Using the CRISPR/Cas9 System to Derive an Isogenic BRCA1 Mutant (187delAG) Strain From a Wild Type BRCA1 Mammary Epithelial Strain

The Harvard community has made this article openly available. Please share how this access benefits you. Your story matters

<table>
<thead>
<tr>
<th>Citation</th>
<th>Greene-Colozzi, April. 2016. Using the CRISPR/Cas9 System to Derive an Isogenic BRCA1 Mutant (187delAG) Strain From a Wild Type BRCA1 Mammary Epithelial Strain. Master’s thesis, Harvard Extension School.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Citable link</td>
<td><a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:33797319">http://nrs.harvard.edu/urn-3:HUL.InstRepos:33797319</a></td>
</tr>
<tr>
<td>Terms of Use</td>
<td>This article was downloaded from Harvard University’s DASH repository, and is made available under the terms and conditions applicable to Other Posted Material, as set forth at <a href="http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA">http://nrs.harvard.edu/urn-3:HUL.InstRepos:dash.current.terms-of-use#LAA</a></td>
</tr>
</tbody>
</table>
Using the CRISPR/Cas9 System to Derive an Isogenic BRCA1 Mutant (187delAG) Strain from a Wild Type BRCA1 Mammary Epithelial Strain

April Greene-Colozzi

A Thesis in the Field of Biotechnology

for the Master of Liberal Arts Degree in Extension Studies

Harvard University

May 2016
Abstract

A powerful new genomic engineering technique, the Clustered Regular Interspersed Short Palindromic Repeats (CRISPR), was co-opted from the immune system of bacteria and archaea, and uses specific guide RNAs to bring the Cas-9 endonuclease to a targeted DNA sequence, which then creates double-stranded breaks in the genome where a new sequence can be inserted. CRISPR is notable particularly for the relative ease with which one can manipulate the system to generate almost any sequence of interest, and the high specificity and efficiency for its target (Chu et al., 2015).

The goal of this project was to introduce a well-studied founder mutation (two base pair deletion), (187delAG; BRCA1mut/+) into BRCA1 wild type (WT; mutation-free; BRCA1+/+) human mammary epithelial cells (HMECs), to generate isogenic WT and BRCA1mut/+ cell strains. Such isogenic strains, with no variability between the wild type and mutant strain except the mutation of interest, would provide an important tool for understanding the contribution of BRCA1 mutation towards those early changes in the cells that lead to breast and ovarian cancer in women with BRCA1 mutations (Silver & Livingston, 2012). This study demonstrated the use of several methods for screening single cell colonies for the small two base pair deletion, which precludes the use of conventional selection methods. During the process of screening, a reliable strategy was developed; however, we were not successful in isolating the right clone even after screening roughly 350 clones. Nevertheless, this study offers suggestions for future attempts to introduce mono allelic mutations of the kind described in this study. Among other things, strategies such as enhancing the high fidelity repair mechanism, blocking
the error-prone repair pathway, and employing alternate methods of transfection and cell screening are strongly recommended for future researchers.
Dedication

This manuscript is dedicated to my parents and sister
Acknowledgements

I would like to acknowledge the following people:

Dr. Shailja Pathania, for her unending patience, generosity of time, and encouragement.

Dr. Andrea Richardson, for providing the resources and allowing me the opportunity to pursue this project in her lab

Drs. JB Lazaro and Bose Kochupurakkal, for their ideas and help with strategy

Dr. Christian Colin, for donating the lentiCRISPR v2 plasmid

Ada Watson, for helping me jump through hoops so I could continue to work on this after leaving DFCI

Rachel Reed and Haohui Duan, for their patience and help

My sister, for her encouragement and expert guidance in making figures

My parents, for their constant encouragement and interest
**Table of Contents**

- Dedication ...................................................................................................................... v
- Acknowledgements ...................................................................................................... vi
- List of Tables ................................................................................................................ ix
- List of Figures ................................................................................................................ x
- I. Introduction ................................................................................................................ 1
- II. Materials and Methods ............................................................................................. 15
  - Guide RNA Design .................................................................................................... 15
  - Target Sequence Cloning ........................................................................................ 16
  - Transfection and Cloning of HMECS ....................................................................... 16
  - PCR and Sequencing .................................................................................................. 18
  - Restriction Digest - DNA .......................................................................................... 19
  - Restriction Digest - Cell Lysate ................................................................................ 20
  - AGT-Specific PCR ..................................................................................................... 20
- III. Results ....................................................................................................................... 22
  - Guide RNA and Initial Strategy ............................................................................... 22
  - Screening Strategy 1 ................................................................................................. 24
  - Screening Strategy 2 ................................................................................................. 26
  - Screening Strategy 3 ................................................................................................. 28
  - Cell Health ................................................................................................................ 33
IV. Discussion..............................................................34
References...............................................................47
List of Tables

Table 1. BRCA1 sequences ................................................................. 15
Table 2. Primers for restriction digest ............................................... 18
Table 3. AGT-specific primers ............................................................. 19
Table 4. Guide RNA sequences ............................................................ 21
Table 5. Summary table ..................................................................... 30
List of Figures

