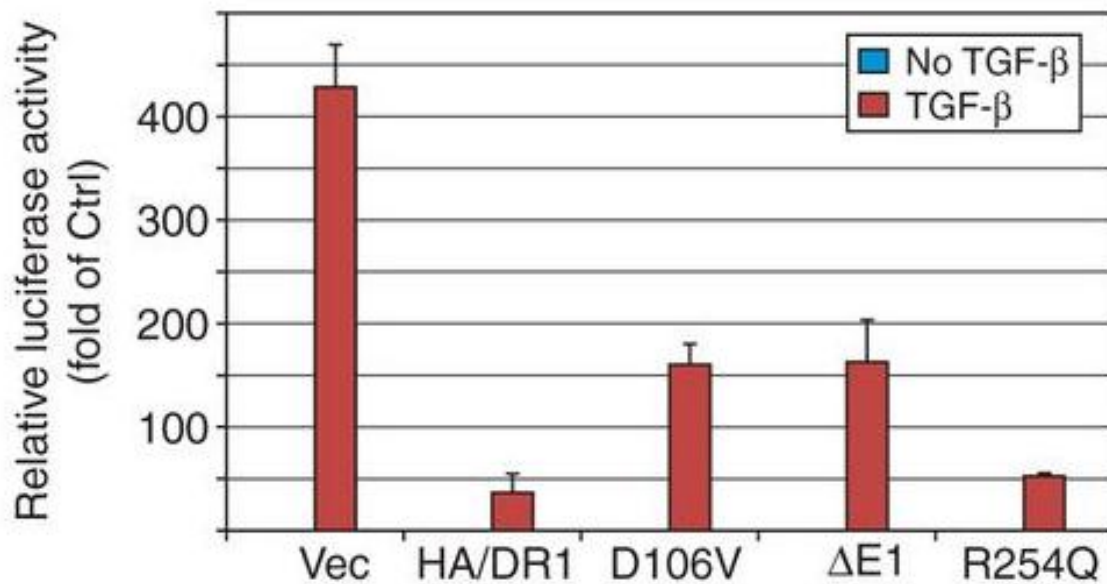


Figure 11- Mutation of *DEARI* Affects TGF- β and SMAD3 Signal Transduction. This figure shows the effect of tumor-derived mutation of *DEARI* on TGF- β signal transduction. Various *DEARI* mutations found by sequencing of breast and pancreatic cancer (D106V, R254Q) along with artificial deletion of the 1st exon of *DEARI* (Δ E1) were co-transfected into HEK293T cells with CAGA12 reporter. After 24 hours, cells were treated with or without TGF- β (1 ng/mL) for 24 hours and luciferase activity was measured. Figure taken from Chen et al. 2013².

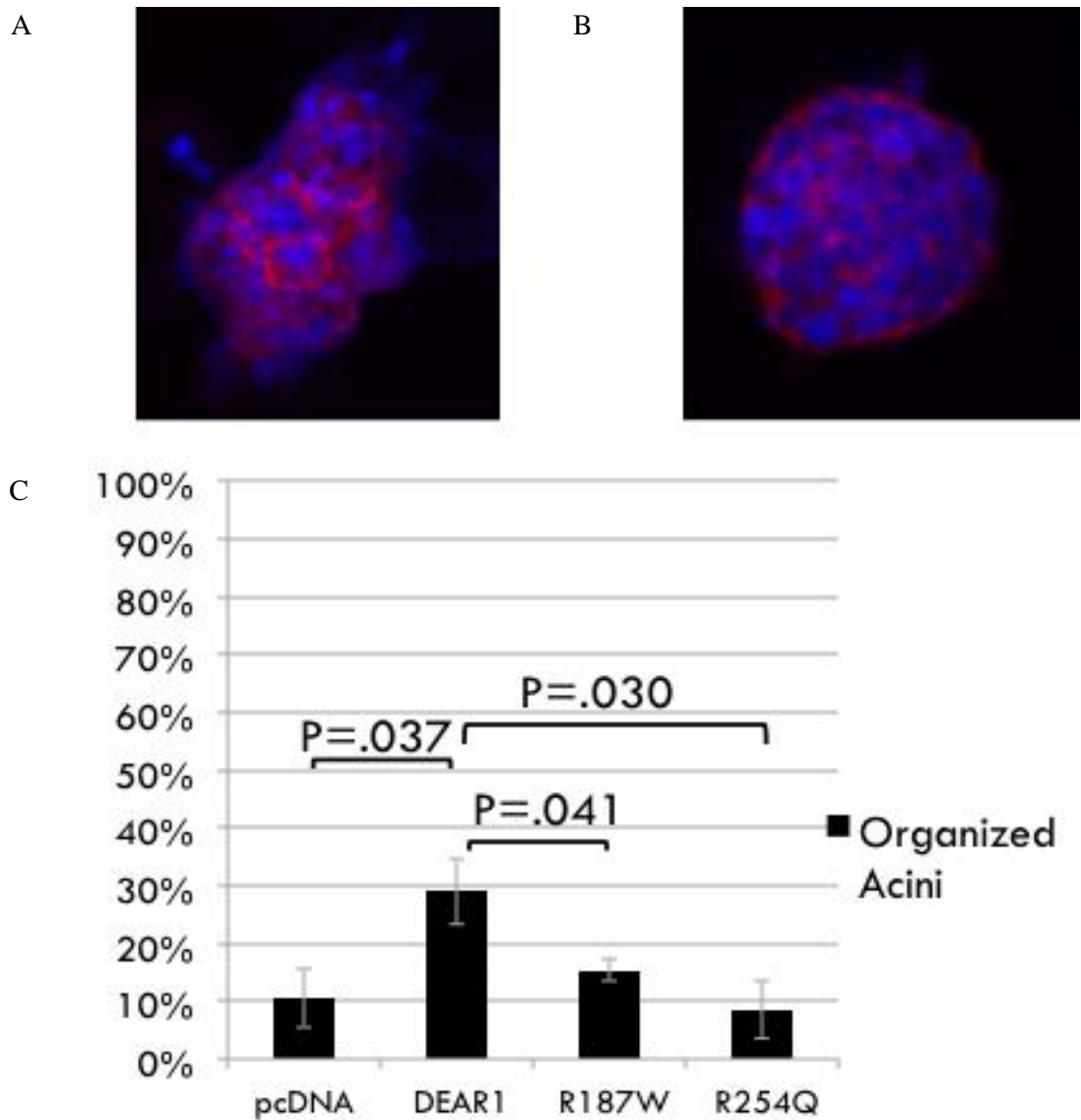


Figure 12- *DEAR1* Variants Can Effect Acinar Morphology in SKBR3 in 3D Culture.

SKBR3 cells stable expressing empty plasmid control, wildtype or mutant *DEAR1* in a pooled manner were grown in 3D culture and immunostained with Alpha-6-integrin (red) and DAPI (blue). Example of disorganized, grape cluster like nature of SKBR3 cells grown in 3D culture shown in A. A proportion of the cultures exhibited a more organized morphology shown in B. C) Quantification of organized population of SKBR3 stable cells.

nuclear grade, and surgical margins²⁴⁶. One of the more promising molecular biomarker studies, the DCIS Score, a derivative of the OncoType-DX predictive biomarker test, was shown to be able to use the gene expression signature of 7 cancer genes to predict recurrence in the non-randomized, prospective, multi-center trial ECOG E5194²⁴⁷. It is important to note that the DCIS Score was tested in a relatively small trial and under specific patient selection criteria. This gene expression biomarker panel study is promising but does have its limitations and has not been currently validated in a large clinical trial. Moreover, the gene expression signature used in the DCIS Score has not resulted in a better understanding of the biology of the disease. Therefore, it is important to continue to search for biomarkers which not only inform on recurrence potential but also on the mechanisms of recurrence, of which can be important for future therapeutic developments. In accordance with this, our lab has sought to understand the molecular and genetic role of *DEAR1* in DCIS progression as well as its possible ability to act as a prognostic marker for DCIS. *DEAR1*, a dominant regulator of polarity and acinar morphogenesis, is known to undergo downregulation at the protein level during the transition from normal breast ductal tissue to DCIS¹³¹. Moreover, *DEAR1* is an E3 ubiquitin ligase for SMAD3, a key signaling molecule in TGF β pathway driven EMT^{2, 220}. As loss of polarity and acquisition of invasive capacity are known hallmarks in tumor acquired EMT, it is possible that *DEAR1* can play a role in the progression of DCIS and recurrence of the disease. As such, the expression of *DEAR1* was found in a cohort of early onset breast cancer to be associated with a significantly higher rate of recurrence free survival¹³¹. It is plausible that alteration of *DEAR1* expression or loss of its normal function via mutations may act as a driver in recurrence. In accordance, *DEAR1* has been shown previously to be mutated and chromosomally lost, by both heterozygous and homozygous deletions, in multiple epithelial cancers including invasive breast cancer^{2, 131}. With this knowledge, it was proposed that genetic alterations of *DEAR1* may play an important role in DCIS progression. To determine this, a custom Ampliseq panel was designed to target the genomic locus encompassing *DEAR1*. This panel was used to complete ultra-deep targeted sequencing of *DEAR1* on the Ion Torrent PGM and Proton platforms. Sequencing of Pure DCIS and DCIS/INV samples with this custom Ampliseq panel found that *DEAR1* is mutated in 71% of DCIS. Other genetic drivers known to be mutated in DCIS, however at lower frequencies, include *PI3KCA*

(50%, n=12), *p53* (21.6%, n=70), *BRCA1* (0.8%, n=369), and *BRCA2* (2.4%, n=369)²⁴⁸⁻²⁵⁰. The degree to which *DEARI* is altered in DCIS highlights the potential importance of genetic alterations in this tumor suppressor in early breast cancer. It is important to note, however, that the techniques utilized in these studies did not have the high sensitivity that the ultra-deep targeted next generation sequencing (NGS) method used in this study had. The application of ultra-deep NGS allows one to see rare sub-clonal populations that may be missed by other sequencing platforms, like Sanger, or other lower coverage applications of NGS, like whole genome or exome sequencing, due to the high depth of coverage able to be achieved. This is also a potential reason for the *DEARI* genetic alteration frequency being significantly higher in our cohort versus sequencing data of IDC from the TCGA cohort (0.1%; n=981) as well as our own previous sequencing data of IDC (13%; n=55)^{131, 145, 146}.

The high degree of sensitivity able to be achieved by ultra-deep targeted sequencing can be important for identifying potentially clinically relevant sub-clonal variants. The use of ultra-deep NGS for sequencing *DEARI* has revealed that many of the variants harbored within the gene occur as rare (<10% variant allele frequency) sub-clonal events. Current debate exists concerning the clinical and functional importance these rare sub-clone variants. It has been suggested that the high degree of intra-tumor heterogeneity and presence of rare sub-clonal populations can serve as a method to increase tumor survival by making the tumor able to withstand different environmental and therapeutic assaults via its high degree of genetic diversity, a known population based survival mechanism within ecology¹⁹⁵. Further, evidence has shown that genetically demarcated sub-clones can have differential functional and morphological features²⁵¹. Multiple factors can influence the determination of which sub-clone acquires dominancy, including therapeutic applications and the acquisition of other genetic drivers that give the clone a large functional advantage. It is the ability of the tumor, by the presence of these sub-clonal populations, to be in a constant state of flux that allows the tumor to have a large advantage towards survival and progression.

Though ultra-deep NGS represents a novel method for detecting these rare sub-clones, a trade-off exists as the high degree of sensitivity able to be achieved by this method can allow errors to appear

to be valid variants which may be introduced during library construction or sequencing. Therefore, due to the use of ultra-deep targeted sequencing to find rare sub-clonal mutations within *DEARI*, stringent variant filtering techniques was utilized to reduce the degree of errors that exist within the data (see methods for detailed filtering techniques). Multiple techniques for variant filtering stringency were tested in order to find the technique that minimized the degree of possible erroneous variant calls while still, in keeping with the discovery nature of the project, maximized the inclusion of variants than can be potentially deleterious (see Appendix III for differential filtering strategies tested and their effect on variant counts). Use of the clinically applicable requirement of at least 1,000x coverage at the variant locus reduced the possibility of spurious variant calls due to areas of low coverage while the requirement that the variant needed to obtain at least 60 supporting reads allowed for a high degree of substantiation of the variant yet still permitting the identification of sub-clonal variants that could be potentially clinically important. Further, variant calls were filtered against data compiled by the 1000 Genome's project to remove any catalogued Single Nucleotide Polymorphisms (SNP) from the final variant list as common polymorphisms are less likely to be pathogenic²⁰⁴. Use of the population variation data assembled by the 1000 Genome's project to assess the pathogenicity based on the variant frequency in a control population has been encouraged by the American College of Medical Genetics and Genomics (ACMG) and the Association for Molecular Pathology (AMP)²⁵². The 1000 Genome's project theoretically accounts for all SNPs that occur in at least 1% of the populations studied, as stated by the project (www.1000genomes.org). However, a small number of variants that are known to occur in less than 1% of the population are still harbored within the data. Therefore, variant filtration to remove common population variants from the sequencing data by filtering against variants catalogued by the 1000 Genomes data will not only remove common SNPs but also may remove rare variants that could possibly be pathogenic. Due to this, variant files were also assessed to determine the presence of variants annotated by the 1000 genomes project which were known to occur in less than 1% of the matched ethnic population of the clinical sample. Variants matching these criteria were retained in the final variant table. Further, half of the clinical samples used in this sequencing study also had matched normal samples, either normal adjacent breast or lymph tissue, sequenced as well. Germline variants were

removed from the final variant table for clinical samples in which the matched normal was able to be sequenced. For those clinical samples for which a matched normal quality DNA sample was unable to be acquired, suspected potential germline variants were unable to be removed due to the lack of confirmatory evidence to their germline status. Suspected germline variants were identified by allele frequencies around 50% (heterozygous) or 100%. These potential germline variants have been marked in the full variant table within the appendix (Appendix X). However, other potential germline variants may exist that cannot be identified by the canonical heterozygous and homozygous allele frequencies, as variances in ploidy and heterogeneity within tumors can affect the variant allele frequencies identified by sequencing. For example, if a germline heterozygous variant undergoes loss of heterozygosity in 30% of tumor cells sequenced, the variant will be present in 65% of sequencing reads of the tumor bulk rather than the previous 50% variant allele frequency that was present at the germline. Therefore, for those samples in which a normal was unable to be sequenced, variants identified in these samples may include germline variants that were unable to be extracted.

The variant spectrum found in Pure DCIS and DCIS with invasive components showed that the nature of the variants found in the lesion types are quite different. Pure DCIS lesions exhibited few exonic variants and of those that existed, the variants seemed to affect a limited number of sites. However, the DCIS associated with an adjacent invasive lesion showed a high degree of variability not only in the functionality of the variants but also in the degree of sites affected by variation. The increase in the number of exonic variants in the samples of DCIS with an invasive component compared to the Pure DCIS could be due to multiple factors. It is possible that DCIS lesions associated with invasive components experience different mutational processes. Both the DCIS and the invasive component of the DCIS/INV samples exhibited an 89-91% reduction in mean variant counts after the removal of C>T and G>A variants versus 40% reduction in the Pure DCIS. A larger enrichment of the C>T transition in the DCIS with invasive component may indicate a role in mutational mechanisms that enrich for these transitions in the pathogenesis of DCIS and its associated invasive lesion. In support of this idea, multiple proposed mutational signatures, many of which are associated with C>T transitions, have been

linked to different mutational mechanism like aging, APOBEC editing, UV radiation, Temozolomide treatment, and mismatch repair defects^{253, 254}. Some of these signatures have been associated with certain subgroups of cancer, like the Signature B single-nucleotide substitution processes as described by Stephens et al. 2012 and its association with ER+ breast tumors²⁵⁵. Moreover, these samples also underwent LCM, which utilizes a UV laser to assist in the removal of the tissue. The use of the UV laser may have induced C>T variants in *DEARI* in those cells exposed to the laser and the small fraction of these associated variants could have been detected due to the use of highly sensitive ultra-deep sequencing. This is supported, as previously mentioned, by the reduction of variant counts per sample when C>T variants are removed, which are also known to be the predominant substitution associated with UV A exposure (Fig 7b). Further, it is important to note that the median rate of follow-up for our Pure DCIS samples was about 5 years. Recent evidence suggests that the rate of recurrence is twice as high within 15 years post-primary tumor development than at 5 years²⁵⁶. Therefore the possibility cannot be ruled out that the Pure DCIS lesions used in this study were later diagnosed with an invasive recurrence.

For those variants that were found by our ultra-deep targeted sequencing of *DEARI*, many of these discovered were predicted by PolyPhen2 and SIFT functional prediction tools to be deleterious. The potential functionality of these variants was also supported by molecular functional studies completed with these variants showing their importance in maintaining acinar morphology and regulation of TGF β -induced signaling. As previously mentioned, our lab has identified DEAR1 as an E3 ubiquitin ligase for TGF β pathway member SMAD3, promoting SMAD3's proteosomal degradation². Multiple TGF β antagonists are currently in various phases of development, such as the TGF- β neutralizing antibody, 2G7, and antisense TGF β oligonucleotide, AP 12009, which are in phase III clinical trials²⁵⁷. *DEARI* mutation, such as the D106V variant, or deletion of the locus, which is known to be common in multiple cancers including invasive breast cancer, may be able to act as a predictive biomarker and thus stratify patients for the efficacy of therapies targeting the TGF β pathway.

Our sequencing efforts of DCIS have uncovered not only novel variants, but also variants that have been previously described by other sequencing efforts to occur in cancer, including breast cancer. The first nonsense variants in *DEAR1* in breast cancer were also described herein. Interestingly, the nonsense variants were restricted to those DCIS samples associated with an invasive component. Nonsense variants are essentially considered to be loss of function variants as they produce a codon change that instills a premature coding stop. The vast majority of transcript products containing nonsense variants are degraded via nonsense-mediated mRNA decay as a quality control method employed by the cell. These variants are one example of how the tumor cell can cause loss of DEAR1 expression, thereby allowing for deregulation of the pathways that DEAR1 regulates. Further, the detection of repetitively altered codons in *DEAR1*, occurring in both Pure DCIS and DCIS associated with invasive components, has been described. The recurring nature of these variants could indicate their importance in breast cancer. The repetitive variants found in our Pure DCIS samples excitingly had been previously described in past sequencing efforts in cancer, including two that are known to occur in breast cancer. Moreover, the incidence of recurring variants in DCIS associated with an adjacent invasive component were found to be, interestingly, restricted only to the invasive portion of the lesion and were all within the SPRY domain of exon 5. This data could suggest not only the significance of these particular variants in invasive breast tumors but also the importance of the SPRY domain in regulating essential pathways in IDC. Further work is needed to understand the detrimental effects of these variants and the role they play in DCIS.

Contrary to expectation, the DCIS samples associated with an invasive component shared very few variants in common between the *in situ* and invasive components. If the DCIS progressed according to the “branched” model of progression, as described in Kaur et al. 2013, it would be expected to find that the vast majority of variants would be shared between the DCIS and invasive components, as has been found for copy number alterations in a number of studies^{195, 235, 248} (Fig. 13a). However, our data supports a model in which the DCIS and invasive components share a very early ancestor and then

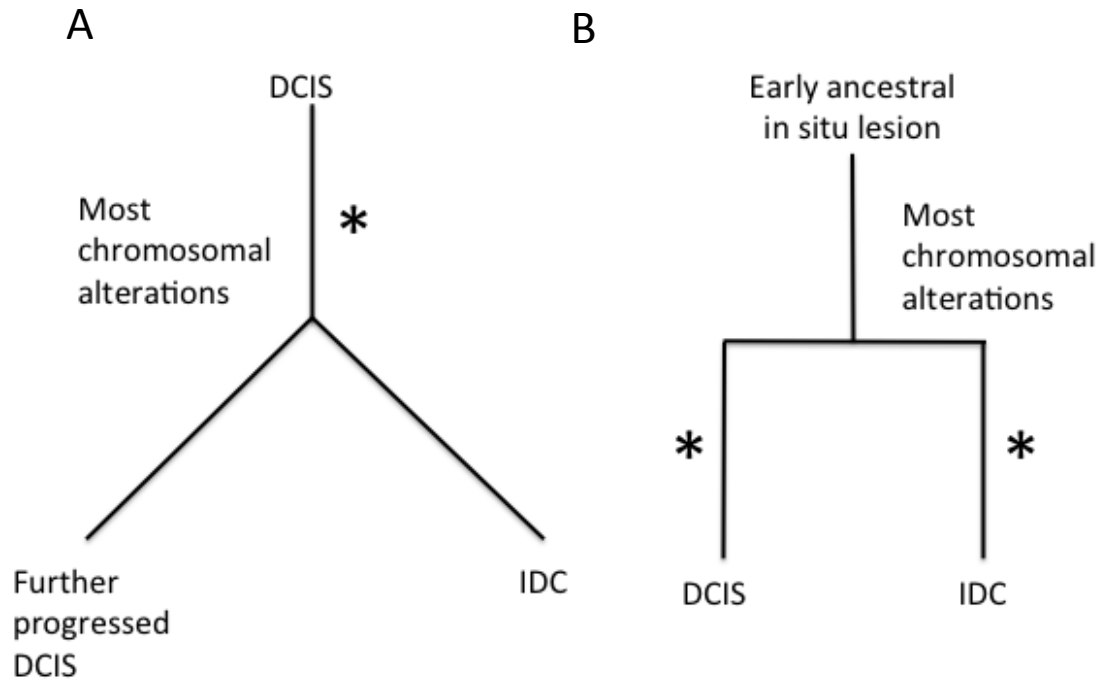


Figure 13- Model of *DEARI* Variants Based on DCIS Progression Models. Both branched and parallel models of DCIS progression have been proposed. The * indicates where *DEARI* mutations would be expected in DCIS within each model. As previous work has shown by other labs, most chromosomal changes are common between both lesions and are therefore an early event in DCIS progression A) This model is representative of the branched model of progression with invasive components (IDC) being a direct lineage from DCIS. Our data is unresponsive of this model as *DEARI* variants were found to mostly exist as private changes in both DCIS and IDC, which is unable to occur if DCIS is a direct ancestor of IDC. B) This model is representative of the parallel model of progression with an earlier lesions acting as the common ancestor for both the in situ and invasive lesions. In this model, the common ancestor splits off to form both DCIS and IDC. Very little variation is expected to be shared between these lesions in this model as the DCIS and the IDC continue to progress in independent yet parallel mechanisms, both accumulating different genetic alterations after the split. Our data supports this parallel model as very few variants in *DEARI* were shared between the two components.

evolve independently. Due to the minute degree to which *DEARI* variants were shared between the DCIS and invasive lesions, the data indicate that the *DEARI* variants mostly occurred after the split from the ancestral clone (Fig. 13b). Due to the few number of variants shared between the lesions, conclusions regarding DEAR1's role in progression from DCIS to invasive disease is unable to be determined. However, our data seems to support the "parallel" model that was first described by Sontag and Axelrod in 2005 describing independent yet parallel evolution of the DCIS and invasive lesions²³⁷. Their model has been empirically supported by data from Newburger et al. 2013, in which they found evidence for "parallel" evolution in 2 of the 3 DCIS lesions sequenced, with the DCIS and invasive tumors sharing more in common with an earlier neoplasia lesion than with each other²⁵⁸. Therefore, future ultra-deep sequencing and molecular studies of earlier neoplasia lesions may reveal sub-clones which distinct genetic alterations that exhibit an increased invasive potential. The clinical implications of the "parallel" model may alter the way that DCIS lesions are currently treated, as the evidence purports that the DCIS lesions are not always the lesion of origin that progresses to invasive carcinoma but rather, the potential for invasion capacity may exist within a sub-clone of a much earlier cell of origin.

Our study has described the high rate of existence of sub-clonal variants within an important tumor suppressor, *DEARI*, including the identification of loss of function variants, in the earliest form of breast cancer through the use of ultra-deep targeted sequencing. Sequencing with our targeted 48kb Ampliseq panel has also provided evidence for the "parallel" model of the relationship between DCIS and IDC. Further studies are needed to definitively identify and empirically show the presence of rare sub-clones within an earlier cell of origin, potentially in a very early neoplastic lesion, which may give rise to a population with an increased propensity to be invasive. Understanding the nature and mechanisms of the sub-clonal populations known to harbor variants in genes important in polarity and invasion, which are critical attributes of cells undergoing EMT, may help us to better understand breast cancer progression and decipher better way of treating this disease.

Discussion

Introduction

Normal cellular homeostasis is dependent on a delicate balance between the expression of genes regulating growth, cell death, and polarity. Interruptions to the balance of these factors can lead to dysplasia and potentially the formation of a carcinoma. The importance of these factors are reflected in their inclusion as components of the hallmarks of cancer⁵². Combination of genetic alterations including deletions and/or loss of function mutations in genes that suppress tumor formation or amplifications and/or translocations in genes which promote oncogenesis can act as integral drivers of tumorigenesis. The cell has multiple mechanisms to inhibit the process of tumor formation. Broadly, genes important to maintaining critical pathways involved in inhibiting growth promotion and invasive capacity are called tumor suppressors.

The classical method of tumor suppressor inactivation was first described by Alfred Knudson in 1971³⁰. Knudson's two hit model based on *Retinoblastoma* inactivation describes the requirement of alteration to both alleles of a tumor suppressor gene in order to complete its full inactivation and the potentiation of tumorigenesis. Twenty-seven years after the description of the two hit model, the discovery of haploinsufficient genes, in which loss of only one allele of the tumor suppressor is required to inactivate the gene, was reported, changing the paradigm of how tumor suppressors are described²⁵⁹.²⁶⁰ Tumor suppressor genes can be disabled by multiple mechanisms including chromosomal deletion of one or both alleles or inactivating mutations like nonsense or frameshift alterations. Further, epigenetic alterations and overexpression of targeting microRNAs have also been described as mechanisms, beyond altering the genetic sequence, which can lead to the loss of expression of tumor suppressor genes. Besides altering expression of these critical tumor suppressors, another method to modify the regulatory nature of these genes is through the existence of single nucleotide changes inducing missense mutations, which may alter the ability of tumor suppressor genes to bind to the gene they regulate. No matter the

mechanism of inactivating the gene, alteration of tumor suppressors can lead to large cellular changes and may affect clinical outcome, as typified by germline mutations in tumor suppressors *TAp53*, *PTEN*, and *BRCA1/2* which can greatly increase the lifetime risk of cancer for these individuals²⁶¹⁻²⁶³.

The TRIM family of annotated E3 ubiquitin ligases, though mostly known for their tumor promoting activities, contain noted tumor suppressors, like PML (TRIM19) which has been shown to be important for the promotion of DNA fidelity by its recruitment of vital DNA repair factors to nuclear bodies after DNA damage and cellular stress²⁶⁴. Examples of other members associated with tumor suppressive activities in cancer are TRIM13 in chronic lymphocytic leukemia, TRIM33 in chronic myelomonocytic leukemia, TRIM24 in hepatocellular carcinoma, and TRIM40 in colorectal cancer (see chapter 1 for further descriptions)⁷². Our lab has previously implicated another important TRIM family protein, DEAR1 (TRIM62), as having strong tumor suppressive activities. In invasive breast cancer (IDC), *DEAR1* expression was significantly associated with local recurrence free survival and was also found to be mutated and homozygously deleted¹³¹. Moreover, *Dear1*^{-/-} and *Dear1*^{-/+} mice exhibited late onset tumor formation in multiple tissues, including lung, mammary, and lymph, indicating the importance of bi-allelic expression of *Dear1*². Our lab has further elucidate two of the significant molecular mechanisms behind *DEAR1*'s tumor suppressive activities including its role as a master regulator of acinar morphogenesis and negative regulator of TGF β mediated migration in breast cancer².

131, 220

Discussion

The work presented herein further implicates *DEAR1* as a tumor suppressor through its characterization of the genetic mechanisms of *DEAR1* inactivation. Our work has described the occurrence of chromosomal loss and mutation of the *DEAR1* locus in multiple cancers including breast, lung, and colorectal adenocarcinomas². Lending support for the clinical importance of chromosomal alterations involving the *DEAR1* locus in these tumor types was the ability of the *Dear1* mouse model to recapitulate the tumor spectrum associated with chromosome 1p loss in humans². Loss of chromosome

arm 1p is a frequent event in epithelial tumors and has been shown to be clinically relevant, as 1p loss is associated with reduced overall survival in breast cancer^{142, 143}. As cancer progression occurs by an accumulation of multiple genetic alterations, the ability for *DEAR1* loss to cooperate with other alterations to affect survival was ascertained. Through database analysis, it was shown that the heterozygous deletion of the *DEAR1* locus could cooperate with amplification of a known epithelial to mesenchymal transition (EMT) driver, *SNAI2*, to predict overall survival in Invasive Breast Cancer (IDC), further highlighting the clinical utility of *DEAR1* chromosomal alterations². As migration and invasion are important steps in late tumor progression, understanding the aberrant genetic changes which promote the movement of cells beyond the basement membrane can help unravel the underlying mechanisms regulating these processes. *SLUG* (*SNAI2*) has been identified as a prominent regulator of EMT, known to cooperate with other drivers of migration such as *Twist* to promote tumorigenesis²⁶⁵. *SNAI2* is amplified and overexpressed in a variety of tumors and its expression has been shown to promote tumor formation in mice²⁶⁶. Combination of alterations effecting different pathways but driving the same cellular process of migration can be powerful mechanisms in the progression of tumorigenesis. This is supported by the strong and significant synergistic effect of the chromosomal heterozygous loss of *DEAR1* and the alteration of *SNAI2* on predicting IDC survival rates. This association has led to the possibility for future development of the use of chromosomal alterations of these two genes, *DEAR1* and *SNAI2*, to act as a prognostic indicator for IDC to help in the stratification of patients for prognosis.

As *DEAR1* has been shown to be important for migration and invasion as well as a possible prognostic indicator in IDC, a better understanding of *DEAR1*'s role in the early *in situ* form of breast cancer is needed. To begin to understand if *DEAR1* is genetically altered in *in situ* breast cancer, ultra-deep targeted sequencing of a 48kb locus encompassing *DEAR1* was completed in Pure Ductal Carcinoma *in Situ* (DCIS) and DCIS associated with invasive components. For this project, a custom targeted *DEAR1* Ampliseq panel was generated, which exhibited strong precision capabilities, as well as a high degree of sensitivity and specificity. The *DEAR1* Ampliseq panel showed a high degree of sensitivity and specificity in regards to single nucleotide variants. The sensitivity and specificity data

also showed our inability to be confident in Insertion/Deletions (INDEL) calls from the Ion Torrent platforms. The low sensitivity of Ion Torrent sequencing platforms to call INDELs has been shown in a variety of publications^{212, 213}. This, however, did not affect our ability to understand the variation in *DEARI* in DCIS as previous analysis has shown that majority of variants seen in the *DEARI* locus, besides loss of the entire p arm of chromosome 1, are single base-pair substitutions (described in chapter 2). Due to the high frequency to which chromosome 1p is lost in epithelial cancer as well as data showing that chromosomal alterations in DCIS are often conserved in its invasive counterpart, it was hypothesized that the vast majority of informative information concerning how *DEARI* variants evolve during breast cancer progression could be found in looking solely at nucleotide alterations^{2, 142, 143, 233}. Further, the analytical performance of the custom sequencing panel generated was characterized by a novel spike in assay designed to assess the accuracy of the allele frequencies of the variants identified in the sequencing data. The novel spike-in assay showed a relatively high rate of recapitulation for the experimental variant allele frequencies in regards to the expected spiked-in allele frequencies. Moreover, our development of this method for determining the accuracy of variant allele frequencies for a sequencing panel can be utilized as an important analytical metric for any targeted sequencing approach. We hope that future research studies as well as potential clinical assays featuring targeted sequencing panels will employ this performance measurement to help researchers fully understand the capabilities of their sequencing panels.

Ultra-deep targeted sequencing of *DEARI* revealed the high frequency (71%) to which the gene is altered in DCIS. Few noted genes have been shown to exhibit mutations at this degree in DCIS, such as *Tp53*, *PIK3CA*, and *BRCA1* (as previously discussed in chapter 4). The high frequency of *DEARI* mutations in combination with the frequent loss of chromosomal 1p may cooperate to fully inactivate *DEARI* in these tumors. Variants discovered during sequencing contained notable exonic variants that were found during previous sequencing completed by our lab and other large sequencing efforts like The Cancer Genome Atlas (TCGA) in epithelial cancers, including breast cancer. Moreover, the presence of the first known nonsense alterations in *DEARI* in breast cancer was described as well as

a potential germline mutation. Potential functionality for many of these variants were suggested by deleterious prediction tools, SIFT and PolyPhen2, as well as by functional assays completed by our lab. It is important to note that particular factors, such as environmental cues, may modify the deleterious nature of some of the *DEARI* variants. For example, the D106V mutation, found both in DCIS and IDC, was shown to potentiate TGF β signaling in the presence of the TGF β ligand, in comparison to wild-type *DEARI*². Variants in *DEARI* that modulate *DEARI*'s inhibitory nature regarding the TGF β pathway may be able to intensify downstream phenotypic effects such as increasing motility and invasive capacity. Further, *DEARI* loss of function mutations may be able to synergize with other genetic defects in cancer to further promote tumorigenesis. For example, it has been shown that heterozygous loss of *DEARI* can cooperate with mutant *Kras* can promote tumorigenesis and metastasis in a Non-Small Cell Lung Cancer mouse model¹³⁸. Moreover, the identification of nonsense and germline mutations in *DEARI* represent hallmarks of classical mechanisms for tumor suppressor inactivation and provide further evidence for the important role that *DEARI* plays in cancer. Additionally, the identification of germline mutations within *DEARI* hint at the potential for *DEARI* to act as a potential stratifier for inherited breast cancer cases. Future work in deciphering *DEARI*'s roles in hereditary breast cancer need to be completed to better understand the frequency of *DEARI* alterations at the germline level. Sequencing for *DEARI* in large cohorts of breast cancer patients with strong family histories and whose genetic predisposition is currently unsolved may help in the molecular diagnosis of these patients. Altogether, this data indicates the significance of *DEARI* variants contributing to the loss of function of this important tumor suppressor in an early stage of breast disease.

Currently, the American Cancer Society has identified a small number of genes known to increase the lifetime risk of breast cancer²²¹. This list includes *BRCA1*, *BRCA2*, *ATM*, *TP53*, *CHEK2*, *PTEN*, *CHD1*, *STK11*, and *PALB2*. Most notable on this list are in *BRCA1* and *BRCA2* as defects in these genes are known to carry the highest risk of breast cancer estimated to be on average between 55-65%²²¹. Despite the known increased risk of breast cancer associated with alterations in these genes, much discussion still exists in the field regarding how these variants should be used clinically. *BRCA1* and

BRCA2 mutation catalogues, family case studies, and experimental *in vitro* testing have been able to identify over a 1000 variants within these genes that are classified as pathogenic and come with the recommendation of increased surveillance and/or potential prophylactic surgery. However, 10-15% of variants discovered during full gene sequencing of *BRCA1* and *BRCA2* are classified as variants of unknown significance (VUS)²⁶⁷. VUS variants are defined as a change in the DNA sequence with unknown effects on the gene product and disease risk²⁶⁸. These variants pose a challenge clinically as there is a possibility that these variants can be deleterious and alter disease risk but yet no clinical action can be recommended as there is no current substantiated evidence for their pathogenicity.

Reclassification of these variants can occur as more information is gleaned from large population variant catalogues and as future additional family case studies, pathological profiles, and functional assays are completed²⁶⁹. Moreover, molecular studies have shown that the location of the variants within *BRCA1* and *BRCA2* can inform on pathogenicity, with the highest levels of pathogenicity being associated with variants within the RING, BRCT, and DNA binding domains²⁶⁹. These domain structures are known to be important for *BRCA1* and *BRCA2*'s functional roles in DNA repair and cell cycle regulation. These cases exemplify the use of molecular based knowledge to inform on the possible deleterious nature and clinical relevance of variants within cancer-related genes. As our study has identified variants in *DEARI* in early stage breast cancer as well as a possible germline variant, understanding how these variants could be important clinically is critical. *DEARI* has been found previously to be a dominant regulator of polarity, acinar morphogenesis and EMT^{2, 131}. Through these studies, the RING finger and the B-Box Coiled Coil (BBC) domains have been found to be integral to the regulation of these pathways. The RING finger has been shown to be necessary for DEAR1's regulation of SMAD3 and TGF β induced EMT². Moreover, genetic complementation studies of a variant in the BBC domain have shown that this domain of DEAR1 can act in a dominant manner in the regulation of acinar morphogenesis¹³¹. Thus, variants within these domains may be more likely to be exert pathogenic effects. Sequencing of *DEARI* in DCIS revealed multiple variants within these two domain structures. Mutations in codons 187 and 240 not only occur in the BBC domain but are predicted to be deleterious by prediction tools PolyPhen2 and SIFT, which use amino acid conservation and structural analysis to predict the deleterious nature of

variants. These BBC domain variants have also been found in both IDC and DCIS samples of multiple breast cancer patients. Moreover, variants at codon 240 have been found to occur within the germline of 2 breast cancer patients and molecular analysis has shown that variance at codon 187 can effect acinar morphogenesis, further giving credence to the pathogenicity of these variants. Another example of a variant with a high probability of pathogenicity is the occurrence of a nonsense variant at codon 19 in the RING domain. Nonsense variant are typically identified as loss of function variants as truncated transcripts which result from nonsense variants are often degraded through nonsense mediated decay. Variants in breast cancer patients occurring in *DEARI* at these codons or possibly within these domains may stratify patients for therapies targeting TGF β or polarity pathways as well as designate the need for increased surveillance for patients who exhibit these variants within the germline. More research however is required to fully grasp the potential ability of *DEARI* variants in informing on clinical aspects, including therapy and breast cancer risk. As more knowledge is gained into how *DEARI* regulates vital cellular processes and how these processes are abrogated in cancer, as well as a continued effort to catalogue variants in *DEARI*, reclassification of these variants from VUS to clinically actionable may occur.

Many of the somatic variants found in *DEARI* in DCIS were often sub-clonal yet private to either the matched *in situ* or invasive lesions within DCIS samples associated with invasive components. This was an unexpected finding as many studies have hinted at the similar profiles of DCIS and invasive lesions^{223, 233, 236,235}. The sequencing of DCIS samples associated with invasive lesions indicated that *DEARI* mutations occur after both the major chromosomal alterations and the evolutionary split from the ancestral clone, thereby giving rise to both the DCIS and the IDC. The private nature of these variants supports the less-heralded “parallel” model of DCIS progression, which described the independent yet parallel evolution of breast tumors that is supported by a few studies^{237, 258}. The study of chromosomal alterations in DCIS and matched IDC has so far proven unfruitful in producing markers of DCIS progression. Recent evidence, including our work, has shown that the study of nucleotide variants may prove to be more beneficial in the search for progression biomarkers. More work, including whole

genome sequencing studies of many DCIS and IDC samples, as well as earlier lesions of breast hyperplasia, are needed to fully elucidate the evolutionary nature of DCIS progression. However, if future studies give further support for an independent evolutionary model for DCIS and IDC, this can have large implications within the clinic. For example, lesions which have been thoroughly vetted to contain no evidence of histological invasiveness may be candidates for lumpectomy only or long-term surveillance as the risk of the development of an invasive lesion in pure DCIS, given the independent evolutionary model, is very low. Further, lesions in which any invasive component exists will require much more extensive treatment like lumpectomy plus radiation as there is a high possibility that the invasive component can develop into a full IDC. It is important to note that there exists a chance that small sub-clonal populations with invasive features may be missed during histological evaluation of the lesions as current practices only include the observation of a few representative histological slides, possibly leading to misdiagnosis of the lesions. If future practices stratify the intensiveness of the treatment options based on the presence or absence of diagnosed invasive components, misdiagnoses of DCIS/INV as pure DCIS in combination with positive surgical margins can increase the risk of recurrence in these patients. Therefore, better and more sensitive histological and molecular markers as well as more intensive histological practices are needed in order to fully vet DCIS lesions for the presence of invasiveness. In all, our work has provided demonstrated the presence of variants in *DEARI*, a tumor suppressor integral in regulating polarity and migration, in the earliest form of breast cancer which may act as drivers in the progression of this disease as well as providing evidence for a “parallel” based evolution model for DCIS and IDC.

Future Directions

Further work to understand the comprehensive effects of *DEARI* variants in DCIS, however, is needed. RNAseq and immunohistochemistry on matched DNA sequencing samples will help to provide knowledge about the expression status of the variants, especially to determine if these variants induce a novel loss or gain of function. Dogma suggests that the nonsense variants found in our DCIS samples associated with invasive components induce a loss of function in *DEARI* as nonsense mutations are

typically thought to induce mRNA decay, leading to decreased gene expression and complete loss of function of the gene in those alleles. Further, evidence supporting the possibility that some of these variants are associated with a novel gain of function has been shown by previous studies completed by the TCGA in which some *DEARI* variants can coincide with chromosomal aberrations or increased gene expression of the variants^{145, 146}. One possible advantage of missense mutations associated with moderate to high gene expression is the possibility that these variants fall in areas important for protein-protein interactions which can potentially disrupt *DEAR1*'s critical negative regulation of tumor promoting pathways, like the TGF β pathway. Moreover, further sequencing analysis of important tumor suppressors and oncogenes can help to define new pathways that *DEAR1* is associated with by delineating whether these genes exhibit a co-occurring or mutually exclusive nature with *DEARI* variants.

Determining if *DEARI* variants have the ability to stratify populations at higher risk is also a critical undertaking. Most of the samples within the current study showed hormone receptor status associated with luminal breast cancer. Luminal type breast cancer is the most common subtype of breast cancer, accounting for 60% of all tumors, with luminal A subtype being more prevalent than luminal B²⁷⁰. A large sample population representing equal numbers of the breast cancer subtypes would be helpful in understanding if *DEARI* experiences differential mutation rate or spectrum in the different subtypes. This is especially important to understand if *DEARI* variants showed a different mutation pattern in the more aggressive breast cancer subtypes, like HER2 and basal, and therefore could be used to potentially further stratify these populations for prognosis²⁷¹. Further, racial disparity is a known epidemiological risk factor in breast cancer. African American women show a somewhat similar incident rates of breast cancer as Caucasians; however, breast cancer within the African American community tends to be more aggressive and incur higher mortality rates than Caucasians²⁷². Determining the disparate molecular and genetic drivers unique to breast cancer in African American women will be very important to understanding what factors lead to a more aggressive behavior of these tumors. The current study is limited in this scope due to its inclusion of mostly Caucasians. As mutation or loss of *DEARI* expression has been shown to allow for an increased propensity for migration and invasion, describing

DEAR1's potential role in the more aggressive phenotype of African American breast cancer will be important. Defining new molecular markers for sub-groups of breast cancer can be vital for future therapeutic and prognosis stratification efforts.

The interplay between the tumor and the microenvironment surrounding it may also be important to understanding why invasive lesions associated with a DCIS counterpart tend to be less aggressive than pure IDC lesions. Wong et al. 2010 has shown that the nuclear grade and degree of proliferation is significantly reduced in IDC associated with DCIS in comparison to pure IDC²⁷³. Further, the same group found an inverse relationship between the size of the DCIS component and aggressiveness of its matched adjacent IDC. Pure IDC has also been associated with larger tumors and earlier onset of metastasis than DCIS-IDC²⁷⁴. These studies suggest a large degree of interaction between the DCIS and IDC components and their potential to regulate each other. Understanding how the DCIS component can limit the aggressiveness of the invasive lesion may help us to uncover new, targetable areas important to inhibiting the aggressive phenotype.

The field as a whole has much to learn in order for us to fully grasp the mechanism of progression of breast cancer, as fully comprehending the pathway taken by these lesions can have large effects on how these tumors are clinically treated. A study conducted by Newburger et al. 2013 suggested that some invasive breast cancers have more in common with their very early neoplastic lesion counterpart than with a matched DCIS lesion²⁵⁸. Though the sequencing efforts within this study focused on a single gene, *DEAR1*, in matched DCIS and invasive tumors, the limited number of shared variants between the lesions may give some credence to the independent "parallel" progression model. If a "parallel" model of breast cancer progression exists, with atypical ductal hyperplasia or even earlier lesions hosting the common cells of origins for both DCIS and IDC, this may indicate the need to identify factors within these very early, benign states of breast disease that may predispose the lesions to spurring off independent DCIS and/or invasive lesions. Moreover, Newburger et al 2013 also identified a subset of invasive lesions which showed a high degree of commonality with their matched DCIS neoplasia. This idea is supported by other molecular and cytogenetic based studies which have indicated

a preservation of chromosomal aberrations, degree of nuclear atypia, and immunohistochemical expression of hormone receptor status between the DCIS and IDC tumors^{233, 275}. Understanding what intrinsic features are differentially associated with clonal or independent, parallel evolution are important to comprehend the underlying mechanisms of breast tumor progression as well as for stratifying patients for differential treatment strategies. Recently, it has been suggested that the tumor microenvironment can play a role in DCIS progression^{276, 277}. Further work in this area to delineate the potential ability of stromal and immunological factors to influence breast cancer progression, and how these pathways interplay with inherent genetic alterations within these lesions, is greatly needed.

It is obvious that multiple factors are critical in breast cancer progression and our knowledge of this area is still in its infancy. How genetic and immune factors, along with tumor microenvironment signals, integrate to drive the mechanisms behind distinct evolutionary patterns seen in breast cancer is a major question in this field which can have huge clinical implications. Much more work is needed in this area in order to fully comprehend the molecular mechanisms driving tumor progression in breast cancer. Defining these pathways can potentially lead to better prognostic methods and new strategies for treating this disease. Our work has discovered novel variants in DCIS in the *DEARI* tumor suppressor which is important in the regulation of polarity and migration, early essential steps in EMT. The pattern of these variants have also revealed the distinct independent evolution of DCIS and their adjacent invasive components. Future larger studies are needed to fully comprehend the ability of *DEARI* to act as a predictive marker for breast cancer progression as well as to understand the factors important to determining the evolutionary pathway central to DCIS and IDC progression. Despite this, the current study described herein provides the first step to understanding the possible effects of *DEARI* genetic alterations in breast cancer and how these variants potentially evolve during the progression of DCIS.

Appendix 1. Spike-In Plasmid Sequence

FASTA sequenced of the plasmid spiked into NA12878 for determination of allele frequency sensitivity. Substitutions and insertions deviant from the reference sequence are bolded/underlined while deletions are highlighted by underlined spaces. Original plasmid design created by Dr. Steven Lott.

>13AAT37P_1351815

CACTATAGGGCGAATTGGCGGAAGGCCGTCAAGGCCGCATTTTTAGGCTAATGGGCTCTGC
ATCTAAGGTAACGATCAGCTTTGGGTATACCCTCACCTGTCCCCCAGCCTGTGGGCAGGA
AGGTGGTGGCACATGACTATAATGTGTGGACTCACCGACTACCAGGTGCCTTCCACAGGCT
CCAGCATCATCACTGACCAGATCGTTACCTTAGCAGATCAGTGGGAGAAGTAGGAACAGGA
GCCAGACTGACTTCACACTAAGGTAACGATGGCCCCACAGACAAAAAAAGGGGCCAGCC
AACCAGGCTGCTGCTCTCTGAACTCCCCAGGGG_CTCTGCTCTGGGCAGGACCGTGGAGCAC
AGTCTGTTCCCTCCCGCAGCTCCTCTCTGGCATCGTTACCTTAGAGGCAGCTTTCTCTAGAGC
AGGGGTTTCAGGAAGTCTTGGAGATAACTAAGGTAACGATTCATTATACCATCTGC_
_GGTGAAGAAGACAGGGAGCTTTTCCAATGTGTCTGCGGATG
CACAGCACTCAGCACACAGTAAAGGTGTGTAGCTGAATAACGCGCATCAACGCCCTGGATC
GTTACCTTAGGTTGCAGGAGGAAGGAAAGACTCCAGGCTCACCTTGGTCTCTAAGGTAACG
ATCCGCCAGTTGTCGCTTGAGCAGCTGCAGCGC_ _ _
GGTGTGTTCCCGCTCGCTGTCTTGAAGGGCCTGTAGTTGGTCCTTCAGCTCCCTCTATCGTTA
CCTTAG
GAAACACACACAGGGCCGCTGGGCCTCATGGGCCTTCCGCTCACTGCCCGCTTTCCAG

Appendix II. R Code to Extract Barcoded Reads from Unmapped BAM Files

Novel spike-in assay for determination of the accuracy of variant allele frequencies required the extraction of barcoded reads representing the presence of the artificial variants spiked into normal control DNA NA12878. R programming code represents the method used to extract these barcoded reads from unmapped BAM files. R programming code originally written by Dr. Steven Lott.

Extract Barcoded Reads from Unmapped BAM Files

Load the necessary library

```
library(ShortRead)
```

```
## Loading required package: BiocGenerics
## Loading required package: parallel
##
## Attaching package: 'BiocGenerics'
##
## The following objects are masked from 'package:parallel':
##
##   clusterApply, clusterApplyLB, clusterCall, clusterEvalQ,
##   clusterExport, clusterMap, parApply, parCapply, parLapply,
##   parLapplyLB, parRapply, parSapply, parSapplyLB
##
## The following object is masked from 'package:stats':
##
##   xtabs
##
## The following objects are masked from 'package:base':
##
##   Filter, Find, Map, Position, Reduce, anyDuplicated, append,
##   as.data.frame, as.vector, cbind, colnames, duplicated, eval,
##   evalq, get, intersect, is.unsorted, lapply, mapply, match,
##   mget, order, paste, pmax, pmax.int, pmin, pmin.int, rank,
##   rbind, rep.int, rownames, sapply, setdiff, sort, table,
##   tapply, union, unique, unlist
##
## Loading required package: IRanges
## Loading required package: GenomicRanges
## Loading required package: XVector
## Loading required package: Biostrings
## Loading required package: lattice
## Loading required package: Rsamtools
```

```
library(Rsamtools)
```

Import the BAM file to be examined

```
setwd("~/Documents/DEAR1_Data_Analysis/PGM_Data/NCR-11")
bam_file <- scanBam("R_2013_03_04_12_05_26_user_NCR-11-Killary_test_Auto_user_NCR-11-
Killary_test_12.bam")
```

This command will identify the sequences with a barcode still attached and output the results in a file named "modified.sam".

```
hits = vcountPattern("CACACTAAGGTAACGAT", bam_file[[1]]$seq, max.mismatch = 3,
  with.indels = TRUE)
sum(hits)
```

```
## [1] 47
```

```
inds = which(hits %in% c(1))
x <- list(qname = bam_file[[1]]$qname[inds], flag = bam_file[[1]]$flag[inds],
  rname = bam_file[[1]]$rname[inds], strand = bam_file[[1]]$strand[inds],
  pos = bam_file[[1]]$pos[inds], qwidth = bam_file[[1]]$qwidth[inds], mapq =
bam_file[[1]]$mapq[inds],
  cigar = bam_file[[1]]$cigar[inds], mrnm = bam_file[[1]]$mrnm[inds], mpos =
bam_file[[1]]$mpos[inds],
  isize = bam_file[[1]]$isize[inds], seq = bam_file[[1]]$seq[inds], qual =
```

```
bam_file[[1]]$qual[inds])
summary(x)
```

##	Length	Class	Mode
## qname	47	-none-	character
## flag	47	-none-	numeric
## rname	47	factor	numeric
## strand	47	factor	numeric
## pos	47	-none-	numeric
## qwidth	47	-none-	numeric
## mapq	47	-none-	numeric
## cigar	47	-none-	character
## mrnm	47	factor	numeric
## mpos	47	-none-	numeric
## isize	47	-none-	numeric
## seq	47	DNAStrngSet	S4
## qual	47	PhredQuality	S4

```
for (i in inds) {
  cat(c(bam_file[[1]]$qname[i], bam_file[[1]]$flag[i], bam_file[[1]]$rname[i],
        bam_file[[1]]$strand[i], bam_file[[1]]$pos[i], bam_file[[1]]$qwidth[i],
        bam_file[[1]]$mapq[i], bam_file[[1]]$cigar[i], bam_file[[1]]$mrnm[i],
        bam_file[[1]]$mpos[i], bam_file[[1]]$isize[i], as.character(bam_file[[1]]$seq[i]),
        as.character(bam_file[[1]]$qual[i]), "\n"), file = "modified.sam", separator = "\t",
      append = TRUE)
}
```

The resulting "modified.sam" file can be converted back to a bam file with the following command using SamTools:

```
$ samtools view -b -S file.sam>file.bam
```

Appendix III- Differential Testing of Various Variant Filtering Stringencies. Table details the reduction of variant counts through various variant filtering strategies. Variant filtering occurred after the use of the automated variant calling by the Torrent Variant Caller using parameters listed in the methods in chapter 3. Multiple strategies were tested in order to reduce spurious variants with low levels of evidence per variant. Ultimately, coverage depth of at least 1000x with greater or equal to 60 reads representing the variant with at least 30 supporting reads on each forward and reverse strands. Acronyms: DP=Coverage Depth; AO=Alternative allele observations; SAF/SAR=alternative allele observations on the Forward and Reverse strands, respectively; FDP= Flow evaluator Coverage Depth; FAO= Flow evaluator Alternative allele observations; FSAF/FSAR= flow evaluator alternative allele observations on the Forward and Reverse strands

Sample	PD3 Pure DCIS	D05 INV	PD9 Pure DCIS	D11 INV
DP>200, AO>20, SAF/SAR>5	32	303	33	229
FDP>1000, FAO>20, FSAF/FSAR>10	16	262	30	204
FDP>1000, FAO>40, FSAF/FSAR>10	16	262	27	204
FDP>1000, FAO>50, FSAF/FSAR>10	14	262	26	202
FDP>1000, FAO>40, FSAF/FSAR>20	12	256	21	199
FDP>1000, FAO>80, FSAF/FSAR>20	10	254	18	197
FDP>1000, FAO>60, FSAF/FSAR>30	11 66% reduction	249 18% reduction	18 46% reduction	197 14% reduction
FDP>1000, FAO>130, FSAF/FSAR>30	9	238	14	189
FDP>1000, FAO>100, FSAF/FSAR>50	8	235	15	182
FDP>1000, FAO>200, FSAF/FSAR>50	6	212	10	162

71% reduction

30% reduction

70% reduction

29% reduction

Appendix IV. R Code to Calculate Coverage By Amplicon.

As designed sequencing amplicons often produced slightly overlapping fragments, calculation of sequencing coverage by amplicon required manual determination via R programming code. R programming code originally written by Dr. Steven Lott.

Calculate Coverage By Amplicon

To generate coverage statistics, the following bash script should be created using nano or another editor and saved in the folder containing the mapped BAM files.

```
for file in *.bam
do
    coverageBed -abam $file -b IAD33712_Designed.bed -d > temp
    awk -v name="$(basename "$file" .bam)" '{print $0, "\t", \
name}' temp > "$(basename "$file" .bam)_coverage.txt"
    rm temp
done
cat *coverage.txt > final_summary.txt
```

After running the batch coverage script, load the concatenated summary file into R as the variable x.

```
x=read.table("final_summary.txt", head=FALSE)
```

Also, load the standard ggplot2 library for use in upcoming plotting functions.

```
library(ggplot2)
library(grid)
```

Not all of the data is needed so the necessary subset is extracted. In addition, the column names are changed.

```
subTot=subset(x, select=c("V4", "V8", "V9"))
colnames(subTot)[1]="Amplicon"
colnames(subTot)[2]="Coverage"
colnames(subTot)[3]="Sample"
```

To verify, the column names are checked.

```
names(subTot)
```

```
## [1] "Amplicon" "Coverage" "Sample"
```

Next, load the amplicon location file that summarizes bases covered by each amplicon in the BED file.

```
regions=read.table("~/Documents/DEAR1_Coverage/amplicons_byRegion_final.txt", head=TRUE)
```

A merged table is generated to incorporate the gene location (ex Exon 1, Intron, etc) for each amplicon.

```
merged=merge(subTot, regions, by="Amplicon")
```

A summary table is generated that contains the amplicon and sample names followed by the mean coverage. The mean coverage column is renamed MeanCov and a the data is reordered based on increasing MeanCov.

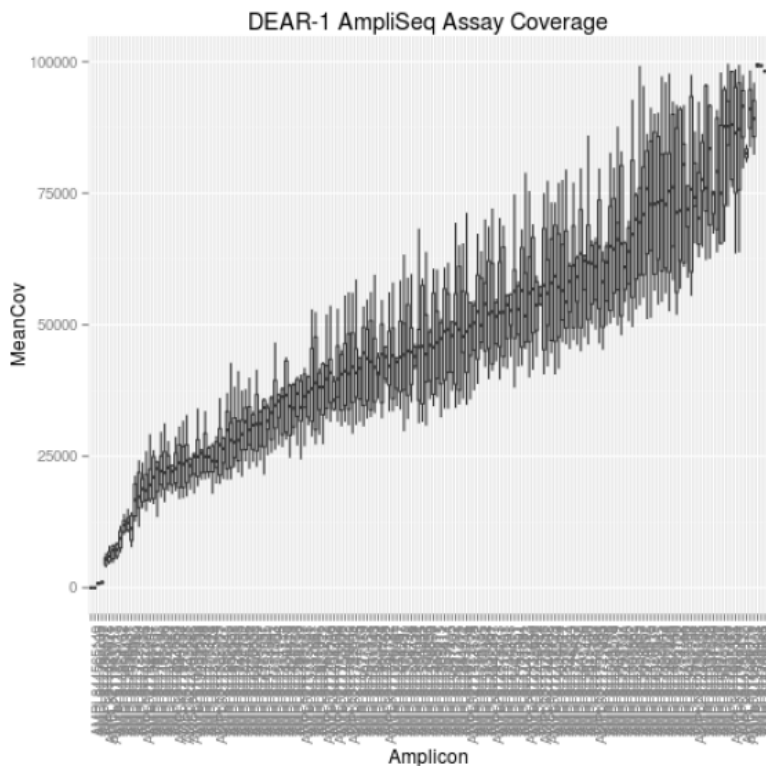
```
library("plyr")
summary=ddply(merged, .(Amplicon, Sample), summarise, mean=mean(Coverage, na.rm=TRUE))
colnames(summary)[3]<-"MeanCov"
summary$Amplicon<-with(summary, reorder(Amplicon, MeanCov, mean))
```

Finally, a plot is generated that shows mean coverage by amplicon as a boxplot.

```
library("ggplot2")
qplot(Amplicon, MeanCov, data=summary, geom="boxplot", na.rm=TRUE,
      outlier.colour="transparent")+
  theme(axis.text.x=element_text(angle=90, hjust=1))+
```



```
ylim(0,100000)+labs(title="DEAR-1 AmpliSeq Assay Coverage")
```



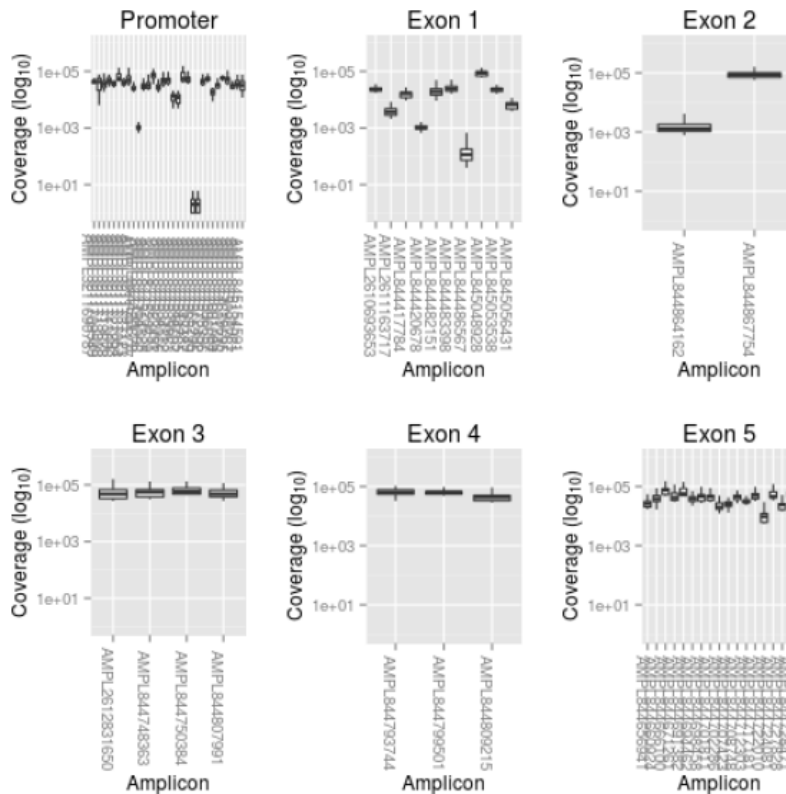
Additionally, the following code will generate a multipanel plot showing amplicon coverage by region of the DEAR1 gene. See end of document for Multiplot.R code that should be saved as a separate file within the working directory.

```
source("Multiplot.R")
promoter <- qplot(Amplicon, Coverage, geom="boxplot", na.rm=TRUE,
  data=subset(merged, Location=="Promoter"),
  outlier.colour="transparent")+
  scale_y_log10(limits=c(1,1e6))+
  theme(axis.text.x=element_text(angle=-90))+
  ggtitle("Promoter")+
  labs(y=expression(paste('Coverage (log'[10], ')'))))
exon1 <- qplot(Amplicon, Coverage, geom="boxplot", na.rm=TRUE,
  data=subset(merged, Location=="Exon 1"), outlier.colour="transparent")+
  scale_y_log10(limits=c(1,1e6))+
  theme(axis.text.x=element_text(angle=-90))+
  ggtitle("Exon 1")+
  labs(y=expression(paste('Coverage (log'[10], ')'))))
exon2 <- qplot(Amplicon, Coverage, geom="boxplot", na.rm=TRUE,
  data=subset(merged, Location=="Exon 2"), outlier.colour="transparent")+
  scale_y_log10(limits=c(1,1e6))+
  theme(axis.text.x=element_text(angle=-90))+
  ggtitle("Exon 2")+
  labs(y=expression(paste('Coverage (log'[10], ')'))))
exon3 <- qplot(Amplicon, Coverage, geom="boxplot", na.rm=TRUE,
  data=subset(merged, Location=="Exon 3"), outlier.colour="transparent")+
  scale_y_log10(limits=c(1,1e6))+
  theme(axis.text.x=element_text(angle=-90))+
  ggtitle("Exon 3")+
  labs(y=expression(paste('Coverage (log'[10], ')'))))
```

```

exon4 <- qplot(Amplicon, Coverage, geom="boxplot", na.rm=TRUE,
  data=subset(merged, Location=="Exon 4"), outlier.colour="transparent")+
  scale_y_log10(limits=c(1,1e6))+
  theme(axis.text.x=element_text(angle=-90))+
  ggtitle("Exon 4")+
  labs(y=expression(paste('Coverage (log' [10], ')')))
exon5 <- qplot(Amplicon, Coverage, geom="boxplot", na.rm=TRUE,
  data=subset(merged, Location=="Exon 5"), outlier.colour="transparent")+
  scale_y_log10(limits=c(1,1e6))+
  theme(axis.text.x=element_text(angle=-90))+
  ggtitle("Exon 5")+
  labs(y=expression(paste('Coverage (log' [10], ')')))
multiplot(promoter, exon3, exon1, exon4, exon2, exon5, cols=3)

```



MultiPlot.R Code

```

# Multiple plot function
#
# ggplot objects can be passed in ..., or to plotlist (as a list of ggplot objects)
# - cols: Number of columns in layout
# - layout: A matrix specifying the layout. If present, 'cols' is ignored.
#
# If the layout is something like matrix(c(1,2,3,3), nrow=2, byrow=TRUE),
# then plot 1 will go in the upper left, 2 will go in the upper right, and
# 3 will go all the way across the bottom.
#
multiplot <- function(..., plotlist=NULL, file, cols=1, layout=NULL) {
  require(grid)

  # Make a list from the ... arguments and plotlist

```

```

plots <- c(list(...), plotlist)

numPlots = length(plots)

# If layout is NULL, then use 'cols' to determine layout
if (is.null(layout)) {
  # Make the panel
  # ncol: Number of columns of plots
  # nrow: Number of rows needed, calculated from # of cols
  layout <- matrix(seq(1, cols * ceiling(numPlots/cols)),
                   ncol = cols, nrow = ceiling(numPlots/cols))
}

if (numPlots==1) {
  print(plots[[1]])
} else {
  # Set up the page
  grid.newpage()
  pushViewport(viewport(layout = grid.layout(nrow(layout), ncol(layout))))

  # Make each plot, in the correct location
  for (i in 1:numPlots) {
    # Get the i,j matrix positions of the regions that contain this subplot
    matchidx <- as.data.frame(which(layout == i, arr.ind = TRUE))

    print(plots[[i]], vp = viewport(layout.pos.row = matchidx$row,
                                   layout.pos.col = matchidx$col))
  }
}
}

```

Appendix V. R Code to Graph Digital PCR Data.