Figure 1. CRISPR mechanism.................................................................10
Figure 2. LentiCRISPRv2 schematic.......................................................11
Figure 3. Single Cell dilution schematic...................................................17
Figure 4. Flowchart of experimental procedure...........................................20
Figure 5. BRCA1 WT sequence with guide RNAs........................................22
Figure 6. Restriction digest sites and enzymes.............................................23
Figure 7. Restriction digest test..................................................................24
Figure 8. Screening Strategies 1 and 2 PCR and digest....................................25
Figure 9. Screening Strategy 3 on WT and BRCA1........................................26
Figure 10. Screening Strategy 3 on polyclonal cells.......................................27
Figure 11. Schematic of Screening Strategy 3..............................................28
Figure 12. Screening Strategy 3 Results.....................................................29
Chapter I
Introduction

Cancer is a complex disease, and historically, one of the least understood. The
direction of cancer research itself has evolved significantly over the last century, moving
away from “finding a cure to end cancer,” to a more individualized approach. The
classical chemotherapies that kill all rapidly dividing cells are still in use, but clinicians
now look more to a patient’s unique medical history and tumor characteristics for
treatment options. The last twenty-five years have seen huge advances that not only
answer, “how do we treat it,” but also “how did it start?” Combined with advances in
genetic testing and the elucidation of cancer subtypes, particularly in breast cancer, it is
clear that an understanding of the molecular basis of cancer is vital to any further
advances in the field. This is especially true for hereditary breast cancer where women
with mutation in one of the breast cancer tumor suppressor genes are highly predisposed
to breast cancer (Nik-Zainal et al., 2012; Silver & Livingston, 2012).

Breast cancer is one of the leading causes of death among Western women.
However, statistics gathered in 2013 by the National Cancer Institute (NCI) point to a
34% decrease in mortality rates from breast cancer since 1990. This can be attributed to a
number of factors, namely advances in screening technology and early detection methods
such as mammography and the emphasis on self-examination; a national focus on
improving health and combating obesity; an increased number of effective chemotherapy
and more patient-oriented, targeted drugs on the market; and finally, improvements in
and availability of mutation testing (Cope, 2013; DeSantis et al., 2013). Despite this
decrease in mortality, approximately 1 in 8 women will be diagnosed with breast cancer
at some point in their lives (DeSantis et al., 2013), and the NCI estimated more than
200,000 deaths from breast cancer in 2013.

There is now an abundant body of evidence that points to hereditary mutations as
being a root cause for at least 10% of these cancers (Cope, 2013; Couch et al., 2013;
Desantis, Bryan, & Jamal, 2014). The most well-studied are mutations in the high
penetrance genes, BRCA1 and BRCA2, though there are a plethora of high penetrance
and intermediate penetrance genes that are also strongly associated with a breast cancer
phenotype and not surprisingly, interact with the protein products of the BRCA1 and
BRCA2 genes. Examples of these are PTEN, p53, and CHEK2 mutations, all of which
play roles in cell cycle regulation (Shiovitz & Korde, 2015). The BRCA1 gene was
identified as being genetically linked to familial breast cancers in the early 1990’s by a
group at Berkeley led by Mary Claire King. This instrumental finding was not only
revolutionary for breast cancer research- it proved that there could be a link between
complex disease like cancer and a person’s genetics and thus a whole new approach to
prevention and disease management.

The BRCA1 gene is located on the q (long) arm of chromosome 17. It is a large
gene with 24 exons that generates a few different proteins, though the most commonly
studied is the protein product p220, a huge protein with 1863 amino acids that is involved
in tumor suppression (Silver & Livingston, 2012). Tumor suppressors function to inhibit
over-activation of the cell cycle, counteract the activity of oncogenes, inhibit cell growth
and proliferation, and are typically inactive (or recessive, in a familial setting) in cancer, via loss of heterozygosity (LOH), DNA methylation that silences gene promoters, or gene mutation (Weinberg, 2013). BRCA1 is involved in many complex pathways that contribute to the maintenance of cell cycle regulation. The cell cycle regulates both the timing and frequency of DNA replication; loss of this regulation may lead to unchecked cell growth, proliferation, and inhibition of apoptosis— all hallmarks of cancer cells. Many common tumor suppressor proteins like retinoblastoma protein and p53 control the cell cycle. p53 is in fact, the most commonly mutated tumor suppressor and its inactivation is necessary for the development of a cancer phenotype (Vogelstein, Sur, & Prives, 2010). Not surprisingly, it interacts closely with BRCA1. Other pathways BRCA1 is involved in include checkpoint control, repair of double-stranded DNA breaks via homologous recombination (non-error prone), repair of stalled replication forks and R-loops (Pathania et al., 2014; Hatchi et al., 2015) and chromatin remodeling (Silver & Livingston, 2012; Larsen et al., 2014).