Variants found by sequencing for both control and clinical samples on the Ion Torrent Sequencing platforms were verified by digital PCR for the presence and allele frequencies. Figure shows the R programming code written for the creation of scatter plot plotting sequencing allele frequencies versus digital PCR allele frequencies of the same sample. 95% confidence interval as well as Pearson's correlation coefficient was calculated.

```
###Spike In Confirmation with Digital PCR
NA128.dPCR.v.seq. <- read.csv("~/R fig and code/NA128 dPCR v seq .csv")

#to determine Pearson's product-moment correlation
#focus on print out of p value and sample estimates:cor value
NA128.5.10SpikeIn <- with(dPCRseq, cor.test(expected, observed))
cor.coef.from.slope <- NA128.5.10SpikeIn$estimate
names(cor.coef.from.slope) <- NULL
NA128.5.10SpikeIn

#to create scatter plot with line of best fit and 95% confidence interval shaded
area
ggplot(NA128.dPCR.v.seq., aes(x=sequencing, y=observed)) + geom_text(aes(data
=NULL, x = 9, y = 14, label = paste("R^{2}==.4459")), parse = TRUE) +
geom_point(aes(col=expected)) + stat_smooth(method=lm) + xlab("Observed
Frequency of Spike-Ins from Sequencing") + ylab("Digital PCR Observed
Frequency") + ggtitle("Variant Frequency Comparison of Sequencing Data and
Digital PCR Data")

#####FFPE Variant Confirmation with Digital PCR
FFPE.dPCR.v.seq. <- read.csv("~/R fig and code/FFPE dPCR v seq .csv")

#to determine Pearson's product-moment correlation
#focus on print out of p value and sample estimates:cor value
FFPE <- with(FFPE.dPCR.v.seq., cor.test(sequencing, observed))
cor.coef.from.slope <- FFPE$estimate
names(cor.coef.from.slope) <- NULL
FFPE

#to create scatter plot with line of best fit and 95% confidence interval shaded
area
ggplot(FFPE.dPCR.v.seq., aes(x=sequencing, y=observed)) +
geom_point(aes(col=Sample)) + stat_smooth(method=lm) + geom_text(aes(data
=NULL, x = 10, y = 30, label = paste("R^{2}==.3382")), parse = TRUE) +
xlab("Observed Frequency of Variants from Sequencing") + ylab("Digital PCR
Observed Frequency") + ggtitle("Variant Frequency Comparison of Sequencing
Data and Digital PCR Data")
```

Appendix VI. R Code for Determination of Sensitivity and Specificity.

Sensitivity and specificity of the DEAR1 Ampiseq panel was determined by the R program epiR using the following R programming code. A 95% confidence interval was determined for the performance measures using epiR as well.

```
##R Code for Determination of Sensitivity and Specificity of DEAR1 Ampiseq Panel
  Using epiR

library("epiR", lib.loc="/Library/Frameworks/R.framework/Versions/3.0/Resources/
  library")
read.csv("~/Volumes/plm14/KillaryLab/Databases/DeepSeqData/Sensitivity_Data/
  Sens_PGMFFPEVNIST")
View(Sens_PGMFFPEVNIST)
df <- (Sens_PGMFFPEVNIST)

##outputs data as a 2x2 contingency table
table(df)

##input numbers in matrix(c(Obs+Exp+,Obs+Exp-,Obs-Exp+,Obs-Exp-))
dat<- as.table(matrix(c(62,39,6,233), nrow=2, byrow = TRUE))

##apply the catagories to the matrix
colnames(dat) <- c("Dis+", "Dis-")
rownames(dat) <- c("Test+", "Test-")

##Run the epiR program epi.tests
epi.tests(dat, conf.level = 0.95)
```

Appendix VII –Catalogue of Clinical Samples Sequenced. Table contains a detailed list of sample and sample type/components sequenced for each patient.

	Sample	Path.	Normal Lymph or Adjacent	DCIS	Invasive
1	PD1	Pure DCIS		X	n/a
2	PD2	Pure DCIS		X	n/a
3	PD3	Pure DCIS		X	n/a
4	PD4	Pure DCIS	X-Lymph	X	n/a
5	PD5	Pure DCIS	X-Lymph	X	n/a
6	PD6	Pure DCIS		X	n/a
7	PD7	Pure DCIS		X	n/a
8	PD8	Pure DCIS	X-Lymph	X	n/a
9	PD9	Pure DCIS		X	n/a
10	PD10	Pure DCIS		X	n/a
11	PD11	Pure DCIS		X	n/a
12	PD12	Pure DCIS		X	n/a
13	PD13	Pure DCIS		X	n/a
14	PD14	Pure DCIS		X	n/a
15	PD15	Pure DCIS	X- Lymph	X	n/a
16	PD16	Pure DCIS	X-Adjacent Breast	X	n/a
17	PD17	Pure DCIS	X-Lymph	X	n/a
1	D01	DCIS/ INV	X-Lymph	X	X
2	D02	DCIS/ INV	X-Lymph	X	X
3	D03	DCIS/ INV	X-Lymph	X	X
4	D04	DCIS/ INV	X-Lymph	X	X
5	D05	DCIS/ INV	X-Lymph	X	X
6	D06	DCIS/ INV	X-Lymph	X	X
7	D07	DCIS/ INV	X-Lymph	X	X
8	D08	DCIS/ INV	X-Lymph	X	X
9	D09	DCIS/ INV		X	X
10	D10	DCIS/ INV	X-Lymph	X	X
11	D11	DCIS/ INV		X	X
12	D12	DCIS/ INV		X	X
13	D13	DCIS/ INV		X	X
14	D14	DCIS/ INV		X	X
15	D15	DCIS/ INV		X	X
16	D16	DCIS/ INV	X-Lymph	X	X
17	D17	DCIS/ INV	X-Lymph	X	X
18	D18	DCIS/ INV		Unable to complete Variant Calling	X
19	D19	DCIS/ INV		Unable to complete Variant Calling	X

Appendix IIX. R Code for Creation of Venn Diagram Using the VennDiagram R Program
A Venn diagram was created with R package VennDiagram to visualize the degree of shared variants between the adjacent DCIS and invasive components of matched samples. The following R code was used within the VennDiagram package to complete the graph.

```
##Load VennDiagram Package
> library("VennDiagram", lib.loc="/Library/Frameworks/R.framework/Versions/3.0/
Resources/library")

##Input data for the VennDiagram
##draw.pairwise.venn(Total_Invasive, Total_DCIS, Total Shared)
> draw.pairwise.venn(524, 213, 25, category = c("Invasive", "DCIS"), lty =
  rep("blank",2), fill = c("light blue", "pink"), alpha = rep(0.5, 2), cat.pos
  = c(0,0), cat.dist = rep(0.025, 2), scaled = FALSE)
```

Appendix IX. R Code for Variant Count Per Sample Boxplots.

Visualization of variant counts per sample were created using ggplot2 R package. R programming code used to create the box-plots are shown below as well as the code to determine the mean of each value set.

```
##Loading the Variant Counts
Vcount <- read.csv("~/Desktop/Variant_count_by_sample.csv")
View(Vcount)

#Loading ggplot2 Package
library("ggplot2", lib.loc="/Library/Frameworks/R.framework/Versions/3.0/
Resources/library")

q <- qplot(Lesion.Type, Variant.Count, data=Vcount, geom=c("boxplot","jitter"),
  fill=Lesion.Type, main="Variant Counts per Sample", xlab="", ylab="Variants
per Sample")

#to reorder x axis
q + scale_x_discrete(limit = c("Pure DCIS", "DCIS", "Invasive"))
p <- q + scale_x_discrete(limit = c("Pure DCIS", "DCIS", "Invasive"))

#to zoom in and set ylimit axis to 25
p + coord_cartesian(ylim = c(0,25))

##Calclate the Mean of each X axis catagory
PD <- Vcount[c(1:17),]
D <- Vcount[c(18:34),]
I <- Vcount[c(35:51),]
mean(PD$Variant.Count)
mean(D$Variant.Count)
mean(I$Variant.Count)

#Graphing Variant Count in Lesion Types between Full 48 KB locus, DEAR1 38 KB
locus, and DEAR1 38KB with FFPE Associated Variants Removed
##DataFrame Load
PDgraph <- read.csv("~/Desktop/PureDCISVariantCountLocus.csv")
Dgraph <- read.csv("~/Desktop/DCISVariantCountLocus.csv")
Igraph <- read.csv("~/Desktop/InvasiveVariantCountLocus.csv")
##Calclate the Mean of each X axis catagory
FLPD <- PDgraph[c(1:17),]
D1PD <- PDgraph[c(18:34),]
FFPEPD <- PDgraph[c(35:51),]
FLD <- Dgraph[c(1:17),]
D1D <- Dgraph[c(18:34),]
FFPED <- Dgraph[c(35:51),]
FLI <- Igraph[c(1:17),]
D1I <- Igraph[c(18:34),]
FFPEI <- Igraph[c(35:51),]
mean(FLPD$Variant.Count)
mean(D1PD$Variant.Count)
mean(FFPEPD$Variant.Count)
mean(FLD$Variant.Count)
mean(D1D$Variant.Count)
mean(FFPED$Variant.Count)
mean(FLI$Variant.Count)
mean(D1I$Variant.Count)
mean(FFPEI$Variant.Count)
##Creation of graph
PDplot <- qplot(Locus.Area, Variant.Count, data=PDgraph,
  geom=c("boxplot","jitter"), fill=Locus.Area, main="Variant Counts per Sample
in Pure DCIS", xlab="", ylab="Variants per Sample")
```


Appendix 10 - Detailed List of All Variants Harbored Within the 48kb Locus Involving *DEARI* Found in Ductal Carcinoma *In Situ*. This table contains all variants that were found via sequencing within the 48kb locus harboring *DEARI* and regulatory regions upstream of the gene. Variants that may be associated with formalin fixation are in bold. The criteria for possible fixation induced variants were that the nucleotide change was associated with C>T/G>A nucleotide changes and variant allele frequencies were shown to be at 0.04 or less. C>T/G>A variants have been previously reported to be enriched in formalin fixed samples due to the process causing cytosine deamination. The chosen variant frequency of 0.04 or less for the possible formalin induced changes was used due to the very low possibility that a particular nucleotide would undergo cytosine deamination and having shown that variants at 0.05 variant frequencies were able to be verified by orthogonal methods. Potential germline variants not excluded due to lack of sequenced matched normal sample are italicized.

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within <i>DEARI</i>
PD1	Pure DCIS	33629668	TC> CA	<i>c.504+1404 GA>TG</i>	0.93		<i>Intronic</i>
		33644798	AA> GC	<i>c.408+1828 TT>GC</i>	1.00		<i>Intronic</i>
		33645661	GCT> G	<i>c.408+963_408+964 delAG</i>	0.56		<i>Intronic</i>
		33645665	AA> CT	<i>c.408+961 TT>AG</i>	0.56		<i>Intronic</i>
		33648859	T> C	Upstream	0.08		Upstream
PD2	Pure DCIS	33644798	AA> GC	<i>c.408+1828 TT>GC</i>	1.00		<i>Intronic</i>
		33645661	GCT> G	<i>c.408+963_408+964 delAG</i>	1.00		<i>Intronic</i>
		33645665	AA> CT	<i>c.408+961 TT>AG</i>	1.00		<i>Intronic</i>
PD3	Pure DCIS	33612040	C> G	<i>c.*738 G>C</i>	0.16		3'UTR
		33612944	T> C	p.Asp421Gly <i>c.1262A>G</i>	0.04		Exonic
		33613286	C> T	p.Arg307His c.920G>A	0.03	Failed to validate	Exonic
		33613337	G> C	<i>c.878-9 C>G</i>	0.03		<i>Intronic</i>
		33614654	C> T	<i>c.878-1326 G>A</i>	0.13		<i>Intronic</i>
		33626451	A> T	<i>c.505-906 T>A</i>	0.05		<i>Intronic</i>
		33629610	A> G	<i>c.504+1462 T>C</i>	0.06		<i>Intronic</i>
		33630031	A> G	<i>c.504+1041 T>C</i>	0.05		<i>Intronic</i>
		33633168	A> T	<i>c.409-2001 T>A</i>	0.06		<i>Intronic</i>
		33634506	C> T	c.409-3339 G>A	0.04		Intronic
		33638079	A> T	<i>c.409-6912 T>A</i>	0.14		<i>Intronic</i>
PD4	Pure DCIS	33625332	C> T	p.Asp240Asn <i>c.718G>A</i>	0.07		Exonic
		33633168	A> T	<i>c.409-2001 T>A</i>	0.07		<i>Intronic</i>

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEAR1
PD5	Pure DCIS	33636180	G> A	c.409-5013 C>T	0.05		Intronic
		33638079	A> T	c.409-6912 T>A	0.25		Intronic
		33611279	C> T	c.*1499 G>A	0.05		3'UTR
		33612510	A> G	c.*268 T>C	0.05		3'UTR
		33612944	T> C	p.Asp421Gly c.1262A>G	0.68	Yes	Exonic
		33623802	T> G	c.877+52 A>C	0.03		Intronic
		33638079	A> T	c.409-6912 T>A	0.06		Intronic
		33641976	C> T	c.408+4650 G>A	0.04		Intronic
		33642042	C> T	c.408+4584 G>A	0.04		Intronic
		33645209	C> T	c.408+1417 G>A	0.03		Intronic
		33646717	T> A	p.Asp106Val c.317A>T	0.04		Exonic
		33648831	A> T	Upstream	0.04		Upstream
PD6	Pure DCIS	33612040	C> G	c.*738 G>C	0.10		3'UTR
		33612944	T> C	p.Asp421Gly c.1262A>G	0.06	Yes	Exonic
		33625571	A> G	c.505-26 T>C	0.06		Intronic
		33625581	G> C	c.505-36 C>G	0.04		Intronic
		33633168	A> T	c.409-2001 T>A	0.08		Intronic
		33638074	A> G	c.409-6907 T>C	0.03		Intronic
		33638079	A> T	c.409-6912 T>A	0.19		Intronic
		33641731	C> G	c.408+4895 G>C	0.08		Intronic
33651167	A> G	Upstream	0.03		Upstream		
PD7	Pure DCIS	33611279	C> T	c.*1499 G>A	0.05		3'UTR
		33641485	G> A	c.408+5141 C>T	0.03		Intronic
		33641976	C> T	c.408+4650 G>A	0.03		Intronic
		33646717	T> A	p.Asp106Val c.317A>T	0.04		Exonic
		33648235	G> A	Upstream	0.05		Upstream
PD8	Pure DCIS	33611279	C> T	c.*1499 G>A	0.03		3'UTR
		33611936	C> T	c.*842 G>A	0.03		3'UTR
		33614654	C> T	c.878-1326 G>A	0.07		Intronic
		33616024	C> T	c.878-2696 G>A	0.03		Intronic
		33617588	G> A	c.878-4260 C>T	0.03		Intronic

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEARI
PD8	Pure DCIS	33623802	T> G	c.877+52 A>C	0.07		Intronic
		33625332	C> T	p.Asp240Asn c.718G>A	0.04	Failed to validate	Exonic
		33626876	G> T	c.505-1331 C>A	0.06		Intronic
		33627151	G> A	c.505-1606 C>T	0.03		Intronic
		33627310	C> T	c.505-1765 G>A	0.04		Intronic
		33633167	C> T	c.409-2000 G>A	0.03		Intronic
		33634758	G> A	c.409-3591 C>T	0.03		Intronic
		33634770	G> A	c.409-3603 C>T	0.05		Intronic
		33639129	T> A	c.408+7497 A>T	0.04		Intronic
		33642042	C> T	c.408+4584 G>A	0.05		Intronic
		33644997	G> T	c.408+1629 C>A	0.04		Intronic
		33645290	G> A	c.408+1336 C>T	0.04		Intronic
		33645659	A> T	c.408+967 T>A	0.03		Intronic
		33645665	AA> CT	c.408+961 TT>AG	1.00		Intronic
		33646717	T> A	p.Asp106Val c.317A>T	0.04	Yes	Exonic
		33647348	G> A	Upstream	0.05		Upstream
		33648831	A> T	Upstream	0.14		Upstream
PD9	Pure DCIS	33612520	C> T	c.*258 G>A	0.04		3'UTR
		33621086	T> C	c.877+2768 A>G	0.05		Intronic
		33623802	T> G	c.877+52 A>C	0.05		Intronic
		33626876	G> T	c.505-1331 C>A	0.09		Intronic
		33626913	A> G	c.505-1368 T>C	0.05		Intronic
		33638079	A> T	c.409-6912 T>A	0.17		Intronic
		33639129	T> A	c.408+7497 A>T	0.05		Intronic
		33639131	C> T	c.408+7495 G>A	0.05		Intronic
		33642042	C> T	c.408+4584 G>A	0.05		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645659	A> T	c.408+967 T>A	0.04		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	1.00		Intronic
		33645665	AA> CT	c.408+961 TT>AG	1.00		Intronic
		33645916	C> T	c.408+710 G>A	0.03		Intronic

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEARI
PD9	Pure DCIS	33645918	C> G	c.408+708 G>C	0.08		Intronic
		33646083	C> T	c.408+543 G>A	0.04		Intronic
		33648831	A> T	Upstream	0.12		Upstream
		33649303	A> C	Upstream	0.04		Upstream
PD10	Pure DCIS	33618680	G> T	c.877+5174C>A	0.07		Intronic
		33629668	TC> CA	c.504+1404GA>TG	0.36		Intronic
		33636199	C> A	c.409-5032G>T	0.09		Intronic
PD11	Pure DCIS	33612419	G> A	c.*359 C>T	0.03		3' UTR
		33618738	G> A	c.877+5116 C>T	0.04		Intronic
		33623802	T> G	c.877+52 A>C	0.05		Intronic
		33629668	TC>CA	c.504+1404 GA>TG	0.06		Intronic
		33639129	T> A	c.408+7497 A>T	0.05		Intronic
		33642042	C> T	c.408+4584 G>A	0.04		Intronic
		33642082	G> A	c.408+4544 C>T	0.07		Intronic
		33648831	A> T	Upstream	0.15		Upstream
PD12	Pure DCIS	33618680	G> T	c.877+5174 C>A	0.09		Intronic
		33618799	A> T	c.877+5055 T>A	0.04		Intronic
		33625332	C> T	p.Asp240Asn c.718G>A	0.81	Yes (germline)	Exonic
		33625491	G> A	p.Arg187Trp c.559C>T	0.13	Yes	Exonic
PD13	Pure DCIS	33644798	AA>GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	1.00		Intronic
		33645665	AA>CT	c.408+961 TT>AG	1.00		Intronic
		33648859	T> C	Upstream	0.05		Upstream
PD14	Pure DCIS	33621086	T> C	c.877+2768 A>G	0.04		Intronic
		33623802	T> G	c.877+52 A>C	0.04		Intronic
		33626876	G> T	c.505-1331 C>A	0.11		Intronic
		33628265	T> C	c.505-2720 A>G	0.03		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	0.97		Intronic

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEARI
PD14	Pure DCIS	33645120	A> G	c.408+1506 T>C	0.05		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	0.48		Intronic
		33645665	AA> CT	c.408+961 TT>AG	0.48		Intronic
PD15	Pure DCIS	33611941	G>A	c.*837C>T	0.99		3' UTR
		33645278	G>A	c.408+1348C>T	0.98		Intronic
PD16	Pure DCIS	33612362	A> G	c.*416 T>C	0.05		3' UTR
		33625581	G> C	c.505-36 C>G	0.04		Intronic
		33633168	A> T	c.409-2001 T>A	0.04		Intronic
PD17	Pure DCIS	33611279	C> T	c.*1499 G>A	0.10		3' UTR
		33612132	G> A	c.*646 C>T	0.03		3' UTR
		33612419	G> A	c.*359 C>T	0.04		3' UTR
		33612422	C> T	c.*356 G>A	0.05		3' UTR
		33613346	G> A	c.878-18 C>T	0.03		Intronic
		33616298	G> C	c.878-2970 C>G	0.03		Intronic
		33616750	G> A	c.878-3422 C>T	0.03		Intronic
		33617411	G> T	c.878-4083 C>A	0.03		Intronic
		33627170	C> T	c.505-1625 G>A	0.03		Intronic
		33627359	C> T	c.505-1814 G>A	0.03		Intronic
		33629796	G> A	c.504+1276 C>T	0.03		Intronic
		33631379	C> T	c.409-212 G>A	0.05		Intronic
		33636180	G> A	c.409-5013 C>T	0.08		Intronic
		33638079	A> T	c.409-6912 T>A	0.14		Intronic
		33638577	G> A	c.409-7410 C>T	0.06		Intronic
		33638745	C> T	c.409-7578 G>A	0.04		Intronic
		33638780	T> C	c.409-7613 A>G	0.06		Intronic
		33641731	C> G	c.408+4895 G>C	0.05		Intronic
		33641905	C> T	c.408+4721 G>A	0.03		Intronic
D01	DCIS	33611230	A> G	c.*1548 T>C	0.09		3'UTR
		33612117	G> A	c.*661 C>T	0.03		3'UTR
		33613159	G> A	p.Gly349Gly c.1047C>T	0.04		Exonic
		33629577	G> A	c.505-1682 C>T	0.03		Intronic
		33630773	G> A	c.504+1495 C>T	0.04		Intronic

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEARI
D01	DCIS	33630780	G> A	c.504+299 C>T	0.05		Intronic
		33630813	G> A	c.504+259 C>T	0.03		Intronic
		33631291	C> T	c.409-124 G>A	0.04		Intronic
		33633418	C> T	c.409-2251 G>A	0.04		Intronic
		33638005	G> A	c.409-6838 C>T	0.05		Intronic
		33638013	G> A	c.409-6846 C>T	0.04		Intronic
		33638085	G> A	c.409-6918 C>T	0.03		Intronic
		33638937	C> T	c.408+7689 G>A	0.03		Intronic
		33639060	G> A	c.408+7566 C>T	0.06		Intronic
		33639129	T> A	c.408+7497A>T	0.03		Intronic
		33641447	C> T	c.408+5179 G>A	0.04		Intronic
		33645853	C> T	c.408+773 G>A	0.03		Intronic
		33646036	T> C	c.408+590 A>G	0.05		Intronic
		33647588	C> T	Upstream	0.03		Upstream
		33649491	C> T	Upstream	0.04		Upstream
		33649569	G> A	Upstream	0.04		Upstream
		33649571	G> A	Upstream	0.04		Upstream
		33653042	G> A	Upstream	0.03		Upstream
		33681449	G> A	Upstream	0.03		Upstream
	INV	33639129	T> A	c.408+7497 A>T	0.07		Intronic
		33648206	T> C	Upstream	0.16		Upstream
		33648250	G> A	Upstream	0.04		Upstream
D02	DCIS	33639129	T >A	c.408+7497 A>T	0.04		Intronic
		33648237	A> G	Upstream	0.05		Upstream
	INV	33639129	T >A	c.408+7497 A>T	0.04		Intronic
D03	INV	33638085	G> A	c.409-6918 C>T	0.04		Intronic
		33639129	T> A	c.408+7497 A>T	0.03		Intronic
		33646036	T> C	c.408+590 A>G	0.03		Intronic
D04	DCIS	33611114	G> A	c.*1664 C>T	0.03		3'UTR
		33630779	G >A	c.504+293 C>T	0.03		Intronic
		33630928	C> T	c.504+144 G>A	0.03		Intronic
		33638086	G> A	c.409-6919 C>T	0.04		Intronic

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEARI	
D04	DCIS	33638087	G> A	c.409-6920 C>T	0.04		Intronic	
		33641958	C> T	c.408+4668 G>A	0.02		Intronic	
		33648250	G> A	Upstream	0.10		Upstream	
		33648859	T> C	Upstream	0.04		Upstream	
			33658110	G> A	Upstream	0.05		Upstream
	INV	33611115	G> A	c.*1663 C>T	0.03		3'UTR	
		33612419	G> A	c.*359 C>T	0.04		3'UTR	
		33613056	G> A	p.Arg384Cys c.1150C>T	0.03		Exonic	
		33616887	G> A	c.878-3559 C>T	0.03		Intronic	
		33619885	C> T	c.877+3969 G>A	0.03		Intronic	
		33630779	G >A	c.504+293 C>T	0.03		Intronic	
		33648250	G> A	Upstream	0.05		Upstream	
		33658110	G> A	Upstream	0.04		Upstream	
		D05	DCIS	33612172	T> A	c.*606 A>T	0.04	
33612419				G> A	c.*359 C>T	0.04		3'UTR
33616711	G> A			c.878-3383 C>T	0.03		Intronic	
33616870	G> A			c.878-3542 C>T	0.04		Intronic	
33616875	G> A			c.878-3547 C>T	0.04		Intronic	
33617291	C> T			c.878-3963 G>A	0.04		Intronic	
33618447	G> A			c.878-5119 C>T	0.03		Intronic	
33618473	G> A			c.878-5145 C>T	0.04		Intronic	
33618913	G> A			c.877+4941 C>T	0.03		Intronic	
33628228	G> A			c.505-2683 C>T	0.03		Intronic	
33629192	C> T			c.504+1880 G>A	0.04		Intronic	
33629215	C> T			c.504+1857 G>A	0.03		Intronic	
33629313	G> A			c.504+1759 C>T	0.04		Intronic	
33630776	C> T			c.504+296 G>A	0.03		Intronic	
33630850	C> T	c.504+222 G>A	0.04		Intronic			
33630908	C> T	c.504+164 G>A	0.03		Intronic			
		33630924	C> T	c.504+148 G>A	0.05		Intronic	
		33630930	C> T	c.504+142 G>A	0.06		Intronic	

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEAR1
D05	DCIS	33633418	C> T	c.409-2251 G>A	0.05		Intronic
		33634348	C> T	c.409-3181 G>A	0.03		Intronic
		33634759	G> A	c.409-3592 C>T	0.03		Intronic
		33636288	G> A	c.409-5121 C>T	0.03		Intronic
		33637967	G> A	c.409-6800 C>T	0.03		Intronic
		33638016	G> A	c.409-6849 C>T	0.05		Intronic
		33638082	G> A	c.409-6915 C>T	0.04		Intronic
		33638391	G> A	c.409-7224 C>T	0.05		Intronic
		33638864	C> T	c.409-7697 G>A	0.06		Intronic
		33638955	C> T	c.408+7671 G>A	0.05		Intronic
		33640229	C> T	c.408+6397 G>A	0.03		Intronic
		33641723	C> T	c.408+4903 G>A	0.03		Intronic
		33641954	C> T	c.408+4672 G>A	0.05		Intronic
		33643863	G> A	c.408+2763 C>T	0.04		Intronic
		33644807	C> T	c.408+1819 G>A	0.03		Intronic
		33645245	G> A	c.408+1381 C>T	0.04		Intronic
		33645855	C> T	c.408+771 G>A	0.02		Intronic
		33645870	C> T	c.408+756 G>A	0.04		Intronic
		33645977	T> C	c.408+649 A>G	0.03		Intronic
		33646036	T> C	c.408+590 A>G	0.09		Intronic
		33647350	G> A	Upstream	0.05		Upstream
		33648206	T> C	Upstream	0.03		Upstream
		33648250	G> A	Upstream	0.09		Upstream
		33649483	G> A	Upstream	0.03		Upstream
		33653025	G> A	Upstream	0.03		Upstream
		33653135	G> A	Upstream	0.03		Upstream
	INV	33611942	G> A	c.*836 C>T	0.03		3'UTR
		33612099	CCCT>GGGA	c.*676 AGGG>TCCC	0.43		3'UTR
		33612922	G> A	p.Ile428Ile c.1284C>T	0.04		Exonic
		33612977	C> T	p.Arg410Gln c.1229G>A	0.03		Exonic

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEARI
D05	INV	33612994	C> T	p.Trp404* c.1212G>A	0.03		Exonic
		33613048	G> A	p.Phe386Phe c.1158C>T	0.07		Exonic
		33613079	C> T	p.Gly376Asp c.1127G>A	0.04		Exonic
		33613086	G> A	p.Arg374Cys c.1120C>T	0.05		Exonic
		33613124	G> C	p.Thr361Ser c.1082C>G	0.09		Exonic
		33613140	C> T	p.Val356Met c.1066G>A	0.06		Exonic
		33615544	C> T	c.878-2216 G>A	0.03		Intronic
		33615553	C> T	c.878-2225 G>A	0.03		Intronic
		33615721	C> T	c.878-2393 G>A	0.04		Intronic
		33615741	C> T	c.878-2413 G>A	0.03		Intronic
		33616139	C> G	c.878-2811 G>C	0.08		Intronic
		33616143	G> A	c.878-2815 C>T	0.16		Intronic
		33616177	G> A	c.878-2849 C>T	0.09		Intronic
		33616199	G> A	c.878-2871 C>T	0.04		Intronic
		33616214	G> A	c.878-2886 C>T	0.09		Intronic
		33616247	G> A	c.878-2919 C>T	0.07		Intronic
		33616252	C> T	c.878-2924 G>A	0.03		Intronic
		33616495	C> G	c.878-3167 G>C	0.03		Intronic
		33616610	G> A	c.878-3282 C>T	0.03		Intronic
		33616615	G> A	c.878-3287 C>T	0.05		Intronic
		33616667	C> T	c.878-3339 G>A	0.05		Intronic
		33616672	G> A	c.878-3344 C>T	0.05		Intronic
		33616704	G> A	c.878-3376 C>T	0.03		Intronic
		33616712	G> A	c.878-3384 C>T	0.05		Intronic
		33616739	G> A	c.878-3411 C>T	0.04		Intronic
		33616752	G> A	c.878-3424 C>T	0.06		Intronic

Sample	Lesion Type	Chromosome 1 Hg19 location	Nucleotide Change	Variant	Variant Frequency	Validated	Area within DEARI
D05	INV	33616805	A> G	c.878-3477 T>C	0.04		Intronic
		33616808	G> A	c.878-3480 C>T	0.03		Intronic
		33616822	G> A	c.878-3494 C>T	0.03		Intronic
		33616829	G> A	c.878-3501 C>T	0.07		Intronic
		33616871	G> A	c.878-3543 C>T	0.06		Intronic
		33616890	G> A	c.878-3562 C>T	0.07		Intronic
		33616897	G> A	c.878-3569 C>T	0.06		Intronic
		33616903	G> A	c.878-3575 C>T	0.06		Intronic
		33616908	G> A	c.878-3580 C>T	0.06		Intronic
		33616994	G> A	c.878-3666 C>T	0.03		Intronic
		33617168	C> T	c.878-3840 G>A	0.04		Intronic
		33617287	G> A	c.878-3959 C>T	0.04		Intronic
		33617289	C> T	c.878-3961 G>A	0.03		Intronic
		33617296	G> A	c.878-3968 C>T	0.04		Intronic
		33617325	C> G	c.878-3997 G>C	0.03		Intronic
		33617378	C> T	c.878-4050 G>A	0.06		Intronic
		33617393	T> C	c.878-4065 A>G	0.03		Intronic
		33617396	C> T	c.878-4068 G>A	0.06		Intronic
		33617416	C> T	c.878-4088 G>A	0.05		Intronic
		33617428	C> G	c.878-4100 G>C	0.09		Intronic
		33617605	C> T	c.878-4277 G>A	0.03		Intronic
		33617613	C> T	c.878-4285 G>A	0.04		Intronic
		33617617	G> A	c.878-4289 C>T	0.03		Intronic
		33618734	G> A	c.877+5120 C>T	0.04		Intronic
		33618849	C> T	c.877+5005 G>A	0.04		Intronic
		33618852	G> A	c.877+5002 C>T	0.07		Intronic
		33618876	C> T	c.877+4978 G>A	0.06		Intronic
		33618907	C> T	c.877+4947 G>A	0.06		Intronic
		33618961	C> T	c.877+4893 G>A	0.05		Intronic
		33618964	C> T	c.877+4890 G>A	0.05		Intronic
		33618968	C> T	c.877+4886 G>A	0.07		Intronic
		33621043	G> A	c.877+2811 C>T	0.04		Intronic

<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
		33621066	G> A	c.877+2788 C>T	0.04		Intronic
D05	INV	33621087	C> T	c.877+2767 G>A	0.03		Intronic
		33621105	C> T	c.877+2749 G>A	0.05		Intronic
		33623968	G> A	p.Leu255Phe c.763C>T	0.04		Exonic
		33623984	G> A	c.762-15 C>T	0.04		Intronic
		33624019	C> T	c.762-50 G>A	0.05		Intronic
		33625353	G> T	p.Gln233Lys c.697C>A	0.03		Exonic
		33625444	C> T	p.Ala202Ala c.606G>A	0.04		Exonic
		33625495	C> T	p.Leu185Leu c.555G>A	0.03		Exonic
		33625503	C> T	p.Glu183Lys c.547G>A	0.04		Exonic
		33626859	G> A	c.505-1314 C>T	0.04		Intronic
		33627078	G> A	c.505-1533 C>T	0.03		Intronic
		33627136	G> A	c.505-1591 C>T	0.04		Intronic
		33627170	C> T	c.505-1625 G>A	0.03		Intronic
		33627283	G> A	c.505-1738 C>T	0.04		Intronic
		33627339	C> T	c.505-1794 G>A	0.03		Intronic
		33627370	C> T	c.505-1825 G>A	0.03		Intronic
		33627375	G> A	c.505-1830 C>T	0.04		Intronic
		33627387	C> A	c.505-1842 G>T	0.03		Intronic
		33627405	C> T	c.505-1860 G>A	0.03		Intronic
		33627445	C> T	c.505-1900 G>A	0.03		Intronic
		33628205	G> A	c.505-2660 C>T	0.06		Intronic
		33628224	G> A	c.505-2679 C>T	0.03		Intronic
		33628238	G> A	c.505-2693 C>T	0.05		Intronic
		33629313	G> A	c.504+1759C>T	0.05		Intronic
		33629340	G> T	c.504+1732 C>A	0.07		Intronic
		33629363	G> A	c.504+1709 C>T	0.06		Intronic
		33629494	G> A	c.504+1578 C>T	0.04		Intronic

<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
		33629572	G> A	c.504+1500 C>T	0.05		Intronic
D05	INV	33629715	G> A	c.504+1357 C>T	0.04		Intronic
		33629773	C> T	c.504+1299 G>A	0.04		Intronic
		33629776	G> A	c.504+1296 C>T	0.04		Intronic
		33629777	C> T	c.504+1295 G>A	0.05		Intronic
		33629778	C> T	c.504+1294 G>A	0.04		Intronic
		33629801	G> A	c.504+1271 C>T	0.04		Intronic
		33629843	G> A	c.504+1229 C>T	0.03		Intronic
		33629844	G> A	c.504+1228 C>T	0.05		Intronic
		33629857	C> T	c.504+1215 G>A	0.04		Intronic
		33629872	GG>AA	c.504+1200 CC>TT	0.03		Intronic
		33629897	C> T	c.504+1175 G>A	0.04		Intronic
		33630292	G> A	c.504+780 C>T	0.03		Intronic
		33630716	G> A	c.504+356 C>T	0.04		Intronic
		33630759	G> A	c.504+313 C>T	0.04		Intronic
		33630775	C> T	c.504+297 G>A	0.1		Intronic
		33630778	G> A	c.504+294 C>T	0.09		Intronic
		33630843	C> T	c.504+229 G>A	0.05		Intronic
		33630850	C> T	c.504+222 G>A	0.04		Intronic
		33630881	G> A	c.504+191 C>T	0.05		Intronic
		33630900	C> T	c.504+172 G>A	0.06		Intronic
		33630906	C> T	c.504+166 G>A	0.06		Intronic
		33630913	C> T	c.504+159 G>A	0.03		Intronic
		33631195	G> A	c.409-28 C>T	0.03		Intronic
		33631232	C> T	c.409-65 G>A	0.04		Intronic
		33631277	C> T	c.409-110 G>A	0.05		Intronic
		33631302	C> T	c.409-135 G>A	0.05		Intronic
		33631305	C> T	c.409-1897 G>A	0.06		Intronic
		33633064	C> T	c.409-2116 C>T	0.03		Intronic
		33633283	G> A	c.409-2138 G>A	0.06		Intronic
		33633305	C> T	c.409-3128 C>T	0.09		Intronic
		33634295	G> A	c.409-3191 G>A	0.05		Intronic

		33634358	C> T	c.409-3128 C>T	0.05		Intronic
		33634373	C> G	c.409-3206 G>C	0.03		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D05	INV	33634375	G> A	c.409-3208 C>T	0.05		Intronic
		33634646	C> T	c.409-3479 G>A	0.04		Intronic
		33634651	C> T	c.409-3484 G>A	0.05		Intronic
		33634691	G> A	c.409-3524 C>T	0.03		Intronic
		33634693	G> A	c.409-3526 C>T	0.04		Intronic
		33634756	G> A	c.409-3589 C>T	0.06		Intronic
		33634793	G> A	c.409-3626 C>T	0.03		Intronic
		33634798	G> A	c.409-3631 C>T	0.03		Intronic
		33636146	G> A	c.409-4979 C>T	0.03		Intronic
		33636276	G> A	c.409-5109 C>T	0.05		Intronic
		33636287	A> G	c.409-5120 T>C	0.05		Intronic
		33636289	G> A	c.409-5122 C>T	0.05		Intronic
		33636668	C> T	c.409-5501 G>A	0.03		Intronic
		33636691	G> A	c.409-5524 C>T	0.04		Intronic
		33636713	C> T	c.409-5546 G>A	0.03		Intronic
		33636721	G> C	c.409-5554 C>G	0.03		Intronic
		33636725	C> T	c.409-5558 G>A	0.03		Intronic
		33636738	G> A	c.409-5571 C>T	0.03		Intronic
		33637943	G> A	c.409-6776 C>T	0.05		Intronic
		33637946	G> A	c.409-6779 C>T	0.05		Intronic
		33638005	G> A	c.409-6838 C>T	0.03		Intronic
		33638086	G> A	c.409-6919 C>T	0.05		Intronic
		33638194	C> G	c.409-7027 G>C	0.06		Intronic
		33638199	G> A	c.409-7032 C>T	0.06		Intronic
		33638242	C> T	c.409-7075 G>A	0.08		Intronic
		33638251	C> T	c.409-7084 G>A	0.03		Intronic
		33638258	G> A	c.409-7091 C>T	0.06		Intronic
		33638290	G> A	c.409-7123 C>T	0.06		Intronic
		33638302	G> T	c.409-7135 C>A	0.04		Intronic
		33638370	G> A	c.409-7203 C>T	0.04		Intronic

		33638393	C> T	c.409-7226 G>A	0.04		Intronic
		33638472	C> T	c.409-7305 G>A	0.05		Intronic
		33638475	G> A	c.409-7308 C>T	0.04		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D05	INV	33638518	C> T	c.409-7351 G>A	0.08		Intronic
		33638547	G> A	c.409-7380 C>T	0.07		Intronic
		33638551	CC> TT	c.409-7384 GG>AA	0.07		Intronic
		33638558	C> T	c.409-7391 G>A	0.07		Intronic
		33638608	C> T	c.409-7441 G>A	0.08		Intronic
		33638611	C> T	c.409-7444 G>A	0.04		Intronic
		33638647	C> T	c.409-7480 G>A	0.04		Intronic
		33638744	C> T	c.409-7577 G>A	0.04		Intronic
		33638757	A> C	c.409-7590 T>G	0.05		Intronic
		33638758	C> T	c.409-7591 G>A	0.04		Intronic
		33638762	C> T	c.409-7595 G>A	0.04		Intronic
		33638860	C> T	c.409-7693 G>A	0.05		Intronic
		33638864	C> T	c.409-7697G>A	0.05		Intronic
		33638875	G> A	c.409-7708 C>T	0.04		Intronic
		33638918	C> T	c.408+7708 G>A	0.04		Intronic
		33638982	C> T	c.408+7644 G>A	0.08		Intronic
		33638990	G> C	c.408+7636 C>G	0.08		Intronic
		33639069	C> T	c.408+7557 G>A	0.09		Intronic
		33639083	C> A	c.408+7543 G>T	0.09		Intronic
		33640324	C> T	c.408+6302 G>A	0.04		Intronic
		33640342	G> A	c.408+6284 C>T	0.04		Intronic
		33641453	C> T	c.408+5173 G>A	0.03		Intronic
		33641656	C> G	c.408+4970 G>C	0.03		Intronic
		33641668	C> T	c.408+4958 G>A	0.04		Intronic
		33641874	C> T	c.408+4752 G>A	0.04		Intronic
		33641909	C> T	c.408+4717 G>A	0.04		Intronic
		33641923	C> T	c.408+4703 G>A	0.05		Intronic
		33641994	C> T	c.408+4632 G>A	0.04		Intronic
		33642018	C> T	c.408+4608 G>A	0.04		Intronic

		33642063	C> T	c.408+4563 G>A	0.03		Intronic
		33642109	G> A	c.408+4517 C>T	0.04		Intronic
		33643885	G> A	c.408+2741 C>T	0.04		Intronic
		33643944	G> A	c.408+2682 C>T	0.03		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D05	INV	33644402	C> T	c.408+2224 G>A	0.04		Intronic
		33644432	G> A	c.408+2194 C>T	0.04		Intronic
		33644439	C> T	c.408+2187 G>A	0.05		Intronic
		33644451	C> T	c.408+2175 G>A	0.03		Intronic
		33644474	G> A	c.408+2152 C>T	0.04		Intronic
		33644863	C> T	c.408+1763 G>A	0.04		Intronic
		33644894	G> A	c.408+1732 C>T	0.04		Intronic
		33645020	G> A	c.408+1606 C>T	0.04		Intronic
		33645027	G> A	c.408+1599 C>T	0.04		Intronic
		33645040	C> A	c.408+1586 G>T	0.04		Intronic
		33645090	G> A	c.408+1536 C>T	0.03		Intronic
		33645128	C> T	c.408+1498 G>A	0.07		Intronic
		33645151	C> T	c.408+1475 G>A	0.03		Intronic
		33645183	G> A	c.408+1443 C>T	0.06		Intronic
		33645290	G> A	c.408+1336 C>T	0.04		Intronic
		33645303	C> T	c.408+1323 G>A	0.04		Intronic
		33645307	C> T	c.408+1319 G>A	0.04		Intronic
		33645675	G> A	c.408+951 C>T	0.03		Intronic
		33645696	G> A	c.408+930 C>T	0.04		Intronic
		33645855	C> T	c.408+771G>A	0.04		Intronic
		33645856	C> T	c.408+770 G>A	0.05		Intronic
		33645912	C> T	c.408+714 G>A	0.04		Intronic
		33645932	C> T	c.408+694 G>A	0.04		Intronic
		33645955	C> T	c.408+671 G>A	0.04		Intronic
		33645964	G> C	c.408+662 C>G	0.04		Intronic
		33646035	C> T	c.408+591 G>A	0.03		Intronic
		33646036	T> C	c.408+590A>G	0.06		Intronic
		33646094	C> T	c.408+532 G>A	0.04		Intronic

		33646110	C> T	c.408+516 G>A	0.07		Intronic
		33646271	C> T	c.408+355 G>A	0.03		Intronic
		33647517	G> A	Upstream	0.04		Upstream
		33647526	G> A	Upstream	0.04		Upstream
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D05	INV	33647595	C> T	Upstream	0.04		Upstream
		33647631	C> A	Upstream	0.06		Upstream
		33648206	T> C	Upstream	0.15		Upstream
		33648250	G> A	Upstream	0.09		Upstream
		33648829	G> A	Upstream	0.04		Upstream
		33649473	C> T	Upstream	0.06		Upstream
		33649476	G> A	Upstream	0.06		Upstream
		33649564	C> T	Upstream	0.05		Upstream
		33649574	C> G	Upstream	0.07		Upstream
		33651320	C> T	Upstream	0.05		Upstream
		33651330	C> T	Upstream	0.07		Upstream
		33651397	G> T	Upstream	0.04		Upstream
		33651410	G> A	Upstream	0.06		Upstream
		33653008	C> G	Upstream	0.05		Upstream
		33653022	G> A	Upstream	0.06		Upstream
		33653028	G> A	Upstream	0.07		Upstream
		33653059	G> A	Upstream	0.04		Upstream
		33657843	C> T	Upstream	0.04		Upstream
		33657845	C> T	Upstream	0.06		Upstream
		33657853	C> T	Upstream	0.06		Upstream
		33657922	C> T	Upstream	0.07		Upstream
		33658110	G> A	Upstream	0.03		Upstream
		33658670	G> A	Upstream	0.05		Upstream
		33658796	G> A	Upstream	0.04		Upstream
		33681361	C> T	Upstream	0.04		Upstream
		33681404	C> T	Upstream	0.04		Upstream
		33681418	G> A	Upstream	0.03		Upstream
		33681433	C> T	Upstream	0.04		Upstream

D06	DCIS	33613098	C> T	p.Glu370Lys c.1108G>A	0.04		Exonic
		33625498	C> T	p.Arg184Arg c.552G>A	0.03		Exonic
		33616812	G> A	c.878-3484 C>T	0.03		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D06	DCIS	33638087	G> A	c.409-6920 C>T	0.03		Intronic
		33638582	C> T	c.409-7415 G>A	0.03		Intronic
		33644801	C> T	c.408+1825 G>A	0.04		Intronic
	INV	33611108	G> A	c.*1670 C>T	0.06		3'UTR
		33611283	G> A	c.*1495 C>T	0.05		3'UTR
		33611323	C> T	c.*1455 G>A	0.03		3'UTR
		33612181	C> T	c.*597 G>A	0.03		3'UTR
		33612370	G> A	c.*408 C>T	0.04		3'UTR
		33612419	G> A	c.*359 C>T	0.03		3'UTR
		33613072	G> C	p.Ile378Met c.1134C>G	0.03		Exonic
		33613080	C> T	p.Gly376Ser c.1126G>A	0.03		Exonic
		33613116	C> T	p.Val364Met c.1090G>A	0.03		Exonic
		33616206	G> A	c.878-2878 C>T	0.03		Intronic
		33616478	G> A	c.878-3150 C>T	0.03		Intronic
		33616729	G> A	c.878-3401 C>T	0.04		Intronic
		33616839	G> A	c.878-3511 C>T	0.03		Intronic
		33616870	G> A	c.878-3542 C>T	0.03		Intronic
		33616875	G> A	c.878-3547 C>T	0.03		Intronic
		33616890	G> A	c.878-3562 C>T	0.03		Intronic
		33616898	G> A	c.878-3570 C>T	0.06		Intronic
		33616902	G> A	c.878-3574 C>T	0.04		Intronic
		33619899	C> T	c.877+3955 G>A	0.03		Intronic
		33623976	G> A	c.762-7 C>T	0.03		Intronic
		33625263	G> A	c.761+26 C>T	0.03		Intronic
		33625514	C> T	p.Gly179Asp	0.04		Exonic

<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
		33629315	G> A	c.504+1757 C>T	0.07		Intronic
		33629321	G> A	c.504+1751 C>T	0.04		Intronic
		33629360	G> A	c.504+1712 C>T	0.03		Intronic
D06	INV	33630778	G> A	c.504+294 C>T	0.09		Intronic
		33630913	C> T	c.504+159 G>A	0.03		Intronic
		33630919	C> T	c.504+153 G>A	0.06		Intronic
		33630920	C> T	c.504+152 G>A	0.05		Intronic
		33630930	C> G	c.504+142 G>C	0.03		Intronic
		33633308	C> G	c.409-2141 G>C	0.05		Intronic
		33634282	G> A	c.409-3115 C>T	0.04		Intronic
		33634288	G> A	c.409-3121 C>T	0.04		Intronic
		33634294	G> A	c.409-3127 C>T	0.04		Intronic
		33636640	G> A	c.409-5473 C>T	0.07		Intronic
		33636663	G> A	c.409-5496 C>T	0.03		Intronic
		33638086	G> A	c.409-6919 C>T	0.03		Intronic
		33638217	G> A	c.409-7050 C>T	0.03		Intronic
		33638393	C> T	c.409-7226 G>A	0.03		Intronic
		33638506	C> T	c.409-7339 G>A	0.05		Intronic
		33638844	C> T	c.409-7677 G>A	0.03		Intronic
		33638868	C> T	c.409-7701 G>A	0.03		Intronic
		33638932	C> T	c.408+7694 G>A	0.03		Intronic
		33639142	C> T	c.408+7484 G>A	0.04		Intronic
		33641558	C> T	c.408+5068 G>A	0.04		Intronic
		33641940	C> T	c.408+4686 G>A	0.03		Intronic
		33642164	C> G	c.408+4462 G>C	0.02		Intronic
		33645318	G> A	c.408+1308 C>T	0.03		Intronic
		33645919	C> T	c.408+707 G>A	0.03		Intronic
		33646965	G> A	p.Ser23Ser c.69C>T	0.04		Exonic
		33648722	G> A	Upstream	0.04		Upstream
		33649476	G> A	Upstream	0.04		Upstream

		33649511	G> A	Upstream	0.04		Upstream
		33657845	C> T	Upstream	0.03		Upstream
		33657889	C> T	Upstream	0.03		Upstream
D07	DCIS	33639129	T> A	c.408+7497 A>T	0.04		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D07	INV	33648237	A> G	Upstream	0.05		Upstream
D08	DCIS	33639129	T> A	Upstream	0.05		Upstream
		33648250	G> A	Upstream	0.15		Upstream
	INV	33611230	A> G	c.*1548 T>C	0.04		3'UTR
		33616852	G> A	c.878-3524 C>T	0.04		Intronic
		33629318	C> T	c.504+1754 G>A	0.03		Intronic
		33638876	G> C	c.409-7709 C>G	0.03		Intronic
		33639129	T> A	Upstream	0.05		Upstream
		33657915	A> G	Upstream	0.03		Upstream
D09	DCIS	33612473	T> C	c.*305 A>G	0.14		3'UTR
		33629668	TC> CA	c.504+1404 GA>TG	0.46		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	0.51		Intronic
	INV	33629668	TC> CA	c.504+1404 GA>TG	0.49		Intronic
		33639129	T> A	c.408+7497 A>T	0.05		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	0.52		Intronic
		33648859	T> C	Upstream	0.05		Upstream
D10	DCIS	33639129	T> A	c.408+7497 A>T	0.04		Intronic
	INV	33644798	AA> GC	c.408+1828 TT>GC	1		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	1		Intronic
D11	DCIS	33612419	G> A	c.*359 C>T	0.04		3'UTR
		33612452	G> A	c.*326 C>T	0.03		3'UTR
		33613021	G> A	p.Asn395Asn c.1185C>T	0.03		Exonic
		33613415	G> A	c.878-87 C>T	0.04		Intronic
		33616060	C> T	c.878-2732 G>A	0.03		Intronic
		33616104	C> T	c.878-2732 G>A	0.03		Intronic

		33616166	C> T	c.878-2776 G>A	0.03		Intronic
		33616221	G> A	c.878-2838 G>A	0.03		Intronic
		33616233	G> A	c.878-2893 C>T	0.03		Intronic
		33616342	G> A	c.878-2905 C>T	0.03		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D11	DCIS	33616344	G> A	c.878-3016 C>T	0.04		Intronic
		33616397	G> A	c.878-3069 C>T	0.03		Intronic
		33616513	G> A	c.878-3185 C>T	0.03		Intronic
		33616666	G> A	c.878-3338 C>T	0.05		Intronic
		33616684	G> A	c.878-3356 C>T	0.03		Intronic
		33616712	G> A	c.878-3384 C>T	0.04		Intronic
		33616808	G> A	c.878-3480 C>T	0.04		Intronic
		33616864	G> A	c.878-3536 C>T	0.03		Intronic
		33616887	G> A	c.878-3559 C>T	0.04		Intronic
		33616891	G> A	c.878-3563 C>T	0.04		Intronic
		33618646	G> A	c.877+5208 C>T	0.11		Intronic
		33618721	G> A	c.877+5133 C>T	0.03		Intronic
		33618922	C> T	c.877+4932 G>A	0.03		Intronic
		33621020	C> T	c.877+2834 G>A	0.03		Intronic
		33623961	C> T	p.Gly257Glu c.770G>A	0.03		Exonic
		33623985	G> A	c.762-16 C>T	0.03		Intronic
		33625122	C> T	c.761+167 G>A	0.03		Intronic
		33625142	C> T	c.761+147 G>A	0.04		Intronic
		33625456	C> T	p.Glu77Glu c.231G>A	0.03		Exonic
		33625665	G> A	c.505-120 C>T	0.04		Intronic
		33627101	C> T	c.142-1556 G>A	0.02		Intronic
		33627241	C> T	c.505-1696 G>A	0.04		Intronic
		33627263	G> A	c.505-1718 C>T	0.04		Intronic
		33627291	C> T	c.505-1746 G>A	0.03		Intronic
		33628207	G> A	c.505-2662 C>T	0.04		Intronic
		33629578	G> A	c.504+1494 C>T	0.03		Intronic

		33630772	G> A	c.504+300 C>T	0.08		Intronic
		33630775	C> T	c.504+297 G>A	0.05		Intronic
		33630908	C> T	c.504+164 G>A	0.04		Intronic
		33630925	C> T	c.504+147 G>A	0.05		Intronic
		33633303	C> T	c.409-2136 G>A	0.05		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D11	DCIS	33633416	G> A	c.409-2249 C>T	0.03		Intronic
		33634282	G> A	c.409-3115 C>T	0.03		Intronic
		33634785	C> T	c.409-3618 G>A	0.03		Intronic
		33636295	G> A	c.409-5128 C>T	0.05		Intronic
		33637997	G> A	c.409-6830 C>T	0.03		Intronic
		33638004	G> A	c.409-6837 C>T	0.09		Intronic
		33638021	G> A	c.409-6854 C>T	0.04		Intronic
		33638048	G> A	c.409-6881 C>T	0.03		Intronic
		33638086	G> A	c.409-6919 C>T	0.05		Intronic
		33638087	G> A	c.409-6920 C>T	0.04		Intronic
		33638181	G> A	c.409-7014 C>T	0.03		Intronic
		33638182	G> A	c.409-7015 C>T	0.03		Intronic
		33638210	G> A	c.409-7043 C>T	0.03		Intronic
		33638360	C> T	c.409-7193 G>A	0.04		Intronic
		33638535	C> T	c.409-7368 G>A	0.03		Intronic
		33639014	G> A	c.408+7612 C>T	0.05		Intronic
		33641603	C> T	c.408+5023 G>A	0.03		Intronic
		33641922	C> T	c.408+4704 G>A	0.04		Intronic
		33641929	C> T	c.408+4697 G>A	0.03		Intronic
		33641956	C> T	c.408+4670 G>A	0.05		Intronic
		33641976	C> T	c.408+4650 G>A	0.03		Intronic
		33643945	G> A	c.408+2681 C>T	0.05		Intronic
		33644794	C> T	c.408+1832 G>A	0.04		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33644941	G> A	c.408+1685 C>T	0.03		Intronic
		33645037	G> A	c.408+1589 C>T	0.03		Intronic
		33645055	G> A	c.408+1571 C>T	0.04		Intronic

		33645109	C> T	c.408+1517 G>A	0.03		Intronic
		33645179	C> T	c.408+1447 G>A	0.03		Intronic
		33645185	G> A	c.408+1441 C>T	0.04		Intronic
		33645644	G> A	c.408+982 C>T	0.03		Intronic
		<i>33645661</i>	<i>GCT> G</i>	<i>c.408+963_408+964 delAG</i>	<i>1.00</i>		<i>Intronic</i>
		33645856	C> T	c.408+770 G>A	0.03		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEARI</i>
D11	DCIS	33645897	C> T	c.408+729 G>A	0.04		Intronic
		33645931	C> T	c.408+695 G>A	0.04		Intronic
		33645937	C> T	c.408+689 G>A	0.04		Intronic
		33645997	C> T	c.408+629 G>A	0.03		Intronic
		33646001	C> T	c.408+625 G>A	0.04		Intronic
		33646103	C> T	c.408+523 G>A	0.04		Intronic
		33646143	C> T	c.408+483 G>A	0.03		Intronic
		33646165	C> T	c.408+461 G>A	0.03		Intronic
		33646181	C> T	c.408+445 G>A	0.04		Intronic
		33646186	C> T	c.408+440 G>A	0.04		Intronic
		33646330	C> T	c.408+296 G>A	0.03		Intronic
		33646333	C> T	c.408+293 G>A	0.04		Intronic
		33646347	C> T	c.408+279 G>A	0.03		Intronic
		33647195	G> A	c.-162 C>T	0.03		5'UTR
		33648206	T> C	Upstream	0.14		Upstream
		33649525	C> T	Upstream	0.04		Upstream
		33649543	C> T	Upstream	0.04		Upstream
		33649577	C> T	Upstream	0.03		Upstream
		33651364	C> T	Upstream	0.04		Upstream
		33651366	G> A	Upstream	0.04		Upstream
		33653032	G> A	Upstream	0.05		Upstream
		33656997	C> T	Upstream	0.03		Upstream
		33657853	C> T	Upstream	0.03		Upstream
		33657888	C> T	Upstream	0.04		Upstream
		33657910	C> T	Upstream	0.04		Upstream
		33657918	C> T	Upstream	0.03		Upstream

	INV	33611118	G> A	c.*1660 C>T	0.03		3'UTR
		33611332	G> A	c.*1446 C>T	0.03		3'UTR
		33611892	G> A	c.*886 C>T	0.03		3'UTR
		33612122	G> A	c.*656 C>T	0.03		3'UTR
		33612384	G> A	c.*394 C>T	0.05		3'UTR
		33612468	GGTA> CCAT	c.*307 TACC>ATGG	0.06		3'UTR
		33612547	G> A	c.*231 C>T	0.03		3'UTR
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D11	INV	33612978	G> A	p.Arg410Trp c.1228C>T	0.03		Exonic
		33613058	C> T	p.Ser383Asn c.1148G>A	0.03		Exonic
		33613079	C> T	p.Gly376Asp c.1127G>A	0.04		Exonic
		33613085	C> T	p.Arg374His c.1121G>A	0.05		Exonic
		33613200	C> T	p.Val336Met c.1006G>A	0.03		Exonic
		33613362	G> A	c.878-34 C>T	0.04		Intronic
		33615571	G> A	c.878-2243 C>T	0.04		Intronic
		33615678	C> T	c.878-2350 G>A	0.03		Intronic
		33616092	C> T	c.878-2764 G>A	0.05		Intronic
		33616105	C> T	c.878-2777 G>A	0.04		Intronic
		33616197	G> A	c.878-2869 C>T	0.05		Intronic
		33616324	G> A	c.878-2996 C>T	0.05		Intronic
		33616726	G> A	c.878-3398 C>T	0.04		Intronic
		33616740	G> A	c.878-3412 C>T	0.04		Intronic
		33616897	G> A	c.878-3569 C>T	0.03		Intronic
		33617358	C> T	c.878-4030 G>A	0.04		Intronic
		33618710	G> A	c.877+5144 C>T	0.05		Intronic
		33618712	G> A	c.877+5142 C>T	0.05		Intronic
		33618734	G> A	c.877+5120 C>T	0.04		Intronic
		33619828	G> A	c.877+4026 C>T	0.05		Intronic
		33623981	G> A	c.762-12 C>T	0.04		Intronic

		33623997	C> T	c.762-28 G>A	0.03		Intronic
		33624003	G> A	c.762-34 C>T	0.04		Intronic
		33624016	G> A	c.762-47 C>T	0.03		Intronic
		33624033	C> T	c.762-64 G>A	0.05		Intronic
		33624036	G> A	c.762-67 C>T	0.03		Intronic
		33624050	G> A	c.762-81 C>T	0.03		Intronic
		33625077	C> T	c.761+212 G>A	0.04		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D11	INV	33625655	G> A	c.505-110 C>T	0.03		Intronic
		33625663	G> A	c.505-118 C>T	0.04		Intronic
		33626868	G> A	c.505-1323 C>T	0.03		Intronic
		33626885	G> A	c.505-1340 C>T	0.04		Intronic
		33626928	G> A	c.505-1383 C>T	0.04		Intronic
		33627271	G> A	c.505-1726 C>T	0.07		Intronic
		33627411	C> T	c.505-1866 G>A	0.03		Intronic
		33627448	C> T	c.505-1903 G>A	0.03		Intronic
		33628224	G> A	c.142-2679 C>T	0.03		Intronic
		33628228	G> A	c.505-2683 C>T	0.03		Intronic
		33628249	G> A	c.505-2704 C>T	0.03		Intronic
		33628254	G> A	c.505-2709 C>T	0.03		Intronic
		33628264	G> A	c.505-2719 C>T	0.07		Intronic
		33629174	C> T	c.504+1898 G>A	0.04		Intronic
		33629225	C> T	c.504+1847 G>A	0.05		Intronic
		33629389	C> T	c.504+1683 G>A	0.03		Intronic
		33629399	G> A	c.504+1673 C>T	0.04		Intronic
		33629408	G> A	c.504+1664 C>T	0.03		Intronic
		33629429	C> T	c.504+1643 G>A	0.03		Intronic
		33629544	G> A	c.504+1528 C>T	0.03		Intronic
		33629741	G> A	c.504+1331 C>T	0.04		Intronic
		33629784	G> A	c.504+1288 C>T	0.03		Intronic
		33629829	G> A	c.504+1243 C>T	0.04		Intronic
		33629870	C> T	c.504+1202 G>A	0.03		Intronic
		33630756	C> T	c.504+316 G>A	0.03		Intronic

		33630782	C> T	c.504+290 G>A	0.04		Intronic
		33630846	C> T	c.504+226 G>A	0.03		Intronic
		33630900	C> T	c.504+172 G>A	0.04		Intronic
		33630928	C> T	c.504+144 G>A	0.09		Intronic
		33630929	C> T	c.504+143 G>A	0.06		Intronic
		33631385	C> T	c.409-218 G>A	0.04		Intronic
		33633128	C> T	c.409-1961 G>A	0.04		Intronic
		33633231	C> T	c.409-2064 G>A	0.05		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D11	INV	33633241	C> T	c.409-2074 G>A	0.04		Intronic
		33634280	C> T	c.409-3113 G>A	0.03		Intronic
		33634288	G> A	c.409-3121 C>T	0.03		Intronic
		33634320	G> A	c.409-3153 C>T	0.04		Intronic
		33634322	G> A	c.409-3155 C>T	0.06		Intronic
		33634343	G> A	c.409-3176 C>T	0.04		Intronic
		33634391	G> A	c.409-3224 C>T	0.04		Intronic
		33634592	C> T	c.409-3425 G>A	0.03		Intronic
		33634609	C> T	c.409-3442 G>A	0.04		Intronic
		33634699	G> A	c.409-3532 C>T	0.03		Intronic
		33634705	G> A	c.409-3538 C>T	0.05		Intronic
		33636128	G> A	c.409-4961 C>T	0.04		Intronic
		33636180	G> A	c.409-5013 C>T	0.03		Intronic
		33636640	G> A	c.409-5473 C>T	0.03		Intronic
		33636654	G> A	c.409-5487 C>T	0.04		Intronic
		33636661	C> T	c.409-5494 G>A	0.03		Intronic
		33637967	G> A	c.409-6800 C>T	0.03		Intronic
		33638000	G> A	c.409-6833 C>T	0.06		Intronic
		33638005	G> A	c.409-6838 C>T	0.08		Intronic
		33638065	G> A	c.409-6898 C>T	0.04		Intronic
		33638081	G> A	c.409-6914 C>T	0.07		Intronic
		33638180	G> A	c.409-7013 C>T	0.03		Intronic
		33638199	G> A	c.409-7032 C>T	0.05		Intronic
		33638200	G> A	c.409-7033 C>T	0.03		Intronic

		33638349	C> T	c.409-7182 G>A	0.04		Intronic
		33638384	C> T	c.409-7217 G>A	0.03		Intronic
		33638462	G> A	c.409-7295 C>T	0.04		Intronic
		33638480	C> T	c.409-7313 G>A	0.04		Intronic
		33638575	C> T	c.409-7408 G>A	0.04		Intronic
		33638591	C> T	c.409-7424 G>A	0.05		Intronic
		33638623	C> T	c.409-7456 G>A	0.03		Intronic
		33638637	C> T	c.409-7470 G>A	0.03		Intronic
		33638753	C> T	c.409-7586 G>A	0.04		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D11	INV	33638762	C> T	c.409-7595 G>A	0.04		Intronic
		33638803	C> T	c.409-7636 G>A	0.03		Intronic
		33638808	C> T	c.409-7641 G>A	0.06		Intronic
		33638828	C> T	c.409-7661 G>A	0.04		Intronic
		33638848	C> T	c.409-7681 G>A	0.03		Intronic
		33638864	C> T	c.409-7697 G>A	0.03		Intronic
		33638872	C> T	c.409-7705 G>A	0.03		Intronic
		33638884	C> T	c.409-7717 G>A	0.07		Intronic
		33638889	C> T	c.409-7722 G>A	0.03		Intronic
		33639002	G> A	c.408+7624 C>T	0.03		Intronic
		33639010	G> A	c.408+7616 C>T	0.06		Intronic
		33639044	G> A	c.408+7582 C>T	0.03		Intronic
		33639046	G> A	c.408+7580 C>T	0.04		Intronic
		33639056	G> A	c.408+7570 C>T	0.03		Intronic
		33640222	C> T	c.408+6404 G>A	0.04		Intronic
		33640229	C> T	c.408+6397 G>A	0.04		Intronic
		33640366	C> T	c.408+6260 G>A	0.03		Intronic
		33641447	C> T	c.408+5179 G>A	0.03		Intronic
		33641481	C> T	c.408+5145 G>A	0.04		Intronic
		33641528	C> T	c.408+5098 G>A	0.05		Intronic
		33641543	G> A	c.408+5083 C>T	0.04		Intronic
		33641565	G> A	c.408+5061 C>T	0.04		Intronic
		33641599	C> T	c.408+5027 G>A	0.04		Intronic

		33641650	G> A	c.408+4976 C>T	0.04		Intronic
		33641707	C> T	c.408+4919 G>A	0.04		Intronic
		33641767	C> T	c.408+4859 G>A	0.03		Intronic
		33641874	C> T	c.408+4752 G>A	0.07		Intronic
		33641923	C> T	c.408+4703 G>A	0.04		Intronic
		33641959	C> T	c.408+4667 G>A	0.06		Intronic
		33641962	C> T	c.408+4664 G>A	0.05		Intronic
		33643862	G> A	c.408+2764 C>T	0.03		Intronic
		33643874	G> A	c.408+2752 C>T	0.04		Intronic
		33643882	G> A	c.408+2744 C>T	0.03		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D11	INV	33643887	G> A	c.408+2739 C>T	0.03		Intronic
		33644006	G> A	c.408+2620 C>T	0.03		Intronic
		33644084	G> A	c.408+2542 C>T	0.03		Intronic
		33644433	G> A	c.408+2193 C>T	0.04		Intronic
		33644439	C> T	c.408+2187 G>A	0.03		Intronic
		33644442	G> A	c.408+2184 C>T	0.05		Intronic
		33644467	G> A	c.408+2159 C>T	0.06		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33644908	G> A	c.408+1718 C>T	0.04		Intronic
		33644917	G> A	c.408+1709 C>T	0.04		Intronic
		33645017	C> T	c.408+1609 G>A	0.06		Intronic
		33645020	G> A	c.408+1606 C>T	0.05		Intronic
		33645036	G> A	c.408+1590 C>T	0.03		Intronic
		33645110	C> T	c.408+1516 G>A	0.03		Intronic
		33645161	C> T	c.408+1465 G>A	0.05		Intronic
		33645176	C> T	c.408+1450 G>A	0.03		Intronic
		33645203	G> A	c.408+1423 C>T	0.04		Intronic
		33645207	G> A	c.408+1419 C>T	0.04		Intronic
		33645240	G> A	c.408+1386 C>T	0.03		Intronic
		33645285	G> A	c.408+1341 C>T	0.09		Intronic
		33645318	G> A	c.408+1308 C>T	0.03		Intronic
		33645597	G> A	c.408+1029 C>T	0.05		Intronic

		33645661	GCT> G	c.408+963_408+964 delAG	1.00		Intronic
		33645771	C> T	c.408+855 G>A	0.03		Intronic
		33645897	C> T	c.408+729 G>A	0.03		Intronic
		33645934	C> T	c.408+692 G>A	0.03		Intronic
		33646036	T> C	c.408+590 A>G	0.06		Intronic
		33646277	C> T	c.408+349 G>A	0.03		Intronic
		33646304	C> T	c.408+322 G>A	0.04		Intronic
		33646644	G> A	p.Asp130Asp c.390C>T	0.16		Exonic
		33646650	G> A	p.Ile128Ile c.384C>T	0.03		Exonic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEARI</i>
D11	INV	33647177	G> A	c.-144 C>T	0.03		5'UTR
		33647232	G> A	c.-199 C>T	0.03		5'UTR
		33647319	G> A	Upstream	0.04		Upstream
		33647348	G> A	Upstream	0.05		Upstream
		33647350	G> A	Upstream	0.06		Upstream
		33647588	C> T	Upstream	0.04		Upstream
		33648237	A> G	Upstream	0.03		Upstream
		33648765	G> A	Upstream	0.03		Upstream
		33649376	G> A	Upstream	0.03		Upstream
		33649414	G> A	Upstream	0.03		Upstream
		33649463	G> A	Upstream	0.04		Upstream
		33649470	C> T	Upstream	0.04		Upstream
		33649489	G> A	Upstream	0.03		Upstream
		33649491	C> T	Upstream	0.04		Upstream
		33649492	C> T	Upstream	0.07		Upstream
		33649495	C> T	Upstream	0.04		Upstream
		33649529	G> A	Upstream	0.03		Upstream
		33649588	G> A	Upstream	0.04		Upstream
		33651192	C> T	Upstream	0.05		Upstream
		33651312	C> T	Upstream	0.03		Upstream
		33651315	G> A	Upstream	0.05		Upstream
		33651406	G> A	Upstream	0.05		Upstream

		33651475	C> T	Upstream	0.03		Upstream
		33653001	G> A	Upstream	0.05		Upstream
		33653032	G> A	Upstream	0.05		Upstream
		33653057	G> A	Upstream	0.05		Upstream
		33653093	G> A	Upstream	0.04		Upstream
		33657851	C> T	Upstream	0.04		Upstream
		33657912	C> T	Upstream	0.03		Upstream
		33657981	C> T	Upstream	0.04		Upstream
		33657990	C> T	Upstream	0.03		Upstream
		33658118	G> A	Upstream	0.03		Upstream
D12	DCIS	33612422	C> T	c.*356 G>A	0.04		3'UTR
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEARI</i>
D12	DCIS	33612446	G> A	c.*332 C>T	0.04		3'UTR
		33628228	G> A	c.505-2683 C>T	0.03		Intronic
		33629305	G> A	c.504+1767 C>T	0.02		Intronic
		33629323	G> A	c.504+1749 C>T	0.04		Intronic
		33629668	<i>TC> CA</i>	<i>c.504+1404 GA>TG</i>	<i>0.50</i>		<i>Intronic</i>
		33630843	C> T	c.504+229 G>A	0.04		Intronic
		33638065	G> A	c.409-6898 C>T	0.03		Intronic
		33638085	G> A	c.409-6918 C>T	0.04		Intronic
		33638086	G> A	c.409-6919 C>T	0.05		Intronic
		33644798	<i>AA> GC</i>	<i>c.408+1828 TT>GC</i>	<i>1.00</i>		<i>Intronic</i>
		33645661	<i>GCT> G</i>	<i>c.408+963_408+964 delAG</i>	<i>0.50</i>		<i>Intronic</i>
		33645855	C> T	c.408+771 G>A	0.03		Intronic
		33646979	G> A	p.Gln19* c.55C>T	0.04		Exonic
		33647350	G> A	Upstream	0.04		Upstream
		33648250	G> A	Upstream	0.80		Upstream
		33648859	T> C	Upstream	0.06		Upstream
		33658110	G> A	Upstream	0.06		Upstream
		33612448	G> A	c.*330 C>T	0.04		3'UTR
		33612452	G> A	c.*326 C>T	0.03		3'UTR
		33612948	G> A	p.Leu420Leu c.1258C>T	0.03		Exonic

	INV	33616704	G> A	c.878-3376 C>T	0.03		Intronic
		33616716	G> A	c.878-3376 C>T	0.03		Intronic
		33616791	G> A	c.878-3463 C>T	0.04		Intronic
		33616897	G> A	c.878-3569 C>T	0.08		Intronic
		33625426	C> T	p.Leu208Leu c.624G>A	0.05		Exonic
		33627317	C> T	c.505-1772 G>A	0.04		Intronic
		33629293	G> A	c.504+1779 C>T	0.03		Intronic
		33629364	G> A	c.504+1708 C>T	0.04		Intronic
		33629486	G> A	c.504+1586 C>T	0.03		Intronic
		33629668	TC> CA	c.504+1404 GA>TG	0.5		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D12	INV	33630854	C> T	c.504+218 G>A	0.05		Intronic
		33636289	G> A	c.409-5122 C>T	0.03		Intronic
		33638552	C> T	c.409-7385 G>A	0.03		Intronic
		33638564	C> T	c.409-7397 G>A	0.06		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	0.5		Intronic
		33646088	C> T	c.408+538 G>A	0.04		Intronic
		33646330	C> T	c.408+296 G>A	0.03		Intronic
		33648250	G> A	Upstream	0.06		Upstream
		33648859	T> C	Upstream	0.08		Upstream
		33658110	G> A	Upstream	0.06		Upstream
D13	DCIS	33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.409_410 delGA	1.00		Intronic
		33647350	G> A	Upstream	0.04		Upstream
		33647595	C> T	Upstream	0.03		Upstream
		33648237	A> G	Upstream	0.03		Upstream
	INV	33612099	CCCT>GGGA	c.*676 AGGG>TCCC	0.68		3'UTR
		33612473	T> C	c.*305 A>G	0.64		3'UTR
		33616851	G> A	c.878-3523 C>T	0.03		Intronic
		33618447	G> A	c.878-5119 C>T	0.03		Intronic
		33618956	C> T	c.877+4898 G>A	0.04		Intronic

		33628201	G> A	c.505-2656 C>T	0.04		Intronic
		33630834	C> T	c.504+238 G>A	0.03		Intronic
		33630919	C> T	c.504+153 G>A	0.04		Intronic
		33638087	G> A	c.409-6920 C>T	0.04		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.409_410 delAG	1.00		Intronic
		33645977	T> C	c.408+649 A>G	0.05		Intronic
		33646036	T> C	c.408+590 A>G	0.10		Intronic
		33646374	G> A	c.408+252 C>T	0.04		Intronic
		33646644	G> A	p.Asp130Asp c.390C>T	0.06		Exonic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEARI</i>
D13	INV	33648206	T> C	Upstream	0.29		Upstream
		33648210	T> C	Upstream	0.02		Upstream
		33648234	T> A	Upstream	0.03		Upstream
		33658110	G> A	Upstream	0.03		Upstream
D14	DCIS	33611109	G> A	c.*1669 C>T	0.06		3'UTR
		33612172	T> A	c.*606 A>T	0.04		3'UTR
		33613114	C> T	p.Val364Val c.1092G>A	0.03		Exonic
		33616864	G> A	c.878-3536 C>T	0.05		Intronic
		33618915	C> T	c.877+4939 G>A	0.04		Intronic
		33623992	G> A	c.762-23 C>T	0.05		Intronic
		33625665	G> A	c.505-120 C>T	0.02		Intronic
		33627255	C> T	c.505-1710 G>A	0.03		Intronic
		33627263	G> A	c.505-1718 C>T	0.03		Intronic
		33627382	C> T	c.505-1837 G>A	0.04		Intronic
		33629315	G> A	c.504+1757 C>T	0.05		Intronic
		33629316	G> A	c.504+1756 C>T	0.04		Intronic
		33629860	G> A	c.504+1212 C>T	0.03		Intronic
		33629867	G> A	c.504+1205 C>T	0.04		Intronic
		33630777	C> T	c.504+295 G>C	0.08		Intronic
		33638081	G> A	c.409-6914 C>T	0.04		Intronic
		33638087	G> A	c.409-6920 C>T	0.04		Intronic

		33638177	G> A	c.409-7010 C>T	0.03		Intronic
		33638270	C> T	c.409-7103 G>A	0.04		Intronic
		33638793	C> T	c.409-7626 G>A	0.03		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	1.00		Intronic
		33648237	A> G	Upstream	0.03		Upstream
		33648760	G> A	Upstream	0.04		Upstream
		33653022	G> A	Upstream	0.04		Upstream
		33657846	C> T	Upstream	0.04		Upstream
		33658110	G>A	Upstream	0.03		Upstream
		33681433	C> T	Upstream	0.03		Upstream
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEARI</i>
D14	INV	33611113	G> A	c.*1665 C>T	0.04		3'UTR
		33612422	C> T	c.*356 G>A	0.06		3'UTR
		33629305	G> A	c.504+1767 C>T	0.02		Intronic
		33638005	G> A	c.409-6838 C>T	0.04		Intronic
		33638085	G> A	c.409-6918 C>T	0.05		Intronic
		33638086	G> A	c.409-6919 C>T	0.04		Intronic
		33638753	C> T	c.409-7586 G>A	0.04		Intronic
		33638879	C> T	c.409-7712 G>A	0.04		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	1.00		Intronic
		33648234	T> A	Upstream	0.03		Upstream
D15	DCIS	33612419	G> A	c.*359 C>T	0.04		3'UTR
		33616155	G> A	c.878-2827 C>T	0.04		Intronic
		33616467	G> A	c.878-3139 C>T	0.04		Intronic
		33616858	G> A	c.878-3530 C>T	0.04		Intronic
		33627372	C> T	c.505-1827 G>A	0.53		Intronic
		33629316	G> A	c.504+1756 C>T	0.06		Intronic
		33630847	C> T	c.504+225 G>A	0.05		Intronic
		33643872	G> A	c.408+2754 C>T	0.03		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	1.00		Intronic

		33646036	T> A	c.408+590 A>G	0.04		Intronic
		33648206	T> A	Upstream	0.09		Upstream
	INV	33616482	G> A	c.878-3154 C>T	0.04		Intronic
		33627236	G> A	c.505-1691 C>T	0.09		Intronic
		33627372	C> T	c.505-1827 G>A	0.40		Intronic
		33629305	G> A	c.504+1767 C>T	0.02		Intronic
		33630714	G> A	c.504+358 C>T	0.03		Intronic
		33630924	C> T	c.504+148 G>A	0.05		Intronic
		33638081	G> A	c.409-6914 C>T	0.04		Intronic
		33638082	G> A	c.409-6915 C>T	0.03		Intronic
		33638083	G> A	c.409-6916 C>T	0.04		Intronic
		33638352	G> A	c.409-7185 C>T	0.04		Intronic
<i>Sample</i>	<i>Lesion Type</i>	<i>Chromosome 1 Hg19 location</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Validated</i>	<i>Area within DEAR1</i>
D15	INV	33638568	C> T	c.409-7401 G>A	0.03		Intronic
		33638868	C> T	c.409-7701 G>A	0.03		Intronic
		33639129	T> A	c.408+7497 A>T	0.04		Intronic
		33641960	C> T	c.45+49 G>A	0.03		Intronic
		33642121	G> A	c.408+4505 C>T	0.02		Intronic
		33644798	AA> GC	c.408+1828 TT>GC	1.00		Intronic
		33645661	GCT> G	c.408+963_408+964 delAG	1.00		Intronic
		33645977	T> A	c.408+649A>G	0.03		Intronic
		33647354	G> A	Upstream	0.03		Upstream
		33648250	G> A	Upstream	0.05		Upstream
		33658110	G> A	Upstream	0.04		Upstream
D16	DCIS	33639129	T> A	c.408+7497 A>T	0.03		Intronic
		33648859	T> C	Upstream	0.05		Upstream
	INV	33612040	C> G	c.*738 G>C	0.03		3'UTR
		33639129	T> A	c.408+7497 A>T	0.04		Intronic
		33656988	T> C	Upstream	0.02		Upstream
D17	DCIS	33639129	T> A	c.408+7497 A>T	0.03		Intronic
		33648237	A> G	Upstream	0.03		Upstream
	INV	33648859	T> C	Upstream	0.04		Upstream

Appendix XI- Detailed List of *DEARI* Variants Shared Between the *In Situ* and Invasive Components of Ductal Carcinoma *In Situ*. This table describes the variants that shared between the *in situ* and invasive components of adjacent DCIS lesions. The vast majority of these variants were within *DEARI*'s introns. Shared variants with stable variant frequencies at or around 50% or 100% within both lesion components are contained in samples for which the normal breast tissue was unable to be sequenced, and most likely reflect germline variants.

<i>Sample</i>	<i>Chromosome1 Hg19 location</i>	<i>Lesion Type</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>% Variant Frequency</i>	<i>Area within DEARI</i>
D01	33639129	DCIS	T> A	c.408+7497A>T	0.03	Intronic
		INV			0.07	
D02	33639129	DCIS	T >A	c.408+7497 A>T	0.04	Intronic
		INV			0.04	
D04	33630779	DCIS	G >A	c.504+293 C>T	0.03	Intronic
		INV			0.03	
	33648250	DCIS	G> A	Upstream	0.10	Upstream
		INV			0.05	
	33658110	DCIS	G> A	Upstream	0.05	Upstream
		INV			0.04	
D05	33629313	DCIS	G> A	c.504+1759 C>T	0.04	Intronic
		INV			0.05	
	33630850	DCIS	C> T	c.504+222 G>A	0.04	Intronic
		INV			0.04	
	33638864	DCIS	C> T	c.409-7697 G>A	0.06	Intronic

<i>Sample</i>	<i>Chromosome1 Hg19 location</i>	<i>Lesion Type</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>% Variant Frequency</i>	<i>Area within DEARI</i>
		INV			0.04	
D05	33645855	DCIS	C> T	c.408+771 G>A	0.02	Intronic
		INV			0.04	
	33646036	DCIS	T> C	c.408+590 A>G	0.09	Intronic
		INV			0.06	
	33648206	DCIS	T> C	Upstream	0.03	Upstream
		INV			0.15	
	33648250	DCIS	G> A	Upstream	0.09	Upstream
		INV			0.09	
	33639129	DCIS	T> A	Upstream	0.05	Upstream
		INV			0.05	
D09	33629668	DCIS	TC> CA	c.504+1404 GA>TG	0.46	Intronic
		INV			0.49	
	33644798	DCIS	AA> GC	c.408+1828 TT>GC	1.00	Intronic
		INV			1.00	
	33645661	DCIS	GCT> G	c.408+963_408+964 delAG	0.51	Intronic
		INV			0.52	
D11	33644798	DCIS	AA> GC	c.408+1828 TT>GC	1.00	Intronic

		INV			1.00	
	33645661	DCIS	GCT> G	c.408+963_408+964 delAG	1.00	Intronic
<i>Sample</i>	<i>Chromosome1 Hg19 location</i>	<i>Lesion Type</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>% Variant Frequency</i>	<i>Area within DEARI</i>
D11	33645661	INV			1.00	
	33645897	DCIS	C> T	c.408+729 G>A	0.04	Intronic
		INV			0.03	
	33653032	DCIS	G> A	Upstream	0.05	Upstream
		INV			0.05	
D12	33629668	DCIS	TC> CA	c.504+1404 GA>TG	0.50	Intronic
		INV			0.50	
	33644798	DCIS	AA> GC	c.408+1828 TT>GC	1.00	Intronic
		INV			1.00	
	33645661	DCIS	GCT> G	c.408+963_408+964 delAG	0.50	Intronic
		INV			0.50	
	33648250	DCIS	G> A	Upstream	0.80	Upstream
		INV			0.06	
	33648859	DCIS	T> C	Upstream	0.06	Upstream
		INV			0.08	
	33658110	DCIS	G> A	Upstream	0.06	Upstream

		INV			0.06	
D13	33644798	DCIS	AA> GC	c.408+1828 TT>GC	1.00	Intronic
		INV			1.00	
<i>Sample</i>	<i>Chromosome1 Hg19 location</i>	<i>Lesion Type</i>	<i>Nucleotide Change</i>	<i>Variant</i>	<i>Variant Frequency</i>	<i>Area within DEARI</i>
D13	33645661	DCIS	GCT> G	c.409_410 delGA	1.00	Intronic
		INV			1.00	
D14	33644798	DCIS	AA> GC	c.408+1828 TT>GC	1.00	Intronic
		INV			1.00	
	33645661	DCIS	GCT> G	c.408+963_408+964 delAG	1.00	Intronic
		INV			1.00	
D15	33627372	DCIS	C> T	c.505-1827 G>A	0.53	Intronic
		INV			0.40	
	33644798	DCIS	AA> GC	c.408+1828 TT>GC	1.00	Intronic
		INV			1.00	
	33645661	DCIS	GCT> G	c.408+963_408+964 delAG	1.00	Intronic
		INV			1.00	
D16	33639129	DCIS	T> A	c.408+7497 A>T	0.03	Intronic
		INV			0.04	

Bibliography

Bert Vogelstein, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz Jr. and Kenneth W. Kinzler (2013). "Cancer Genome Landscapes " Science **339**: 1546-1558.

N Chen, S Balasenthil, J Reuther, A Frayna, Y Wang, D S Chandler, L V Abruzzo, A Rashid, J Rodriguez, G Lozano, Y Cao, E Lokken, J Chen, M L Frazier, A A Sahin, I I Wistuba, S Sen, S T Lott and A M Killary (2013). "DEAR1 is a chromosome 1p35 tumor suppressor and master regulator of TGF- β -driven epithelial-mesenchymal transition." Cancer Discov. **3**(10): 1172-1189.

Payton Rous (1911). "A sarcoma of the fowl transmissible by an agent separable from the tumor cells." J Exp Med **13**(4): 397-411.

Susan Reich Weiss, Harold E. Varmus and J. Michael Bishop (1977). "The Size and Genetic Composition of Virus-Specific RNAs in the Cytoplasm of Cells Producing Avian Sarcoma-Leukosis Virus." Cell **12**: 983-992.

Dominique Stehelin, Ramareddy V. Guntaka, Harold E. Varmus and J Michael Bishop (1976). "Purification of DNA complementary to nucleotide sequences required for neoplastic transformation of fibroblasts by avian sarcoma viruses." Journal of Molecular Biology **101**(3): 349-365.

T G Krontiris and G M Cooper (1979). "Transforming activity of human tumor DNAs." Proc Natl Acad Sci U S A. **76**(11): 5714-5718.

C Shih, B Z Shilo, M P Goldfarb, A Dannenberg and R A Weinberg (1979). "Passage of phenotypes of chemically transformed cells via transfection of DNA and chromatin." Proc Natl Acad Sci U S A. **76**(11): 57-14-5718.

Robert J. Huebner and George J. Todaro (1969). "Oncogenes of RNA tumor viruses as determinants of cancer." Proc Natl Acad Sci U S A. **64**(3): 1087-1094.

Harvey Lodish, Arnold Berk, S Lawrence Zipursky, Paul Matsudaira, David Baltimore and James Darnell (2000). Proto-Oncogenes and Tumor-Suppressor Genes. New York, W. H. Freeman.

Ruth Sager (1985). "Genetic Suppression of Tumor Formation." Advances in Cancer Research **44**: 43-68.

H. P. Klinger, A. S. Baim, C. K. Eun, T. B. Shows and F. H. Ruddle (1978). "Human chromosomes which affect tumorigenicity in hybrids of diploid human with heteroploid human or rodent cells." Cytogenetic and Genome Research **22**: 245-249.

Eric J. Stanbridge, Robert R. Flandermeyer, David W. Daniels and Walter A. Nelson-Rees (1981). "Specific chromosome loss associated with the expression of tumorigenicity in human cell hybrids." Somatic Cell Genetics **7**(6): 699-712.

William F. Benedict, Bernard E. Weissman, Corey Mark and Eric J. Stanbridge (1984). "Tumorigenicity of human HT1080 fibrosarcoma × normal fibroblast hybrids: chromosome dosage dependency." Cancer Res **44**(8): 3471-3479.

Eric J Stanbridge (1990). "Human Tumor Suppressor Genes." Annual Review of Genetics **24**: 615-657.

Scott A. Bader, Clare Fasching, Garrett M. Brodeur and Eric J. Stanbridge (1991). "Dissociation of suppression of tumorigenicity and differentiation in vitro effected by transfer of single human chromosomes into human neuroblastoma cells." Cell Growth and Differentiation: the molecular biology journal of the American Association for Cancer Research **2**(5): 245-255.

M. Kaelbling and H. P. Klinger (1986). "Suppression of tumorigenicity in somatic cell hybrids. IV: chromosomes of normal human cells associated with suppression of tumorigenicity in hybrids with D98AH2 carcinoma cells." Cytogenetics and cell genetics **42**(4): 225-235.

F. Wiener, G. Klein and H. Harris (1974). "The Analysis of Malignancy by Cell Fusion V. Further Evidence of the Ability of Normal Diploid Cells to Suppress Malignancy." Journal of cell science **15**(1): 177-183.

WH Lee, R Bookstein, F Hong, LJ Young, JY Shew and EY Lee (1987). "Human retinoblastoma susceptibility gene: cloning, identification, and sequence." Science **235**(4794): 1394-1399.

H. J. Huang, J. K. Yee, J. Y Shew, P.-L. Chen, R. Bookstein, T. Friedmann, E. Y.-P. Lee and W.-H. Lee (1988). "Suppression of the neoplastic phenotype by replacement of the RB gene in human cancer cells." Science **242**: 1563-1566.

Daniel J. Riley, Eva Y.-H.P. Lee and Wen-Hwa Lee (1994). "THE RETINOBLASTOMA PROTEIN: More Than a Tumor Suppressor " Annual Review of Cell Biology **10**: 1-29.

Douglas Hanahan and R A Weinberg (2000). "The hallmarks of cancer." Cell **100**(1): 57-70.

G. M. Cooper (2000). Tumor Suppressor Genes. Sunderland, MA, Sinauer Associates.

Sundaresan Venkatachalam, Stuart D. Tyner, Curtis R. Pickering, Scott Boley, Leslie Recio, John E. French and Lawrence A. Donehower (2001). "Is p53 Haploinsufficient for Tumor Suppression? Implications for the p53^{+/-} Mouse Model in Carcinogenicity Testing." Toxicology Pathology **29 (Suppl.)**: 147-154.

Thierry Soussi (2010). "The history of p53: A perfect example of the drawbacks of scientific paradigms." EMBO Reports **11**(11): 822-826.

Junne Kamihara, Huma Q. Rana and Judy E. Garber (2014). "Germline TP53 Mutations and the Changing Landscape of Li-Fraumeni Syndrome." Human Mutation **35**(6): 654-662.

Franziska Michor, Yoh Iwasa and Martin A. Nowak (2004). "Dynamics of Cancer Progression." Nature Reviews Cancer **4**: 197-205.

Kenneth Kinzler and Bert Vogelstein (1997). "Gatekeepers and caretakers." Nature **386**: 761-763.

Christopher D. McFarland, Kirill S. Korolev, Gregory V. Kryukov, Shamil R. Sunyaev and Leonid A. Mirny (2012). "Impact of deleterious passenger mutations on cancer progression." Proc Natl Acad Sci U S A. **110**(8): 2910-2915.

Andriy Marusyk, Vanessa Almendro and Kornelia Polyak (2012). "Intra-tumour heterogeneity: a looking glass for cancer?" Nature Reviews Cancer **12**: 323-334.

Jr. Alfred G. Knudson (1971). "Mutation and Cancer: Statistical Study of Retinoblastoma." Proc Natl Acad Sci U S A. **68**(4): 820-823.

Marra MA Pon JR (2015). "Driver and passenger mutations in cancer. ." Annu Rev Pathol. **10**: 25-50.

J T Dong (2001). "Chromosomal deletions and tumor suppressor genes in prostate cancer." Cancer and Metastasis Reviews **20**(3-4): 173-193.

Donna G Albertson, Colin Collins, Frank McCormick and Joe W Gray (2003). "Chromosome aberration in solid tumors " Nature Genetics **34**: 369-376.

Curt Stern (1943). "Genic Action as Studied by Means of the Effects of Different Doses and Combinations of Alleles." Genetics **28**(6): 441-475.

Wendy D Cook and Benjamin J McCaw (2000). "Accommodating haploinsufficient tumour suppressor genes in Knudson's model." Oncogene **19**(30): 3434-3438.

Shannon R. Payne and Christopher J. Kemp (2005). "Tumor Suppressor Genetics." Carcinogenesis **26**(12): 2031-2045.

Bernard Kwabi-Addo, Karen Schmidt Dipak Giri, Ramon Parsons Katrina Podsypanina, Norman Greenberg, and Michael Ittmann (2001). "Haploinsufficiency of the Pten tumor suppressor gene promotes prostate cancer progression." Proc Natl Acad Sci U S A. **98**(20): 11563-11568.

Shana N. Wingo, Teresa D. Gallardo, Esra A. Akbay, Mei-Chi Liang, Cristina M. Contreras, Todd Boren, Takeshi Shimamura, David S. Miller, Norman E. Sharpless, Nabeel Bardeesy, David J. Kwiatkowski, John O. Schorge, Kwok-Kin Wong and Diego H. Castrillon (2009). "Somatic LKB1 Mutations Promote Cervical Cancer Progression." PLoS One **4**(4): e5137.

Tapio Visakorpi, Eija Hyytinen, Pasi Koivisto, Minna Tanner, Riitta Keinänen, Christian Palmberg, Aarno Palotie, Teuvo Tammela, Jorma Isola and Olli-P. Kallioniemi (1995). "In vivo amplification of the androgen receptor gene and progression of human prostate cancer." Nature Genetics **9**: 401-406.

Mika Tirkkonen, Oskar Johannsson, Bjarni A. Agnarsson, Hakan Olsson, Sigurdur Ingvarsson, Ritva Karhu, Minna Tanner, Jorma Isola, Rosa B. Barkardottir, Ake Borg and Olli-P. Kaffioniemi (1997). "Distinct Somatic Genetic Changes Associated with Tumor Progression in Carriers of BRCA1 and BRCA2 Germ-line Mutations." Cancer Res **57**: 1222-1227.

Laufey Tryggvadóttir, Linda Vidarsdóttir, Tryggvi Thorgeirsson, Jon Gunnlaugur Jonasson, Elinborg Jona Ólafsdóttir, Gudridur Helga Ólafsdóttir, Thorunn Rafnar, Steinunn Thorlacius, Eiríkur Jonsson, Jorunn Erla Eyfjörð and Hrafn Tulinius (2007). "Prostate Cancer Progression and Survival in BRCA2 Mutation Carriers." Journal National Cancer Institute **99**(12): 929-935.

Lavanya H Palavalli, Todd D Prickett, John R Wunderlich, Xiaomu Wei, Allison S Burrell, Patricia Porter-Gill, Sean Davis, Chenwei Wang, Julia C Cronin, Neena S Agrawal, Jimmy C Lin, Wendy Westbroek, Shelley Hoogstraten-Miller, Alfredo A Molinolo, Patricia Fetsch, Armando C Filie, Michael P O'Connell, Carolyn E Banister, Jason D Howard, Phillip Buckhaults, Ashani T Weeraratna, Lawrence C Brody, Steven A Rosenberg and Yardena Samuels (2009). "Analysis of the matrix metalloproteinase family reveals that MMP8 is often mutated in melanoma." Nature Genetics **41**: 518-520.

Rameen Beroukhi, Craig H. Mermel, Dale Porter, Guo Wei, Soumya Raychaudhuri, Jerry Donovan, Jordi Barretina, Jesse S. Boehm, Jennifer Dobson, Mitsuyoshi Urashima, Kevin T. Mc Henry, Reid M. Pinchback, Azra H. Ligon, Yoon-Jae Cho, Leila Haery, Heidi Greulich, Michael Reich, Wendy Winckler, Michael S. Lawrence, Barbara A. Weir, Kumiko E. Tanaka, Derek Y. Chiang, Adam J. Bass, Alice Loo⁸, Carter Hoffman, John Prensner, Ted Liefeld, Qing Gao¹, Derek Yecies³, Sabina Signoretti, Elizabeth Maher, Frederic J. Kaye, Hidefumi Sasaki, Joel E. Tepper, Jonathan A. Fletcher, Josep Taberero, Jose´ Baselga, Ming-Sound Tsao, Francesca Demichelis, Mark A. Rubin, Pasi A. Janne, Mark J. Daly, Carmelo Nucera, Ross L. Levine, Benjamin L. Ebert, Stacey Gabriel, Anil K. Rustgi, Cristina R. Antonescu, Marc Ladanyi, Anthony Letai, Levi A. Garraway, Massimo Loda, David G. Beer, Lawrence D. True, Aikou Okamoto, Scott L. Pomeroy, Samuel Singer, Todd R. Golub, Eric S. Lander, Gad Getz, William R. Sellers and Matthew Meyerson (2010). "The landscape of somatic copy-number alteration across human cancers." Nature **463**: 899-905.

Marco A. Pierotti, Gabriella Sozzi and Carlo M. Croce (2003). Discovery and identification of oncogenes. Holland-Frei Cancer Medicine. Donald W Kufe, Raphael E Pollock, Ralph R Weichselbaum et al. Ontario B C Decker.

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

Nida Iqbal and Naveed Iqbal (2014). "Imatinib: A Breakthrough of Targeted Therapy in Cancer." Chemotherapy Research and Practice **2014**.

Wen Xuea, Thomas Kitzinga, Stephanie Roessler, Johannes Zuber, Alexander Krasnitz, Nikolaus Schultz, Kate Revill, Susann Weissmuellerd, Amy R. Rappaport, Janelle Simon, Jack Zhang, Weijun Luo, James Hicks, Lars Zendera, Xin Wei Wang, Scott Powers, Michael Wigler and Scott W. Lowe (2012). "A cluster of cooperating tumor-suppressor gene candidates in chromosomal deletions." 109 **21**(8212-8217).

Nicole L. Solimini, Qikai Xu, Craig H. Mermel, Anthony C. Liang, Michael R. Schlabach, Ji Luo, Anna E. Burrows, Anthony N. Anselmo, Andrea L. Bredemeyer, Mamie Z. Li, Rameen Beroukhim, Matthew Meyerson and Stephen J. Elledge (2012). "Recurrent Hemizygous Deletions in Cancers May Optimize Proliferative Potential." Science **337**(6090): 104-109.

Wigard P. Kloostermana, Jan Kosterb and Jan J. Molenaar (2014). "Prevalence and clinical implications of chromothripsis in cancer genomes." Current Opinion **26**(1).

K Crasta, N J Ganem, R Dagher, A B Lantermann, E V Ivanova, Y Pan, L Nezi, A Protopopov, D Chowdhury and D Pellman (2012). "DNA breaks and chromosome pulverization from errors in mitosis." Nature **482**(7383): 53-58.

Rebecca A. Burrell, Nicholas McGranahan, Jiri Bartek and Charles Swanton (2013). "The causes and consequences of genetic heterogeneity in cancer evolution." Nature **501**: 338-345.

Douglas Hanahan and Robert A. Weinberg (2011). "Hallmarks of Cancer: The Next Generation " Cell **144**(5): 646–674.

The Cancer Genome Atlas Research Network, John N Weinstein, Eric A Collisson, Gordon B Mills, Kenna R Mills Shaw, Brad A Ozenberger, Kyle Ellrott, Ilya Shmulevich, Chris Sander and Joshua M Stuart (2013). "The Cancer Genome Atlas Pan-Cancer analysis project." Nature Genetics **45**(10): 1113-1120.

Ludmil B. Alexandrov, Serena Nik-Zainal, David C. Wedge, Samuel A. J. R. Aparicio, Sam Behjati, Andrew V. Biankin, Graham R. Bignell, Niccolo Bolli, Ake Borg, Anne-Lise Børresen-Dale, Sandrine Boyault, Birgit Burkhardt, Adam P. Butler, Carlos Caldas, Helen R. Davies, Christine Desmedt, Roland Eils, Jorunn Erla Eyfjord, John A. Foekens, Mel Greaves, Fumie Hosoda, Barbara Hutter, Tomislav Ilicic, Sandrine Imbeaud, Marcin Imielinski, Natalie Jager, David T. W. Jones, David Jones, Stian Knappskog, Marcel Kool, Sunil R. Lakhani, Carlos Lo´pez-Oti´n, Sancha Martin, Nikhil C. Munshi, Hiromi Nakamura, Paul A. Northcott, Marina Pajic, Elli Papaemmanuil, Angelo Paradiso, John V. Pearson, Xose S. Puente, Keiran Raine, Manasa Ramakrishna, Andrea L. Richardson, Julia Richter, Philip Rosenstiel, Matthias Schlesner, Ton N. Schumacher, Paul N. Span, Jon W. Teague, Yasushi

Totoki, Andrew N. J. Tutt, Rafael Valdes-Mas, Marit M. van Buuren, Laura van't Veer, Anne Vincent-Salomon, Nicola Waddell, Lucy R. Yates, Australian Pancreatic Cancer Genome Initiative, ICGC Breast Cancer Consortium, ICGC MMML-Seq Consortium, ICGC PedBrain, Jessica Zucman-Rossi, P. Andrew Futreal, Ultan McDermott, Peter Lichter, Matthew Meyerson, Sean M. Grimmond, Reiner Siebert, Elias Campo, Tatsuhiro Shibata, Stefan M. Pfister, Peter J. Campbell and Michael R. Stratton (2013).

"Signatures of mutational processes in human cancer." Nature **500**: 415-421.

Michael D. McLellan Cyriac Kandoth, Fabio Vandin, Kai Ye, Beifang Niu, Charles Lu, Mingchao Xie, Qunyuan Zhang, Joshua F. McMichael, Matthew A. Wyczalkowski, Mark D. M. Leiserson, Christopher A. Miller, John S. Welch, Matthew J. Walter, Michael C. Wendl, Timothy J. Ley, Richard K. Wilson, Benjamin J. Raphael and Li Ding (2013). "Mutational landscape and significance across 12 major cancer types." Nature **502**: 333-339.

Levi A. Garraway and Eric S. Lander (2013). "Lessons from the Cancer Genome." Cell **153**(1): 17-37.

Laura D. Wood, D. Williams Parsons, Siân Jones, Jimmy Lin, Tobias Sjöblom, Rebecca J. Leary, Dong Shen, Simina M. Boca, Thomas Barber, Janine Ptak, Natalie Silliman, Zoltan Dezso Steve Szabo, Vadim Ustyanksky, Tatiana Nikolskaya, Yuri Nikolsky, Rachel Karchin, Paul A. Wilson, Joshua S. Kaminker, Zemin Zhang,

Joseph Willis Randal Croshaw, Dawn Dawson, Michail Shipitsin, James K. V. Willson, Saraswati Sukumar, Kornelia Polyak, Ben Ho Park, Charit L. Pethiyagoda, P. V. Krishna Pant, Dennis G. Ballinger, Andrew B. Sparks, James Hartigan, Douglas R. Smith, Erick Suh, Nickolas Papadopoulos, Phillip Buckhaults, Sanford D. Markowitz, Giovanni Parmigiani, Kenneth W. Kinzler, Victor E. Velculescu and Bert Vogelstein (2007). "The Genomic Landscapes of Human Breast and Colorectal Cancers." Science **318**(5853): 1108-1113.

M Gerlinger, A J Rowan, S Horswell, J Larkin, D Endesfelder, E Gronroos and Matthews N P Martinez, Stewart A, Tarpey P, Varela I, Phillimore B, Begum S, McDonald NQ, Butler A, Jones D, Raine K, Latimer C, Santos CR, Nohadani M, Eklund AC, Spencer-Dene B, Clark G, Pickering L, Stamp G, Gore M, Szallasi Z, Downward J, Futreal PA, Swanton C (2012). "Intratumor heterogeneity and branched evolution revealed by multiregion sequencing." The New England Journal of Medicine **366**(10): 883-892.

Nicholas C Turner and Jorge S Reis-Filho (2012). "Genetic heterogeneity and cancer drug resistance." Lancet Oncology **13**: e178-185.

M Gerlinger and C Swanton (2010). "How Darwinian models inform therapeutic failure initiated by clonal heterogeneity in cancer medicine." British Journal of Cancer **103**: 1139-1143.

Weiyi Toy, Yang Shen, Helen Won, Bradley Green, Rita A Sakr, Marie Will, Zhiqiang Li, Kinisha Gala, Sean Fanning, Tari A King, Clifford Hudis, David Chen, Tetiana Taran, Gabriel Hortobagyi, Geoffrey Greene, Michael Berger, José Baselga and Sarat Chandralapaty (2013). "ESR1 ligand-binding domain mutations in hormone-resistant breast cancer." Nature Genetics **45**: 1439-1445.

Maria Kleppe and Ross L Levine (2014). "Assessing the implications." Nature Medicine **20**(4): 342-344.

Davide Rossi, Hossein Khiabani, Valeria Spina, Carmela Ciardullo, Alessio Brusca, Rosella Famà, Silvia Rasi, Sara Monti, Clara Deambrogio, Lorenzo De Paoli, Jiguang Wang, Valter Gattei, Anna

Guarini, Robin Foà, Raul Rabadan and Gianluca Gaidano (2014). "Clinical impact of small TP53 mutated subclones in chronic lymphocytic leukemia " Blood **123**(14): 2139-2147.

Elli Papaemmanuil, Moritz Gerstung, Luca Malcovati, Sudhir Tauro, Gunes Gundem, Peter Van Loo, Chris J. Yoon, Peter Ellis, David C. Wedge, Andrea Pellagatti, Adam Shlien, Michael John Groves, Simon A. Forbes, Keiran Raine, Jon Hinton, Laura J. Mudie, Stuart McLaren, Claire Hardy, Calli Latimer, Matteo G. Della Porta, Sarah O'Meara, Ilaria Ambaglio, Anna Galli, Adam P. Butler, Gunilla Walldin, Jon W. Teague, Lynn Quek, Alex Sternberg, Carlo Gambacorti-Passerini, Nicholas C. P. Cross, Anthony R. Green, Jacqueline Boultonwood, Paresh Vyas, Eva Hellstrom-Lindberg, David Bowen, Mario Cazzola, Michael R. Stratton, Peter J. Campbell and Chronic Myeloid Disorders working group of the International Cancer Genome Consortium (2013). "Clinical and biological implications of driver mutations in myelodysplastic syndromes." Blood **122**(22): 3616-3627.

Dan A. Landau, Scott L. Carter, Petar Stojanov, Aaron McKenna, Kristen Stevenson, Michael S. Lawrence, Carrie Sougnez, Chip Stewart, Andrey Sivachenko, Lili Wang, Youzhong Wan, Wandi Zhang, Sachet A. Shukla, Alexander Vartanov, Stacey M. Fernandes, Gordon Saksena, Kristian Cibulskis, Bethany Tesar, Stacey Gabriel, Nir Hacohen, Matthew Meyerson, Eric S. Lander, Donna Neubergh, Jennifer R. Brown, Gad Getz and Catherine J. Wu (2013). "Evolution and Impact of Subclonal Mutations in Chronic Lymphocytic Leukemia." Cell **152**: 714-726.

Carlo C Maley, Patricia C Galipeau, Jennifer C Finley, V Jon Wongsurawat, Xiaohong Li, Carissa A Sanchez, Thomas G Paulson, Patricia L Blount, Rosa-Ana Risques, Peter S Rabinovitch and Brian J Reid (2006). "Genetic clonal diversity predicts progression to esophageal adenocarcinoma." Nature Genetics **38**(4): 468-473.

Philippe L. Bedard, Aaron R. Hansen, Mark J. Ratain and Lillian L. Siu (2013). "Tumour heterogeneity in the clinic." Nature **501**(355-364).

Li Ding, Timothy J. Ley, David E. Larson, Christopher A. Miller, Daniel C. Koboldt, John S. Welch, Julie K. Ritchey, Margaret A. Young, Tamara Lamprecht, Michael D. McLellan, Joshua F. McMichael, John W. Wallis, Charles Lu, Dong Shen, Christopher C. Harris, David J. Dooling, Robert S. Fulton, Lucinda L. Fulton, Ken Chen, Heather Schmidt, Joelle Kalicki-Veizer, Vincent J. Magrini, Lisa Cook, Sean D. McGrath, Tammi L. Vickery, Michael C. Wendl, Sharon Heath, Mark A. Watson, Daniel C. Link, Michael H. Tomasson, William D. Shannon, Jacqueline E. Payton, Shashikant Kulkarni, Peter Westervelt, Matthew J. Walter, Timothy A. Graubert, Elaine R. Mardis, Richard K. Wilson and John F. DiPersio (2012). "Clonal evolution in relapsed acute myeloid leukaemia revealed by whole-genome sequencing." Nature **481**: 506-510.

Nicholas B. La Thangue and David J. Kerr (2011). "Predictive biomarkers: a paradigm shift towards personalized cancer medicine." Nature Reviews Clinical Oncology **8**: 587-596.

Jose Baselga (2006). "Targeting Tyrosine Kinases in Cancer: The Second Wave." Science **312**(5777): 1175-1178.

Miriam Martini, Loredana Vecchione, Salvatore Siena, Sabine Tejpar and Alberto Bardelli (2012). "Targeted therapies: how personal should we go?" Nature Reviews Clinical Oncology **9**: 87-97.

Shigetsugu Hatakeyama (2011). "TRIM proteins and cancer." Nature Reviews Cancer **11**: 792-804.

Christopher E Berndsen and Cynthia Wolberger (2014). "New insights into ubiquitin E3 ligase mechanism." Nature Structural & Molecular Biology **21**: 301–307

S Lipkowitz and A M Weissman (2011). "RINGs of good and evil: RING finger ubiquitin ligases at the crossroads of tumour suppression and oncogenesis." Nature Reviews Cancer **11**(9): 629-643.

Kazuhiro Ikeda and Satoshi Inoue (2012). "TRIM proteins as RING finger E3 ubiquitin ligases." Adv Exp Med Biol **770**: 27-37.

Valérie Lallemand-Breitenbach and Hugues de Thé (2010). "PML Nuclear Bodies." Cold Spring Harbor perspectives in biology **2**(5): a000661.

Massimiliano Mazza and Pier Giuseppe Pelicci (2013). "Is PML a tumor suppressor?" Frontiers in Oncology **3**.

B Herquel, K Ouarrhni, K Khetchoumian, M Ignat, M Teletin, M Mark, G Béchade, A Van Dorsselaer, S Sanglier-Cianférani, A Hamiche, F Cammas, I Davidson and R Losson (2011). "Transcription cofactors TRIM24, TRIM28, and TRIM33 associate to form regulatory complexes that suppress murine hepatocellular carcinoma " Proc Natl Acad Sci U S A. **108**(20): 8212-8217.

Dhanendra Tomar and Rajesh Singh (2015). "TRIM family proteins: emerging class of RING E3 ligases as regulator of NF- κ B pathway." Biol. Cell **107**: 22–40.

J. J. Luciani, D. Depetris, Y. Usson, C. Metzler-Guillemain, C. Mignon-Ravix, M. J. Mitchell, A. Megarbane, P. Sarda, H. Sirma, A. Moncla, J. Feunteun and M.-G. Mattei (2006). "PML nuclear bodies are highly organised DNA-protein structures with a function in heterochromatin remodelling at the G2 phase." Journal of cell science **119**(12): 2518-2531.

Rosa Bernardi and Pier Paolo Pandolfi (2007). "Structure, dynamics and functions of promyelocytic leukaemia nuclear bodies." Nature reviews Molecular cell biology **8**(12): 1006-1016.

L C Trotman, A Alimonti, P P Scaglioni, J A Koutcher, C Cordon-Cardo and P P Pandolfi (2006). "Identification of a tumor Suppressor Network Opposing Nuclear Akt Function " Nature **441**(7092): 523-527.

Paolo Salomoni and Pier Paolo Pandolfi (2002). "The Role of PML in Tumor Suppression." Cell **108**(2): 165-170.

Fumio Nakahara, Cary N. Weiss and Keisuke Ito (2014). "The role of PML in hematopoietic and leukemic stem cell maintenance." International Journal of Hematology **100**: 18-26.

A Carracedo, D Weiss, A K Leliaert, M Bhasin, V C de Boer, G Laurent, A C Adams, M Sundvall, S J Song, K Ito, L S Finley, A Egia, T Libermann, Z Gerhart-Hines, P Puigserver, M C Haigis, E Maratos-Flier, A L Richardson, Z T Schafer and P P Pandolfi (2012). "A Metabolic Pro-Survival Role for PML in Breast Cancer." Journal of Clinical Investigation **122**(9): 3088-3100.

Sue Haupt, Silvia di Agostino, Inbal Mizrahi, Osnat Alsheich-Bartok, Mathijs Voorhoeve, Alex Damalas, Giovanni Blandino and Ygal Haupt (2009). "Promyelocytic Leukemia Protein is Required for Gain of Function by Mutant p53." Cancer Res **69**: 4818-4826.

Y Liu, R Raheja, N Yeh, D Ciznadija, AM Pedraza, T Ozawa, E Hukkelhoven, H Erdjument-Bromage, P Tempst, NP Gauthier, C Brennan, EC Holland and A Koff (2014). "TRIM3, a tumor suppressor linked to regulation of p21Waf1/Cip1." Oncogene **33**: 308–315.

Radhika Raheja, Yuhui Liu, Ellen Hukkelhoven, Nancy Yeh and Andrew Koff (2014). "The ability of TRIM3 to induce growth arrest depends on RING-dependent E3 ligase activity " Biochemistry Journal **458**: 537–545.

Jean-Louis Boulay, Urs Stiefel, Elisabeth Taylor, Béatrice Dolder, Adrian Merlo and Frank Hirth (2009). "Loss of heterozygosity of TRIM3 in malignant gliomas." BMC Cancer **9**: 71.

Gang Chen, Jun Kong, Carol Tucker-Burden, Monika Anand, Yuan Rong, Fahmia Rahman, Carlos S. Moreno, Erwin G. Van Meir, Constantinos G. Hadjipanayis and Daniel J. Brat (2014). "Human Brat Ortholog TRIM3 Is a Tumor Suppressor That Regulates Asymmetric Cell Division in Glioblastoma " Cancer Res **74**: 4536-4548.

Jie Chao, Xiao-Fei Zhang, Qiu-Zhong Pan, Jing-Jing Zhao, Shan-Shan Jiang , Ying Wang, Jian-Hua Zhang and Jian-Chuan Xia (2014). "Decreased expression of TRIM3 is associated with poor prognosis in patients with primary hepatocellular carcinoma " Med Oncol **31**: 102.

Mariano Francesco Caratozzolo, Lucia Micale, Maria Giuseppina Turturo, Silvia Cornacchia, Carmela Fusco, Flaviana Marzano, Bartolomeo Augello, Anna Maria D'Erchia, Luisa Guerrini, Graziano Pesole, Elisabetta Sbisà, Giuseppe Merla and Apollonia Tullo (2012). "TRIM8 modulates p53 activity to dictate cell cycle arrest " Cell Cycle **11**(3): 511-523.

F Carinci, D Arcelli, L Lo Muzio, F Francioso, D Valentini, R Evangelisti, S Volinia, A D'Angelo, G Meroni, M Zollo, A Pastore, F Ionna, F Mastrangelo, P Conti and S Tetè (2007). "Molecular classification of nodal metastasis in primary larynx squamous cell carcinoma." Translational Research **150**(4): 233-245.

Mariano F. Caratozzolo, Alessio Valletti, Margherita Gigante, Italia Aiello, Francesca Mastropasqua, Flaviana Marzano, Pasquale Ditunno, Giuseppe Carrieri, Hélène Simonnet, Anna Maria D'Erchia, Elena Ranieri, Graziano Pesole, Elisabetta Sbisà and Apollonia Tullo (2014). "TRIM8 anti-proliferative action against chemo-resistant renal cell carcinoma " Oncotarget **5**(17): 7446-7457.

Masashi Watanabe, Tadasuke Tsukiyama and Shigetsugu Hatakeyama (2009). "TRIM31 interacts with p52Shc and inhibits Src-induced anchorage-independent growth." Biochemical and Biophysical Research Communications **388**: 422-427.

Takeyuki Sugiura and Kentaro Miyamoto (2008). "Characterization of TRIM31, upregulated in gastric adenocarcinoma, as a novel RBCC protein " Journal of Cellular Biochemistry **105**(4): 1081–1091.

H Li, Y Zhang, Y Zhang, X Bai, Y Peng and P He (2014). "TRIM31 is downregulated in non-small cell lung cancer and serves as a potential tumor suppressor." Tumour Biology **35**(6): 5747-5752.

Hae Mi Jooa , Ji Young Kima, Jae Boon Jeonga, Ki Moon Seonga, Seon Young Nama, Kwang Hee Yanga, Cha Soon Kima, Hee Sun Kima, Meeseon Jeonga, Sungkwan Anb and Young Woo Jin (2011). "Ret finger protein 2 enhances ionizing radiation-induced apoptosis via degradation of AKT and MDM2." European Journal of Cell Biology **90**: 420–431.

Xuetian Yue, Juan Liu and Zhaohui Feng (2014). "Tumor suppressor p53 and TRIM family proteins " Cancer Cell & Microenvironment **1**: e470.

Dhanendra Tomar and Rajesh Singh (2014). "TRIM13 regulates ubiquitination and turnover of NEMO to suppress TNF induced NF- κ B activation." Cellular Signaling **26**: 2606–2613.

Dhanendra Tomar, Paresch Prajapati, Lakshmi Sripada, Kritarth Singh, Rochika Singh, Arun Kumar Singh and Rajesh Singh (2013). "TRIM13 regulates caspase-8 ubiquitination, translocation to autophagosomes and activation during ER stress induced cell death " Biochimica et Biophysica Acta **1833**: 3134–3144.

Dhanendra Tomar, Rochika Singh, Arun Kumar Singh, Chirayu D. Pandya and Rajesh Singh (2012). "TRIM13 regulates ER stress induced autophagy and clonogenic ability of the cells." Biochimica et Biophysica Acta **1823**: 316–326.

W.J. van Everdinka, A. Baranovab, C. Lummena, T. Tyazhelovab, M.W. Loomana, D. Ivanov, E. Verlind, A. Pestova, H. Faber, A.Y. van der Veen, N. Yankovsky, E. Vellenga and C. H. Buys (2003). "RFP2, c13ORF1, and FAM10A4 are the most likely tumor suppressor gene candidates for B-cell chronic lymphocytic leukemia." Cancer Genet Cytogenet. **146**(1): 48-57.

I Lassot, I Robbins, M Kristiansen, R Rahmeh, F Jaudon, MM Magiera, L Vanhille S Mora, A Lipkin, B Pettmann, J Ham and S Desagher (2010). "Trim17, a novel E3 ubiquitin-ligase, initiates neuronal apoptosis " Cell Death and Differentiation **17**: 1928–1941.

M M Magiera, S Mora, B Mojsa, I Robbins, I Lassot and S Desagher (2013). "Trim17-mediated ubiquitination and degradation of Mcl-1 initiate apoptosis in neurons." Cell Death and Differentiation **20**: 281–292.

B Mojsa, S Mora, J P Bossowski, I Lassot and S Desagher (2015). "Control of neuronal apoptosis by reciprocal regulation of NFATc3 and Trim17." Cell Death and Differentiation **22**: 274–286.

Hiroshi Endo, Kazuhiro Ikeda, Tomohiko Urano, Kuniko Horie-Inoue and Satoshi Inoue (2012). "Terf/TRIM17 stimulates degradation of kinetochore protein ZWINT and regulates cell proliferation " J Biochem **151**(2): 139-144.

Hesed M. Padilla-Nash, Nicole E. McNeil, Ming Yi, Quang-Tri Nguyen, Yue Hu, Danny Wangsa, David L. Mack, Amanda B. Hummon, Chanelle Case, Eric Cardin, Robert Stephens, Michael J. Difilippantonio and Thomas Ried (2013). "Aneuploidy, oncogene amplification and epithelial to mesenchymal transition define spontaneous transformation of murine epithelial cells " Carcinogenesis **34**(8): 1929-1939.

Yeung Sook Ryu, Younglang Lee, Keun Woo Lee, Chae Young Hwang, Jin-Soo Maeng, Jeong-Hoon Kim, Yeon-Soo Seo, Kwan-Hee You, Byeongwoon Song and Ki-Sun Kwon (2011). "TRIM32 Protein Sensitizes Cells to Tumor Necrosis Factor (TNF α)-induced Apoptosis via Its RING Domain-dependent E3 Ligase Activity against X-linked Inhibitor of Apoptosis (XIAP) " The Journal of Biochemistry **286**(29): 25729–25738.

Nai-Jia Huang, Liguang Zhang, Wanli Tang, Chen Chen, Chih-Sheng Yang and Sally Kornbluth (2012). "The Trim39 ubiquitin ligase inhibits APC/CCdh1-mediated degradation of the Bax activator MOAP-1." Journal of Cell Biology **197**(3): 361-367.

Taro Kawai and Shizuo Akira (2011). "Regulation of innate immune signaling pathways by the tripartite motif (TRIM) family proteins." EMBO Mol Med **3**: 513–527.

Michael Karin (2009). "NF- κ B as a Critical Link Between Inflammation and Cancer " Cold Spring Harb Perspect Biol. **1**(5): a000141.

Mio Shibata, Tomonobu Sato, Ryota Nukiwa, Tadashi Ariga and Shigetsugu Hatakeyama (2012). "TRIM45 negatively regulates NF- κ B-mediated transcription and suppresses cell proliferation " Biochemical and Biophysical Research Communications **423**: 104–109.

M Niida, M Tanaka and T Kamitani (2010). "Downregulation of active IKK beta by Ro52-mediated autophagy." Mol Immunol. **47**(14): 2378-2387.

M Shi, W Deng, E Bi, K Mao, Y Ji, G Lin, X Wu, Z Tao, Z Li, X Cai, S Sun, C Xiang and B Sun (2008). "TRIM30 alpha negatively regulates TLR-mediated NF- κ B activation by targeting TAB2 and TAB3 for degradation." Nat Immunol. **9**(4): 369-377.

W S Wu, Z X Xu, W N Hittelman, P Salomoni, P P Pandolfi and K S Chang (2003). "Promyelocytic leukemia protein sensitizes tumor necrosis factor alpha-induced apoptosis by inhibiting the NF-kappaB survival pathway." J Biol Chem. **278**(14): 12294-12304.

Mude Shi, Hyelim Cho, Kyung-Soo Inn, Aerin Yang, Zhen Zhao, Qiming Liang, Gijs A. Versteeg, Samad Amini-Bavil-Olyaei, Lai-Yee Wong, Berislav V. Zlokovic, Hee-Sung Park, Adolfo Garcí'a-Sastre and Jae U. Jung (2014). "Negative regulation of NF-kB activity by brain-specific TRIPartite Motif protein 9 " Nat. Commun. **5**: 4820.

Keita Noguchi, Fumihiko Okumura, Norihiko Takahashi, Akihiko Kataoka, Toshiya Kamiyama, Satoru Todo and Shigetsugu Hatakeyama (2011). "TRIM40 promotes neddylation of IKKg and is downregulated in gastrointestinal cancers." Carcinogenesis **32**(7): 995–1004.

Konstantin Khetchoumian, Marius Teletin, Johan Tisserand, Benjamin Herquel, Khalid Ouararhni and Régine Losson (2008). "Trim24 (Tif1 α): An essential 'brake' for retinoic acid-induced transcription to prevent liver cancer." Cell Cycle **7**(23): 3647-3652.

Konstantin Khetchoumian, Marius Teletin, Johan Tisserand, Manuel Mark, Benjamin Herquel, Mihaela Ignat, Jessica Zucman-Rossi, Florence Cammas, Thierry Lerouge, Christelle Thibault, Daniel Metzger, Pierre Chambon and Régine Losson (2007). "Loss of Trim24 (Tif1 α) gene function confers oncogenic activity to retinoic acid receptor alpha." Nature Genetics **39**: 1500-1506.

D G Ransom, N Bahary, K Niss, D Traver, C Burns, N S Trede, N Paffett-Lugassy, W J Saganic, C A Lim, C Hersey, Y Zhou, B A Barut, S Lin, P D Kingsley, J Palis, S H Orkin and L I Zon (2004). "The

Zebrafish moonshine Gene Encodes Transcriptional Intermediary Factor 1γ , an Essential Regulator of Hematopoiesis." PLoS Biol. 2(8): E237.

Romain Aucagne, Nathalie Droin, Jérôme Paggetti, Brice Lagrange, Anne Largeot, Arlette Hammann, Amandine Bataille, Laurent Martin, Kai-Ping Yan, Pierre Fenaux, Régine Losson, Eric Solary, Jean-Noël Bastie and Laurent Delva (2011). "Transcription intermediary factor 1γ is a tumor suppressor in mouse and human chronic myelomonocytic leukemia." J Clin Invest. 121(6): 2361–2370.

David F. Vincent, Kai-Ping Yan, Isabelle Treilleux, Fabien Gay, Vanessa Arfi, Bastien Kaniewsky, Julien C. Marie, Florian Lepinasse, Sylvie Martel, Sophie Goddard-Leon, Juan L. Iovanna, Pierre Dubus, Stéphane Garcia, Alain Puisieux, Ruth Rimokh, Nabeel Bardeesy, Jean-Yves Scoazec, Régine Losson and Laurent Bartholin (2009). "Inactivation of TIF1 γ Cooperates with KRASG12D to Induce Cystic Tumors of the Pancreas " PLoS Genet. 5(7): e1000575.

Ok-Hee Leea, Jinkyoungh Leeb, Keun Ho Leeb, Yun Mi Wooa, Ju-Hee Kangc, Ho-Geun Yoone, Soo-Kyung Baef, Zhou Songyangg, Seung Hyun Ohc and Youngsok Choi (2015). "Role of the focal adhesion protein TRIM15 in colon cancer development." Biochimica et Biophysica Acta- Molecular Cell Research 1853(2): 409–421.

Yi Cheng, Zhi Yan, Yin Liu, Chengbai Liang, Hong Xia, Junming Feng, Guorong Zheng and Hesheng Luo (2014). "Analysis of DNA methylation patterns associated with the gastric cancer genome." Oncol Lett. 7(4): 1021-1026.

Belamy B Cheung, Jessica Koach, Owen Tan, Patrick Kim, Jessica L Bell, Carla D'andreti, Selina Sutton, Alena Malyukova, Eric Sekyere, Murray Norris, Michelle Haber, Maria Kavallaris, Anne M

Cunningham, Charlotte Proby, Irene Leigh, James S Wilmott, Caroline L Cooper, Gary M Halliday, Richard A Scolyer and Glenn M Marshall (2012). "The retinoid signalling molecule, TRIM16, is repressed during squamous cell carcinoma skin carcinogenesis in vivo and reduces skin cancer cell migration in vitro " J Pathol **226**: 451 – 462.

GM Marshall, JL Bell, J Koach, O Tan, P Kim, A Malyukova, W Thomas, EO Sekyere, T Liu, AM Cunningham, V Tobias, MD Norris, M Haber, M Kavallaris and BB Cheung (2010). "TRIM16 acts as a tumour suppressor by inhibitory effects on cytoplasmic vimentin and nuclear E2F1 in neuroblastoma cells " Oncogene **29**: 6172–6183.

Selina K. Sutton, Jessica Koach, Owen Tan, Bing Liu, Daniel R. Carter, James S. Wilmott, Benafsha Yosufi, Lauren E. Haydu, Graham J. Mann, John F. Thompson, Georgina V. Long, Tao Liu, Grant McArthur, Xu Dong Zhang, Richard A. Scolyer, Belamy B. Cheung and Glenn M. Marshall (2014). "TRIM16 inhibits proliferation and migration through regulation of interferon beta 1 in melanoma cells " Oncotarget **5**(20): 10127-10139.

H D Beer, C Munding, N Dubois, C Mamie, D Hohl and S Werner (2002). "The Estrogen-responsive B Box Protein: ANOVEL REGULATOR OF KERATINOCYTE DIFFERENTIATION " J. Biol. Chem **277**: 20740-20749.

Jessica L. Bell, Alena Malyukova, Maria Kavallaris, Glenn M. Marshall and Belamy B. Cheung (2013). "TRIM16 inhibits neuroblastoma cell proliferation through cell cycle regulation and dynamic nuclear localization " Cell Cycle **12**(6): 889–898.

S T Lott, N Chen, D S Chandler, Q Yang, L Wang, M Rodriguez, H Xie, S Balasenthil, T A Buchholz, A A Sahin, K Chaung, B Zhang, S E Olufemi, J Chen, H Adams, V Band, A K El-Naggar, M L Frazier, K Keyomarsi, K K Hunt, S Sen, B Haffty, S M Hewitt, R Krahe and A M Killary (2009). "DEAR1 is a dominant regulator of acinar morphogenesis and an independent predictor of local recurrence-free survival in early-onset breast cancer." PLoS Med. 6(5): e1000068.

J Liu, B Welm, K M Boucher, M T Ebbert and P S Bernard (2012). "TRIM29 Functions as a Tumor Suppressor in Non-tumorigenic Breast Cells and Invasive ER+ Breast Cancer." Am J Pathol. 180(2): 839-847.

L Ai, W J Kim, M Alpay, M Tang, C E Pardo, S Hatakeyama, W S May, M P Kladde, C D Heldermon, E M Siegel and K D Brown (2014). "TRIM29 Suppresses TWIST1 and Invasive Breast Cancer Behavior." Cancer Res. 74(17): 4875-4887.

T Ernst, M Hergenhahn, M Kenzelmann, C D Cohen, M Bonrouhi, A Weninger, R Klären, E F Gröne, M Wiesel, C Güdemann, J Küster, W Schott, G Staehler, M Kretzler, M Hollstein and H J Gröne (2002). "Decrease and gain of gene expression are equally discriminatory markers for prostate carcinoma: a gene expression analysis on total and microdissected prostate tissue." Am J Pathol. 160(6): 2169-2180.

A Avraham, S Cho, R Uhlmann, M Polak, J Sandbank, I T Karni, Pappo, R Halperin, Z Vaknin, A Sella, S Sukumar and E Evron (2014). "Tissue Specific DNA Methylation in Normal Human Breast Epithelium and in Breast Cancer." PLoS One 9(3): e91805.

H J Burstein, K Polyak, J S Wong, S C Lester and C M Kaelin (2004). "Ductal carcinoma in situ of the breast." The New England Journal of Medicine 350(14): 1430-1441.

Joan Massagué and Qiaoran Xi (2012). "TGF- β control of stem cell differentiation genes." FEBS Letters **586**(14): 1953-1958.

Alfonso Quintás-Cardama, Sean M Post, Luisa M Solis, Shunbin Xiong, Peirong Yang, Nanyue Chen, Ignacio I Wistuba, Ann M Killary and Guillermina Lozano (2014). "Loss of the novel tumour suppressor and polarity gene Trim62 (Dear1) synergizes with oncogenic Ras in invasive lung cancer." J Pathol **234**: 108–119.

Franziska Michor, Yoh Iwasa, Bert Vogelstein, Christoph Lengauer and Martin A. Nowak (2005). "Can chromosomal instability initiate tumorigenesis?" Seminars in Cancer Biology **15**(1): 43-49.

Ildiko Balint, Anke Müller, Anetta Nagy and Gyula Kovacs (2004). "Cloning and characterisation of the RBCC728/TRIM36 zinc-binding protein from the tumor suppressor gene region at chromosome 5q22.3." Gene **332**: 45-50.

John Koreth, Peter B. Bethwaite and James O' D. McGee (1995). "Mutation at chromosome 11q23 in human non-familial breast cancer: A microdissection microsatellite analysis." Journal of Pathology **176**(1): 11-18.

G Ragnarsson, G Eiriksdottir, J Th Johannsdottir, J G Jonasson, V Egilsson and S Ingvarsson (1999). "Loss of heterozygosity at chromosome 1p in different solid human tumours: association with survival." British Journal of Cancer **79**(9-10): 1468-1474.

G Ragnarsson, A Sigurdsson, G Eiriksdottir, R Barkardottir, J Jonasson and S Ingvarsson (1996). "Loss of heterozygosity at chromosome 1p in human breast cancer." International Journal of Oncology 9: 731-736.

The Cancer Genome Atlas Network (2012). "Comprehensive molecular portraits of human breast tumors." Nature 490: 61-70.

Ethan Cerami, Jianjiong Gao, Ugur Dogrusoz, Benjamin E. Gross, Selcuk Onur Sumer, Bülent Arman Aksoy, Anders Jacobsen, Caitlin J. Byrne, Michael L. Heuer, Erik Larsson, Yevgeniy Antipin, Boris Reva, Arthur P. Goldberg, Chris Sander and Nikolaus Schultz (2012). "The cBio Cancer Genomics Portal: An Open Platform for Exploring Multidimensional Cancer Genomics Data." Cancer Discov. 2: 401.

Jianjiong Gao, Bülent Arman Aksoy, Ugur Dogrusoz, Gideon Dresdner, Benjamin Gross, S. Onur Sumer, Yichao Sun, Anders Jacobsen, Rileen Sinha, Erik Larsson, Ethan Cerami, Chris Sander and Nikolaus Schultz (2013). "Integrative Analysis of Complex Cancer Genomics and Clinical Profiles Using the cBioPortal." Science Signaling 6(269): p11.

Kornelia Polyak (2014). "A case for Darwinian tumor evolution." Nature Medicine 20(4): 344-346.

Simon A. Forbes, David Beare, Prasad Gunasekaran, Kenric Leung, Nidhi Bindal, Harry Boutselakis, Minjie Ding, Sally Bamford, Charlotte Cole, Sari Ward, Chai Yin Kok, Mingming Jia, Tisham De, Jon W. Teague, Michael R. Stratton, Ultan McDermott and Peter J. Campbell (2014). "COSMIC: exploring the world's knowledge of somatic mutations in human cancer." Nucleic Acids Research 43(D1): D805-D811.

Patrick Frosk, Tracey Weiler, Edward Nylen, Thangirala Sudha, Cheryl R. Greenberg, Kenneth Morgan, T. Mary Fujiwara and Klaus Wrogemann (2002). "Limb-Girdle Muscular Dystrophy Type 2H Associated with Mutation in TRIM32, a Putative E3-Ubiquitin–Ligase Gene." American Journal of Human Genetics **70**(3): 663-672.

Timothy C. Cox, Lillian R. Allen, Liza L. Cox, Blair Hopwood, Bruce Goodwin, Eric Haan and Graeme K. Suthers (2000). "New mutations in MID1 provide support for loss of function as the cause of X-linked Opitz syndrome." Human Molecular Genetics **9**(17): 2553-2562.

Alfonso Quintás-Cardama, Nianxiang Zhang, Yi Hua Qiu, Sean Post, Chad J. Creighton, Jorge Cortes, Kevin R. Coombes and Steven M. Kornblau (2015). "Loss of TRIM62 Expression Is an Independent Adverse Prognostic Factor in Acute Myeloid Leukemia " Clinical Lymphoma Myeloma and Leukemia **15**(2): 115–127.

O W Petersen, L Rønnev-Jessen, A R Howlett and M J Bissell (1992). "Interaction with basement membrane serves to rapidly distinguish growth and differentiation pattern of normal and malignant human breast epithelial cells." Proc Natl Acad Sci U S A. **89**(19): 9064–9068.

Ruby Yun-Ju Huang, Parry Guilford and Jean Paul Thiery (2012). "Early events in cell adhesion and polarity during epithelial-mesenchymal transition." Journal of cell science **125**(19): 4417-4422.

Z Jiang, R Jones, J C Liu, T Deng, T Robinson, P E Chung, S Wang, J I Herschkowitz, S E Egan, C M Perou and E Zacksenhaus (2011). "RB1 and p53 at the crossroad of EMT and triple-negative breast cancer." Cell Cycle **10**(10): 1563-1570.

Yanhong Zhang, Wensheng Yan and Xinbin Chen (2011). "Mutant p53 Disrupts MCF-10A Cell Polarity in Three-dimensional Culture via Epithelial-to-mesenchymal Transitions." Journal of Biological Chemistry **286**: 16218-16228.

Johanna I. Partanen, Topi A. Tervonen, Mikko Myllynen, Essi Lind, Misa Imai, Pekka Katajisto, Gerrit J. P. Dijkgraaf, Panu E. Kovanen, Tomi P. Mäkelä, Zena Werb and Juha Klefström (2012). "Tumor suppressor function of Liver kinase B1 (Lkb1) is linked to regulation of epithelial integrity." Proc Natl Acad Sci U S A. **109**(7): E388-E397.

H Fenton, B Carlile, E A Montgomery, H Carraway, J Herman, F Sahin, G H Su and P Argani (2006). "LKB1 protein expression in human breast cancer." Appl Immunohistochem Mol Morphol **14**(2): 146-153.

L Zhan, A Rosenberg, K C Bergami, M Yu, Z Xuan, A B Jaffe, C Allred and S K Muthuswamy (2008). "Deregulation of scribble promotes mammary tumorigenesis and reveals a role for cell polarity in carcinoma." Cell **135**(5): 865-878.

Nathan J. Godde, Julie M. Sheridan, Lorey K. Smith, Helen B. Pearson, Kara L. Britt, Ryan C. Galea, Laura L. Yates, Jane E. Visvader and Patrick O. Humbert (2014). "Scribble Modulates the MAPK/Fra1 Pathway to Disrupt Luminal and Ductal Integrity and Suppress Tumour Formation in the Mammary Gland." PLoS Genet. **10**(5): e1004323.

Charles L. Vogel, Melody A. Cobleigh, Debu Tripathy, John C. Gutheil, Lyndsay N. Harris, Louis Fehrenbacher, Dennis J. Slamon, Maureen Murphy, William F. Novotny, Michael Burchmore, Steven Shak, Stanford J. Stewart and Michael Press (2002). "Efficacy and Safety of Trastuzumab as a Single Agent in First-Line Treatment of HER2-Overexpressing Metastatic Breast Cancer." Journal of Clinical Oncology **20**(3): 719-726.

Paul B. Chapman, Axel Hauschild, Caroline Robert, John B. Haanen, Paolo Ascierto, James Larkin, Reinhard Dummer, Claus Garbe, Alessandro Testori, Michele Maio, David Hogg, Paul Lorigan, Celeste Lebbe, Thomas Jouary, Dirk Schadendorf, Antoni Ribas, Steven J. O'Day, Jeffrey A. Sosman, John M. Kirkwood, Alexander M.M. Eggermont, Brigitte Dreno, Keith Nolop, Jiang Li, Betty Nelson, Jeannie Hou, Richard J. Lee, Keith T. Flaherty, Grant A. McArthur and BRIM-3 Study Group (2011). "Improved Survival with Vemurafenib in Melanoma with BRAF V600E Mutation." New England Journal of Medicine **364**: 2507-2516.

Harold J. Burstein, Lyndsay N. Harris, P. Kelly Marcom, Rosemary Lambert-Falls, Kathleen Havlin, Beth Overmoyer, Robert J. Friedlander Jr., Janet Gargiulo, Rochelle Strenger, Charles L. Vogel, Paula D. Ryan, Mathew J. Ellis, Raquel A. Nunes, Craig A. Bunnell, Susana M. Campos, Michele Hallor, Rebecca Gelman and Eric P. Winer (2003). "Trastuzumab and Vinorelbine as First-Line Therapy for HER2-Overexpressing Metastatic Breast Cancer: Multicenter Phase II Trial With Clinical Outcomes, Analysis of Serum Tumor Markers as Predictive Factors, and Cardiac Surveillance Algorithm." Journal of Clinical Oncology **21**(15): 2889-2895.

James Larkin, Paolo A. Ascierto, Brigitte Dréno, Victoria Atkinson, Gabriella Liskay, Michele Maio, Mario Mandalà, Lev Demidov, Daniil Stroyakovskiy, Luc Thomas, Luis de la Cruz-Merino, Caroline

Dutriaux, Claus Garbe, Mika A. Sovak, Ilsung Chang, Nicholas Choong, Stephen P. Hack, Grant A. McArthur and Antoni Ribas (2014). "Combined Vemurafenib and Cobimetinib in BRAF-Mutated Melanoma." New England Journal of Medicine **371**: 1867-1876.

Michael S. Lawrence, Petar Stojanov, Paz Polak, Gregory V. Kryukov, Kristian Cibulskis, Andrey Sivachenko, Scott L. Carter, Chip Stewart, Craig H. Mermel, Steven A. Roberts, Adam Kiezun, Peter S. Hammerman, Aaron McKenna, Yotam Drier, Lihua Zou, Alex H. Ramos, Trevor J. Pugh, Nicolas Stransky, Elena Helman, Jaegil Kim, Carrie Sougnez, Lauren Ambrogio, Elizabeth Nickerson, Erica Shefler, Maria L. Cortés, Daniel Auclair, Gordon Saksena, Douglas Voet, Michael Noble, Daniel DiCara, Pei Lin, Lee Lichtenstein, David I. Heiman, Timothy Fennell, Marcin Imielinski, Bryan Hernandez, Eran Hodis, Sylvan Baca, Austin M. Dulak, Jens Lohr, Dan-Avi Landau, Catherine J. Wu, Jorge Melendez-Zajgla, Alfredo Hidalgo-Miranda, Amnon Koren, Steven A. McCarroll, Jaume Mora, Ryan S. Lee, Brian Crompton, Robert Onofrio, Melissa Parkin, Wendy Winckler, Kristin Ardlie, Stacey B. Gabriel, Charles W. M. Roberts, Jaclyn A. Biegel, Kimberly Stegmaier, Adam J. Bass, Levi A. Garraway, Matthew Meyerson, Todd R. Golub, Dmitry A. Gordenin, Shamil Sunyaev, Eric S. Lander and Gad Getz (2013). "Mutational heterogeneity in cancer and the search for new cancer associated genes." Nature **499**: 214-218.

S Perner, J M Mosquera, F Demichelis, M D Hofer, J Simko P L Paris, C Collins, , T A Bismar, A M Chinnaiyan, A M De Marzo and M A Rubin (2007). "TMPRSS2-ERG fusion prostate cancer: an early molecular event associated with invasion." Am J Surg Pathol. **31**(6): 882-888.

Zhigang C. Wang , Ming Lin, Lee-Jen Wei, Cheng Li, Alexander Miron, Gabriella Lodeiro, Lyndsay Harris, Sridhar Ramaswamy, David M. Tanenbaum, Matthew Meyerson, James D. Iglehart and Andrea

Richardson (2004). "Loss of Heterozygosity and Its Correlation with Expression Profiles in Subclasses of Invasive Breast Cancers." Cancer Res **64**(1): 64-71.

Anindya Bagchi, Cristian Papazoglu, Ying Wu, Daniel Capurso, Michael Brodt, Dailia Francis, Markus Bredel, Hannes Vogel and Alea A. Mills (2007). "CHD5 Is a Tumor Suppressor at Human 1p36." Cell **128**(3): 459-475.

N. Allocati, C. Di Ilio and V. De Laurenzi (2012). "p63/p73 in the control of cell cycle and cell death." Experimental Cell Research **318**(11): 1285-1290.

Mun-Su Jung, Jeanho Yun, Hee-Don Chae, Jeong-Min Kim, Sun-Chang Kim, Tae-Saeng Choi and Deug Y Shin (2001). "p53 and its homologues, p63 and p73, induce a replicative senescence through inactivation of NF-Y transcription factor." Oncogene **20**(41): 5818-5825.

J.R. Sampson, S. Jones, S. Dolwani and J.P. Cheadle (2005). "MutYH (MYH) and colorectal cancer." Biochemical Society Transactions **33**(4): 679-683.

Daniel R Rhodes, Shanker Kalyana-Sundaram, Vasudeva Mahavisno, Radhika Varambally, Jianjun Yu, Benjamin B Briggs, Terrence R Barrette, Matthew J Anstet, Colleen Kincaid-Beal, Prakash Kulkarni, Sooryanaryana Varambally, Debashis Ghosh and Arul M Chinnaiyan (2007). "Oncomine 3.0: Genes, Pathways, and Networks in a Collection of 18,000 Cancer Gene Expression Profiles." Neoplasia **9**(2): 166-180.

I A Adzhubei, S Schmidt, L Peshkin, V E Ramensky, A Gerasimova, P Bork, A S Kondrashov and S R Sunyaev (2010). "A method and server for predicting damaging missense mutations." Nature Methods 7(4): 248-249.

P Kumar, S Henikoff and P C Ng (2009). "Predicting the effects of coding non-synonymous variants on protein function using the SIFT algorithm." Nature Protocols 4(7): 1073-1081.

J M Schwarz, D N Cooper, M Schuelke and D Seelow (2014). "MutationTaster2: mutation prediction for the deep-sequencing age." Nature Methods 11(4): 361-362.

N Stransky, A M Egloff, A D Tward, A D Kostic, K Cibulskis, A Sivachenko, G V Kryukov, M S Lawrence, C Sougnez, A McKenna, E Shefler, A H Ramos, P Stojanov, S L Carter, D Voet, M L Cortés, D Auclair, M F Berger, G Saksena, C Guiducci, R C Onofrio, M Parkin, M Romkes, J L Weissfeld, R R Seethala, L Wang, C Rangel-Escareño, J C Fernandez-Lopez, A Hidalgo-Miranda, J Melendez-Zajgla, W Winckler, K Ardlie, S B Gabriel, M Meyerson, E S Lander, G Getz, T R Golub, L A Garraway and J R Grandis (2011). "The mutational landscape of head and neck squamous cell carcinoma." Science 333(6046): 1157-1160.

Christina Curtis, Sohrab P. Shah, Suet-Feung Chin, Gulisa Turashvili, Oscar M. Rueda, Mark J. Dunning, Doug Speed, Andy G. Lynch, Yinyin Yuan Shamith Samarajiwa, Stefan Gräf, Gavin Ha, Gholamreza Haffari, Ali Bashashati, Roslin Russell, Steven McKinney, METABRIC Group, Anita Langerød, Andrew Green, Elena Provenzano, Gordon Wishart, Sarah Pinder, Peter Watson, Florian Markowitz, Leigh Murphy, Ian Ellis, Arnie Purushotham, Anne-Lise Børresen-Dale, James D. Brenton, Simon Tavaré, Carlos Caldas and Samuel Aparicio (2012). "The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups." Nature 486: 346-352.

SE Baldus, KL Schaefer, R Engers, D Hartleb, NH Stoecklein and HE Gabbert (2010). "Prevalence and Heterogeneity of KRAS, BRAF, and PIK3CA Mutations in Primary Colorectal Adenocarcinomas and Their Corresponding Metastases." Clin Cancer Res **16**(3): 790-799.

S P Shah, R D Morin, J Khattra, L Prentice, T Pugh, A Burleigh, A Delaney, K Gelmon, R Guliany, J Senz, R A Holt C Steidl, S Jones, M Sun, G Leung, R Moore, T Severson, G A Taylor, A E Teschendorff, K Tse, G Turashvili, R Varhol, R L Warren, P Watson, Y Zhao, C Caldas, D Huntsman, M Hirst, M A Marra and S Aparicio (2009). "Mutational evolution in a lobular breast tumour profiled at single nucleotide resolution." Nature **461**(7265): 908-813.

Y Antipin B Reva, C Sander (2007). "Determinants of protein function revealed by combinatorial entropy optimization." Genome Biology **8**(11): R232.

Y Antipin B Reva, C Sander (2011). "Predicting the Functional Impact of Protein Mutations: Application to Cancer Genomics." Nucleic Acids Research **39**(17): e118.

Brian J. Druker, François Guilhot, Stephen G. O'Brien, Insa Gathmann, Hagop Kantarjian, Norbert Gattermann, Michael W.N. Deininger, Richard T. Silver, John M. Goldman, Richard M. Stone, Francisco Cervantes, Andreas Hochhaus, Bayard L. Powell, Janice L. Gabrilove, Philippe Rousselot, Josy Reiffers, Jan J. Cornelissen, Timothy Hughes, Hermine Agis, Thomas Fischer, Gregor Verhoef, John Shepherd, Giuseppe Saglio, Alois Gratwohl, Johan L. Nielsen, Jerald P. Radich, Bengt Simonsson, Kerry Taylor, Michele Baccarani, Charlene So, Laurie Letvak and Richard A. Larson (2006). "Five-Year Follow-up of Patients Receiving Imatinib for Chronic Myeloid Leukemia." New England Journal of Medicine **355**(23): 2408-2417.

Enrique Grande, María-Victoria Bolós and Edurne Arriola (2011). "Targeting Oncogenic ALK: A Promising Strategy for Cancer Treatment Molecular Cancer Therapeutics." Mol Cancer Ther **10**(4): 569-579.

C L Vogel (2001). "Herceptin clinical trials: past, present and future." Breast Cancer Res **3**(Suppl 1): A66.

J Sanjmyatav, K Junker, S Matthes, M Muehr, D Sava, M Sternal, S Wessendorf, M Kreuz, M Gajda, H Wunderlich and C Schwaenen (2011). "Identification of Genomic Alterations Associated With Metastasis and Cancer Specific Survival in Clear Cell Renal Cell Carcinoma " J Urol **186**(5): 2078-2083.

Reza Mirnezami, Jeremy Nicholson and Ara Darzi (2012). "Preparing for Precision Medicine." The New England Journal of Medicine **366**: 489-491.

A.N. Desai and A. Jere (2012). "Next generation sequencing: ready for the clinic?" Clinical Genetics **81**: 503-510.

B Sikkema-Raddatz, L F Johansson, E N de Boer, R Almomani, L G Boven, M P van den Berg, K Y van Spaendonck-Zwarts, R H Sijmons J P van Tintelen, J D Jongbloed and R J Sinke (2013). "Targeted next-generation sequencing can replace Sanger sequencing in clinical diagnostics." Human Mutation **2013**(34): 7.

Heidi L. Rehm (2013). "Disease-targeted sequencing: a cornerstone in the clinic." Nat Rev Genet 14(4): 295-300.

J K Sehn, I S Hagemann, J D Pfeifer, C E Cottrell and C M Lockwood (2014). "Diagnostic Utility of Targeted Next Generation Sequencing in Problematic Cases." Am J Surg Pathol. 38(4): 534-541.

H Beltran, R Yelensky, G M Frampton, K Park, S R Downing, T Y MacDonald, M Jarosz, D Lipson, S T Tagawa, D M Nanus, P J Stephens, J M Mosquera, M T Cronin and M A Rubin (2013). "Targeted next-generation sequencing of advanced prostate cancer identifies potential therapeutic targets and disease heterogeneity." Eur Urol 63 (5): 920-926.

J D Holbrook, J S Parker, K T Gallagher, W S Halsey, A M Hughes, V J Weigman, P F Lebowitz and R Kumar (2011). "Deep sequencing of gastric carcinoma reveals somatic mutations relevant to personalized medicine." J Transl Med 9: 119.

A Bashashati, G Ha, A Tone, J Ding, L M Prentice, A Roth, J Rosner, K Shumansky, S Kalloger, J Senz, W Yang, M McConechy, N Melnyk, M Anglesio, M T Luk, K Tse, T Zeng, R Moore, Y Zhao, M A Marra, B Gilks, S Yip, D G Huntsman, J N McAlpine and S P Shah (2013). "Distinct evolutionary trajectories of primary high grade serous ovarian cancers revealed through spatial mutational profiling " J Pathol 231(1): 21-34.

A B Turke, K Zejnullahu, Y L Wu, Y Song, D Dias-Santagata, E Lifshits, L Toschi, A Rogers, T Mok, L Sequist, N I Lindeman, C Murphy, S Akhavanfard, B Y Yeap, Y Xiao, M Capelletti, A J Iafrate, C Lee, J G Christensen, J A Engelman and P A Jänne (2010). "Preexistence and clonal selection of MET amplification in EGFR mutant NSCLC." Cancer Cell 17(1): 77-88.

K Y Su, H Y Chen, K C Li, M L Kuo, J C Yang, W K Chan, B C Ho, G C Chang, J Y Shih, S L Yu and P C Yang (2012). "Pretreatment epidermal growth factor receptor T790M mutation predicts shorter EGFR tyrosine kinase inhibitor response duration in patients with non-small cell lung cancer." J Clin Oncol **30**(4): 433-440.

Yong Wang, Jill Waters, Marco L. Leung, Anna Unruh, Whijae Roh, Xiuqing Shi, Ken Chen, Paul Scheet, Selina Vattathil, Han Liang, Asha Multani, Hong Zhang, Rui Zhao, Franziska Michor, Funda Meric-Bernstam and Nicholas E. Navin (2014). "Clonal evolution in breast cancer revealed by single nucleus genome sequencing." Nature **512**: 155-160.

Ephrem LH Chin, Cristina da Silva and Madhuri Hegde (2013). "Assessment of clinical analytical sensitivity and specificity of next-generation sequencing for detection of simple and complex mutations." BMC Genetics **14**(1): 6.

Katja Tuononen, Satu Mäki-Nevala, Virinder Kaur Sarhadi, Aino Wirtanen, Mikko Rönty, Kaisa Salonen, Jenny M. Andrews, Aino Telega, Kaerle Sari-Harjula, Sofia Lagström, Pekka Ellonen, Sanna-Maria Järnäs and Niko Knuutila (2013). "A Comparison of Targeted Next-Generation Sequencing (NGS) and Real-Time PCR in the Detection of EGFR, KRAS, and BRAF Mutations on Formalin-Fixed, Paraffin-Embedded Tumor Material of Non-Small Cell Lung Carcinoma—Superiority of NGS." Genes, Chromosomes and Cancer **52**(5): 503-511.

Peter J. Campbell, Erin D. Pleasance, Philip J. Stephens, Ed Dicks, Richard Rance, Ian Goodhead, George A. Follows, Anthony R. Green, P. Andy Futreal and Michael R. Stratton (2008). "Subclonal

phylogenetic structures in cancer revealed by ultra-deep sequencing." Proc Natl Acad Sci U S A. **105**(35): 13081-13086.

Francis S. Collins and Margaret A. Hamburg (2013). "First FDA Authorization for Next-Generation Sequencer." The New England Journal of Medicine **369**: 2369-2371.

Suzanne Clancy (2014). "Thermo Fisher Scientific Announces Listing of the Ion PGM Dx System with the U.S. FDA as Class II Medical Device." Retrieved 2/27/14, 2015, from <http://news.thermofisher.com/press-release/life-technologies/thermo-fisher-scientific-announces-listing-ion-pgm-dx-system-us-fda->.

Mary Budagyan and Phoebe Leong (2013, Oct 16, 2014). "Torrent Variant Caller 4.0." from <http://ioncommunity.lifetechnologies.com/docs/DOC-7426>.

P Cingolani, A Platts, L Wang le, M Coon, T Nguyen, L Wang, S J Land, X Lu and D M Ruden (2012). "A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3." Fly **6**(2): 80-92.

P Cingolani, V M Patel, M Coon, T Nguyen, S J Land, D M Ruden and X Lu (2012). "Using *Drosophila melanogaster* as a model for genotoxic chemical mutational studies with a new program, SnpSift." Front Genet **35**.

The 1000 Genomes Project Consortium (2012). "An integrated map of genetic variation from 1,092 human genomes." Nature **491**: 56-65.

A McKenna, M Hanna, E Banks, A Sivachenko, K Cibulskis, A Kernytsky, K Garimella, D Altshuler, S Gabriel, M Daly and M A DePristo (2010). "The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data " Genome Research **20**: 1297-1303.

Hanbo Chen and Paul C Boutros (2011). "VennDiagram: a package for the generation of highly-customizable Venn and Euler diagrams in R " BMC Bioinformatics **12**: 35.

H. Wickham (2009). ggplot2: elegant graphics for data analysis, Springer Science and Business Media.

Justin M Zook, Brad Chapman, Jason Wang, David Mittelman, Oliver Hofmann, Winston Hide and Marc Salit (2014). "Integrating human sequence data sets provides a resource of benchmark SNP and indel genotype calls." Nature Biotechnology **32**: 246–251.

Bendix Carstensen, Martyn Plummer, Esa Laara and Michael Hills (2014). "Epi: A Package for Statistical Analysis in Epidemiology." R Package version 1.1.67. from <http://CRAN.R-project.org/package=Epi>.

R Core Team (2014). "R: A Language and Environment for Statistical Computing." 2014, from <http://www.R-project.org>.

Kimberly Robasky, Nathan E. Lewis and George M. Church (2014). "The role of replicates for error mitigation in next-generation sequencing." Nature Reviews Genetics **15**: 56-62.

Michael A Quail, Miriam Smith, Paul Coupland, Thomas D Otto, Simon R Harris, Thomas R Connor, Anna Bertoni, Harold P Swerdlow and Yong Gu (2012). "A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers." BMC Genomics **13**: 341.

E J Fox, K S Reid-Bayliss, M J Emond and L A Loeb (2014). "Accuracy of Next Generation Sequencing Platforms." Next Generat Sequenc & Applic **1**(106): e1000106.

Sebastian Jünemann, Fritz Joachim Sedlazeck, Karola Prior, Andreas Albersmeier, Uwe John, Jörn Kalinowski, Alexander Mellmann, Alexander Goesmann, Arndt von Haeseler, Jens Stoye and Dag Harmsen (2013). "Updating benchtop sequencing performance comparison." Nature Biotechnology **31**(294-296).

O Harismendy, L Bao R B Schwab, S Rozenzhak J Olson, S K Kotsopoulos, S Pond, B Crain, M S Chee, K Messer, D R Link and K A Frazer (2011). "Detection of low prevalence somatic mutations in solid tumors with ultra-deep targeted sequencing." Genome Biology **12**(12): R124.

W Qin, PKozlowski, B E Taillon, P Bouffard, A J Holmes, P Janne, S Camposano, E Thiele, D Franz and D J Kwiatkowski (2010). "Ultra deep sequencing detects a low rate of mosaic mutations in tuberous sclerosis complex " Human Genetics **127**(5): 573-582.

Nicholas McGranahan and Charles Swanton (2015). "Biological and Therapeutic Impact of Intratumor Heterogeneity in Cancer Evolution." Cancer Cell **27**(1): 15-26.

David H. Spencer, Jennifer K. Sehn, Haley J. Abel, Mark A. Watson, John D. Pfeifer and Eric J. Duncavage (2013). "Comparison of Clinical Targeted Next-Generation Sequence Data from Formalin-Fixed and Fresh-Frozen Tissue Specimens." The Journal of Molecular Diagnostics **15**(5): 623-633.

Crispin Hiley, Elza C de Bruin, Nicholas McGranahan and Charles Swanton (2014). "Deciphering intratumor heterogeneity and temporal acquisition of driver events to refine precision medicine." Genome Biology **15**: 453.

Nanyue Chen, Seetharaman Balasenthil, Jacquelyn Reuther and Ann McNeill Killary (2014). "DEAR1, a Novel Tumor Suppressor That Regulates Cell Polarity and Epithelial Plasticity." Cancer Res **74**: 5683.

American Cancer Society (2013). Breast Cancer Facts and Figures 2013-2014. Atlanta, American Cancer Society, Inc.

Kalliopi P. Siziopikou (2013). "Ductal Carcinoma In Situ of the Breast: Current Concepts and Future Directions." Archives of Pathology and Laboratory Medicine **137**(4): 462-466.

Catherine F. Cowell, Britta Weigelta, Rita A. Sakrb, Charlotte K.Y. Nga, James Hicksc, Tari A. Kingb and Jorge S. Reis-Filho (2013). "Progression from ductal carcinoma in situ to invasive breast cancer: Revisited." Molecular Oncology **7**(5): 859-869.

K Kerlikowske, A Molinaro, I Cha, B M Ljung, V L Ernster, K Stewart, K Chew, D H Moore 2nd and F Waldman (2003). "Characteristics associated with recurrence among women with ductal carcinoma in situ treated by lumpectomy." Journal National Cancer Institute **95**(22): 1692-1702.

Irene L. Wapnir, James J. Dignam, Bernard Fisher, Eleftherios P. Mamounas, Stewart J. Anderson, Thomas B. Julian, Stephanie R. Land, Richard G. Margolese, J Sandra M. Swain, osep P. Costantino and Norman Wolmark (2011). "Long-Term Outcomes of Invasive Ipsilateral Breast Tumor Recurrences After Lumpectomy in NSABP B-17 and B-24 Randomized Clinical Trials for DCIS." Journal National Cancer Institute **103**(6): 478-488.

Wenjing Zhou, Christine Johansson, Karin Jirström, Anita Ringberg, Carl Blomqvist, Rose-Marie Amini, Marie-Louise Fjallskog and Fredrik Wärnberg (2013). "A Comparison of Tumor Biology in Primary Ductal Carcinoma In Situ Recurring as Invasive Carcinoma versus a New In Situ " International Journal of Breast Cancer **2013**: 582134.

Mathias Worni, Rachel A Greenup, Aimee M Mackey, Igor Akushevich and Shelley E. Hwang (2014). "Trends in treatment patterns and outcomes for DCIS patients: A SEER population-based analysis." J Clin Oncol **32**(5s): suppl; abstr 1007.

Edward Obedian, Diana B. Fischer and Bruce G. Haffty (2000). "Second Malignancies After Treatment of Early-Stage Breast Cancer: Lumpectomy and Radiation Therapy Versus Mastectomy." Journal of Clinical Oncology **18**(12): 2406-2412.

Annabel Goodwina, Sharon Parkerb, Davina Ghersib and Nicholas Wilcken (2009). "Post-operative radiotherapy for ductal carcinoma in situ of the breast – A systematic review of the randomised trials." The Breast **18**(3): 143-149.

M S Moran MS, H X Bai, B G Haffty, E E Harris, D W Arthur, L Bailey, L Carey J R Bellon, M Y Halyard S Goyal, K C Horst, S M MacDonald and Expert Panel on Radiation Oncology-Breast (2011). ACR Appropriateness Criteria® ductal carcinoma in situ. Reston, VA, American College of Radiology: 14.

E Rakovitch, S Nofech-Mozes, S A Narod, W Hanna, D Thiruchelvam, R Saskin, C Taylor, A Tuck, S Sengupta, L Elavathil, P A Jani, S J Done, N Miller, B Youngson, I Kong and L Paszat (2013). "Can we select individuals with low risk ductal carcinoma in situ (DCIS)? A population-based outcomes analysis." Breast Cancer Res Treat 138(2): 581-590.

B L Van Leeuwen, K M Rosenkranz, L L Feng, I Bedrosian, K Hartmann, K K Hunt, H M Kuerer, M Ross, S E Singletary and G V Babiera (2011). "The effect of under-treatment of breast cancer in women 80 years of age and older." Crit Rev Oncol Hematol 79(3): 315-320.

E S Hwang, S DeVries, K L Chew, D H Moore 2nd, K Kerlikowske, A Thor, B M Ljung and F M Waldman (2004). "Patterns of Chromosomal Alterations in Breast Ductal Carcinoma In situ." Clin Cancer Res 10(15): 5160-5167.

C Giardina, G Serio, G Lepore, T Lettini, A M Dalena, A Pennella, G D'Eredità, T Valente and R Ricco (2003). "Pure Ductal Carcinoma in situ and in situ Component of Ductal Invasive Carcinoma of the Breast. A Preliminary Morphometric Study " J Exp Clin Cancer Res 22(2): 279-288.

Hitchintan Kaur, Shihong Mao, Seema Shah, David H. Gorski, Stephen A. Krawetz, Bonnie F. Sloane and Raymond R. Mattingly (2013). "Next-Generation Sequencing: A powerful tool for the discovery of molecular markers in breast ductal carcinoma in situ." Expert Rev Mol Diagn 13(2): 151-165.

C E Johnson, K L Gorringer, E R Thompson, K Opeskin, S E Boyle, Y Wang, P Hill, G B Mann and I G Campbell (2012). "Identification of copy number alterations associated with the progression of DCIS to invasive ductal carcinoma." Breast Cancer Res Treat **133**(3): 889-898.

L Sontag and D E Axelrod (2005). "Evaluation of pathways for progression of heterogeneous breast tumors." J Theor Biol **232**(2): 179-189.

K Ozato, D M Shin, T H Chang and H C Morse 3rd (2008). "TRIM family proteins and their emerging roles in innate immunity " Nat Rev Immunol **8**(11): 849-860.

Jayanta Debnatha, Senthil K. Muthuswamy and Joan S. Brugge (2003). "Morphogenesis and oncogenesis of MCF-10A mammary epithelial acini grown in three-dimensional basement membrane cultures." Methods **30**(3): 256-268.

Samy Lamouille, Jian Xu and Rik Derynck (2014). "Molecular mechanisms of epithelial–mesenchymal transition " Nature reviews Molecular cell biology **15**: 178-196.

The Cancer Genome Atlas Network (2012). "Comprehensive molecular characterization of human colon and rectal cancer." Nature **487**: 330-337.

O D Abaan, E C Polley, S R Davis, Y J Zhu, S Bilke, R L Walker, M Pineda, Y Gindin, Y Jiang, W C Reinhold, S L Holbeck, R M Simon, J H Doroshow, Y Pommier and P S Meltzer (2013). "The Exomes

of the NCI-60 Panel: A Genomic Resource for Cancer Biology and Systems Pharmacology." Cancer Res **73**(14): 4372-4382.

Helit Cohen, Rotem Ben-Hamo, Moriah Gidoni, Ilana Yitzhaki, Renana Kozol and Sol Efroni. Alona Zilberberg (2014). "Shift in GATA3 functions, and GATA3 mutations, control progression and clinical presentation in breast cancer." Breast Cancer Res **16**(6): 464.

Joseph B. Addison, Colton Koontz, James H. Fugett, Chad J. Creighton, Dongquan Chen, Mark K. Farrugia, Renata R. Padon, Maria A. Voronkova, Sarah L. McLaughlin, Ryan H. Livengood, Chen-Chung Lin, J. Michael Ruppert, Elena N. Pugacheva and Alexey V. Ivanov (2014). "KAP1 Promotes Proliferation and Metastatic Progression of Breast Cancer Cells." Cancer Res **75**: 344.

Paraic A. Kenny, Genee Y. Lee, Connie A. Myers, Richard M. Neve, Jeremy R. Semeiks, Paul T. Spellman, Katrin Lorenz, Eva H. Lee, Mary Helen Barcellos-Hoff, Ole W. Petersen, Joe W. Gray and Mina J. Bissell (2007). "The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression." Mol Oncol. **1**(1): 84-96.

Julia YS Tsang and Gary MK Tse (2014). "Narrowing down the focus: what are the main predictive markers for ductal carcinoma in situ? ." Breast Cancer Management **3**(3): 271-279.

L J Solin, R Gray, F L Baehner, S M Butler, L L Hughes, C Yoshizawa, D B Cherbavaz, S Shak, D L Page, G W Sledge Jr, N E Davidson, J N Ingle, E A Perez, W C Wood, J A Sparano and S Badve S (2013). "Multigene Expression Assay to Predict Local Recurrence Risk for Ductal Carcinoma In Situ of the Breast " Journal National Cancer Institute **105**(10): 701-710.

Lucia Hernandez, Paul M Wilkerson, Maryou B Lambros, Adriana Campion-Flora, Daniel Nava Rodrigues, Arnaud Gauthier, Cecilia Cabral, Vidya Pawar, Alan Mackay, Roger A'Hern, Caterina Marchiò, Jose Palacios, Rachael Natrajan, Britta Weigelt and Jorge S Reis-Filho (2012). "Genomic and mutational profiling of ductal carcinomas in situ and matched adjacent invasive breast cancers reveals intra-tumour genetic heterogeneity and clonal selection " Journal of Pathology **227**(1): 42-52.

Elizabeth B. Claus, Stacey Petruzella, Ellen Matloff and Darryl Carter (2005). "Prevalence of BRCA1 and BRCA2 Mutations in Women Diagnosed With Ductal Carcinoma In Situ " Journal of the American Medical Association **293**(8): 964-969.

Xiaoyun Mao, Chuifeng Fan, Jing Wei, Fan Yao and Feng Jin (2010). "Genetic mutations and expression of p53 in non-invasive breast lesions." Molecular Medicine Reports **3**: 929-934.

Jeffery M. Klco, David H. Spencer, Christopher A. Miller, Malachi Griffith, Tamara L. Lamprecht, Michelle O'Laughlin, Catrina Fronick, Vincent Magrini, Ryan T. Demeter, Robert S. Fulton, William C. Eades, Daniel C. Link, Timothy A. Graubert, Matthew J. Walter, Elaine R. Mardis, John F. DiPersio, Richard K. Wilson and Timothy J. Ley (2014). "Functional Heterogeneity of Genetically Defined Subclones in Acute Myeloid Leukemia " Cancer Cell **25**(3): 379-392.

S Richards, N Aziz, S Bale, D Bick, S Das, J Gastier-Foster, W W Grody, M Hegde, E Lyon, E Spector, K Voelkerding and H L Rehm (2015). "Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." Genetics in Medicine.

S Nik-Zainal, L B Alexandrov, D C Wedge, P Van Loo, C D Greenman, K Raine, D Jones, J Hinton, J Marshall, L A Stebbings, A Menzies, S Martin, K Leung, L Chen, C Leroy, M Ramakrishna, R Rance, K W Lau, L J Mudie, I Varela, D J McBride, G R Bignell, S L Cooke, A Shlien, J Gamble, I Whitmore, M Maddison, P S Tarpey, H R Davies, E Papaemmanuil, P J Stephens, S McLaren, A P Butler, J W Teague, G Jönsson, J E Garber, D Silver, P Miron, A Fatima, S Boyault, A Langerød, A Tutt, J W Martens, S A Aparicio, A Borg, A V Salomon, G Thomas, A L Børresen-Dale, A L Richardson, M S Neuberger, P A Futreal, P J Campbell, M R Stratton; and Breast Cancer Working Group of the International Cancer Genome Consortium (2012). "Mutational processes molding the genomes of 21 breast cancers." *Cell* **149**(5): 979-993.

Thomas Helleday, Saeed Eshtad and Serena Nik-Zainal (2014). "Mechanisms underlying mutational signatures in human cancers." *Nature Reviews Genetics* **15**: 585-598.

P. J. Stephens, P. Tarpey, H. Davies, P. Van Loo, C. Greenman, D. C. Wedge, S. Nik Zainal, S. Martin, I. Varela, G. R. Bignell, L. R. Yates, E. Papaemmanuil, D. Beare, A. Butler, A. Cheverton, J. Gamble, J. Hinton, M. Jia, A. Jayakumar, D. Jones, C. Latimer, K. W. Lau, S. McLaren, D. J. McBride, A. Menzies, L. Mudie, K. Raine, R. Rad, M.S. Chapman, J. Teague, D. Easton, A. Langerod, Oslo Breast Cancer Consortium (OSBREAC), M. T. Lee, C. Y. Shen, B. T. Tee, B. W. Huimin, A. Broeks, A. C. Vargas, G. Turashvili, J. Martens, A. Fatima, P. Miron, S. F. Chin, G. Thomas, S. Boyault, O. Mariani, S. R Lakhani, M. van de Vijver, L. van 't Veer, J Foekens, C. Desmendt, C. Sotiriou, A. Tutt, C. Caldas, J. S. Reis-Filho, S. A. Aparicio, A. V. Salomon, A. L. Borresen-Dale, A. L. Richardson, P. J. Campbell, P. Andy Futreal and M. R. Stratton (2012). "The landscape of cancer genes and mutational processes in breast cancer." *Nature* **486**(7403): 400-404.

M G Wallis, K Clements, O Kearins, G Ball, J Macartney and G M Lawrence (2012). "The effect of DCIS grade on rate, type and time to recurrence after 15 years of follow-up of screen-detected DCIS." British Journal of Cancer **106**: 1611-1617.

Pavlos Lampropoulos, Adamantia Zizi-Sermpetzoglou, Spyros Rizosa, Alkiviadis Kostakisc, Nikolaos Nikiteasc and Athanasios G. Papavassiliou (2012). "TGF-beta signalling in colon carcinogenesis." Cancer Letters **314**(1): 1-7.

Daniel E. Newburger, Dorna Kashef-Haghighi, Ziming Weng, Raheleh Salari, Robert T. Sweeney, Alayne L. Brunner, Shirley X. Zhu, Xiangqian Guo, Sushama Varma, Megan L. Troxell, Robert B. West, Serafim Batzoglou and Arend Sidow (2013). "Genome evolution during progression to breast cancer " Genome Research **23**: 1097-1108.

Sundaresan Venkatachalam, Yu-Ping Shi, Stephen N. Jones, Hannes Vogel, Allan Bradley, Dan Pinkel and Lawrence A. Donehower (1998). "Retention of wild-type p53 in tumors from p53 heterozygous mice: reduction of p53 dosage can promote cancer formation." EMBO J **17**(16): 4657-4667.

A Balmain, J Gray and B Ponder (2003). "The genetics and genomics of cancer " Nature Genetics **33**(Suppl): 238-244.

Min-Han Tan, Jessica L. Mester, Joanne Ngeow, Lisa A. Rybicki, Mohammed S. Orloff and Charis Eng (2012). "Lifetime Cancer Risks in Individuals with Germline PTEN Mutations " Clin Cancer Res **18**: 400.

E I Palmero, M I Achatz, P Ashton-Prolla, M Olivier and P Hainaut (2010). "Tumor protein 53 mutations and inherited cancer: beyond Li-Fraumeni syndrome." Curr Opin Oncol **22**(1): 64-69.

Andrew Y. Shuen and William D. Foulkes (2011). "Inherited Mutations in Breast Cancer Genes—Risk and Response." Journal of Mammary Gland Biology and Neoplasia **16**(1): 3-15.

Graham Dellaire and David P. Bazett-Jones (2004). "PML nuclear bodies: dynamic sensors of DNA damage and cellular stress " BioEssays **26**(9): 963–977.

Esmeralda Casas, Jihoon Kim, Andrés Bendesky, Lucila Ohno-Machado, Cecily J. Wolfe and Jing Yang (2011). "Snail2 is an Essential Mediator of Twist1-Induced Epithelial Mesenchymal Transition and Metastasis." Cancer Res **71**: 245.

César Cobaleda, María Pérez-Caro, Carolina Vicente-Dueñas and Isidro Sánchez-García (2007). "Function of the Zinc-Finger Transcription Factor SNAI2 in Cancer and Development." Annual Review of Genetics **41**: 41-61.

PDQ Cancer Genetics Editorial Board (2015, 2/6/15). "Genetics of Breast and Gynecologic Cancers (PDQ®)." Retrieved 3/10/15, from <http://www.cancer.gov/cancertopics/pdq/genetics/breast-and-ovarian/HealthProfessional/page2>.

Noralane M. Lindor, David E. Goldgar, Sean V. Tavtigian, Sharon E. Plon and Fergus J. Couch (2013). "BRCA1/2 Sequence Variants of Uncertain Significance: A Primer for Providers to Assist in Discussions and in Medical Managemen." The Oncologist **18**(5): 518-524.

Fergus J. Couch, Katherine L. Nathanson and Kenneth Offit (2014). "Two Decades After BRCA: Setting Paradigms in Personalized Cancer Care and Prevention " Science **343**(6178): 1466-1470.

Susan B. Komen. from <http://www5.komen.org/BreastCancer/SubtypesofBreastCancer.html>.

K. David Voduc, Maggie C.U. Cheang, Scott Tyldesley, Karen Gelmon, Torsten O. Nielsen and Hagen Kennecke (2010). "Breast Cancer Subtypes and the Risk of Local and Regional Relapse." Journal of Clinical Oncology **28**(1684-1691).

Gloria J. Morris and Edith P. Mitchell (2008). "ACS Breast Cancer Facts & Figures 2013-2014, Higher Incidence of Aggressive Breast Cancers in African-American Women: A Review " J Natl Med Assoc. **100**(698-702).

H Wong, S Lau, T Yau, P Cheung and R J Epstein (2010). "Presence of an in situ component is associated with reduced biological aggressiveness of size-matched invasive breast cancer " Br J Cancer **102**(9): 1391-1396.

M Dieterich, F Hartwig, J Stubert, S Klöcking, G Kundt, B Stengel and B Gerber T Reimer (2014). "Accompanying DCIS in breast cancer patients with invasive ductal carcinoma is predictive of improved local recurrence-free survival." Breast **23**(4): 346-351.

Sharon Steinman, Jianmin Wang, Patricia Bourne, Qi Yang and Ping Tang Ann (2007). "Expression of Cytokeratin Markers, ER-alpha, PR, HER-2/neu, and EGFR in Pure Ductal Carcinoma In Situ (DCIS)

and DCIS with Co-existing Invasive Ductal Carcinoma (IDC) of the Breast." Clin Lab Sci **37**(2): 127-134.

Vessela N. Kristensena, Charles J. Vasked, Josie Ursini-Siegele, Peter Van Loof, Silje H. Nordgarda, Ravi Sachidanandamh, Therese Sørliea, Fredrik Wärnbergi, Vilde D. Haakensena, Åslaug Hellanda, Bjørn Naumec, Charles M. Perouk, David Hausslerd, Olga G. Troyanskayal and Anne-Lise Børresen-Dale (2012). "Integrated molecular profiles of invasive breast tumors and ductal carcinoma in situ (DCIS) reveal differential vascular and interleukin signaling." Proc Natl Acad Sci U S A. **109**(8): 2802-2807.

Magdalena A. Cichon, Amy C. Degnim, Daniel W. Visscher and Derek C. Radisky (2010). "Microenvironmental Influences that Drive Progression from Benign Breast Disease to Invasive Breast Cancer." Journal of Mammary Gland Biology and Neoplasia **15**(4): 389-397.

Vita

Jacquelyn Nicole Reuther was born in Houston, Texas on June 9, 1987, the daughter of Gloria Reuther and Randy Reuther, sibling of Mitchell Reuther. After completing her work at Memorial High School, Houston, Texas in 2005, she entered Texas A & M- Corpus Christi (TAMUCC) in Corpus Christi, Texas. She received the degree of Bachelor of Science with a major in Biology and a concentration in Molecular and Cellular Biology from TAMUCC in May, 2009. In September of 2009 she entered The University of Texas Graduate School of Biomedical Sciences at Houston.

Permanent address:

708 Bering Dr J Houston, TX. 77057