The DNA replication fork is actually a network of proteins responsible for coordinating DNA unwinding (DNA helicase) and DNA synthesis (DNA polymerase) at this juncture (Sabatinos, 2010). These forks may be stalled by DNA damage (UV, radiation, cellular metabolites), and if not repaired by BRCA1, may collapse into double-stranded breaks. These DNA lesions result in impaired replication, termed replication stress, a hallmark of cancer (Gaillard, Garcia-Muse, & Aguilera, 2015). R-loops, a hybrid of RNA, DNA, and ssDNA that form after post-transcriptional modification, particularly in highly transcribed regions of a gene, are a normal part of the transcription process, but are strongly implicated in genomic instability. Hatchi et al., (2015) show that a
BRCA1/SETX complex is recruited to sites of R-loop DNA damage. In short, loss of BRCA1 functionality and thus homologous repair following damage, leads to alternative recombination pathways, like non-homologous end-joining (NHEJ) and base-excision repair, which are highly error-prone (Lee, Ledermann, & Kohn, 2013).

BRCA1 inheritance follows the classical autosomal dominant pattern, with most individuals born as heterozygous for the mutated gene; thus, it is applicable to the classic Knudson two-hit hypothesis for tumor suppressor genes, wherein any second “hit,” or mutation to the wild-type allele of an inherited heterozygous state would result in tumorigenesis (Petrucelli, Daly, & Feldman, 2010; Weinberg, 2013). However, there is strong evidence that the two-hit hypothesis may not hold true for the BRCA1 gene, as cells from heterozygous individuals are already defective in stalled fork repair, rendering them likely drivers of tumorigenesis. The pathway from a normal cell carrying an inherited BRCA1 mutation on a single allele to aggressive tumorigenesis is still relatively unknown, as well as why loss of BRCA1 tumor suppression is directly linked to only breast and ovarian carcinomas as opposed to other cancers, since the gene figures so importantly in tumor suppression across all tissue types (Paul & Paul, 2014).

Several populations are at a higher risk of inheriting a BRCA1 mutation; the most well studied of these groups are the Ashkenazi Jews, who have been associated with higher risk of certain diseases like Tay-Sachs disease, colon cancer, and breast cancer (Petrucelli, Daly, & Feldman, 2010). A 2014 study proposed that due to a fairly recent bottleneck effect, a small group of about 350 people are the ancestors of the 10 million Ashkenazi Jews alive today, which explains the high rate of heritable disorders. In addition, the historically common practice of intermarrying created an ideal setting for
passing on a few specific mutations (Carmi, et al., 2014). However, it is important to remember that this particular culture, though it is genetically less diverse than other Jewish ethnic populations, is extremely supportive of genetic study, and so has been the focus of much scientific research compared to other groups. The two common BRCA1 mutations found in about 2% of the Ashkenazi Jewish population are 187delAG (also commonly referred to as 185delAG) and 5385insC; these each convey a lifetime risk of between 65 and 85% of developing breast cancer before age 70, and a lifetime risk between 14-33% for developing ovarian cancer (Petrucelli, Daly, & Feldman, 2010; Aloraifi et al., 2015).

The phenotype of BRCA1 breast cancers are overwhelmingly high grade, highly invasive, and most often fall into the basal/triple negative phenotype, which generally means that these tumors do not express any hormone receptors (Badve et al., 2011; Aloraifi et al., 2015). These tumors typically metastasize to sites like the brain and lungs, and relapse is common with basal cancers within the first 5-8 years (Badve et al., 2011). Given that BRCA1 mutation driven cancer is triple negative (ER, PR and Her2 negative), targeted or hormone therapy is not effective in these cancers thus making their treatment more challenging. Platinum-based therapies have been moderately successful, particularly in the neoadjuvant (prior to surgery treatment) setting, and more recently, several PARP (Poly(ADP-ribose) polymerase) inhibitors have been the focus of many clinical trials investigating a more efficient treatment for BRCA1 breast and ovarian cancer (Nicum et al., 2014; Tangutoori, Baldwin, & Sridhar, 2015). PARP’s are tissue non-specific enzymes that are recruited to assist in DNA damage repair, primarily in cells that lose their homologous repair function, like BRCA1 deficient cells. These cells are
forced to turn to those alternative pathways of repair, which are facilitated by PARP’s. PARP inhibitors (PARPi) are small molecule that inhibit the polymerase’s mode of action, and therefore block the DNA damage repair response, leading to cell death in BRCA1 deficient tumor cells (Lee, Ledermann, & Kohn, 2013; Nicum et al., 2014; Tangutoori, Baldwin, & Sridhar, 2015). However, despite the strong sensitivity of BRCA1 tumors to PARPi’s, like any other successful drug treatments, resistance can still occur with the acquisition of secondary mutations in other related genes and pathways (Nicum et al., 2014). Thus, a keen understanding of the BRCA1 gene and its development of drug resistance are vital for the development of future targeted therapies.

Since the dawn of modern civilization, humans have unknowingly been contributing to the field of genetic engineering. The domestication of dogs, planting of crops, creation of hybrid plants and animals, and the migration patterns of cultures have all yielded an extremely diverse global flora and fauna. A simple internet search now yields a definition of genetic engineering as “the deliberate modification of the characteristics of an organism by manipulating its genetic material (Google, 2015).” There are a multitude of strategies available to scientists now with an exponential number of uses, ranging from gene therapy for human diseases to creating disease-resistant animal models, to modified cell lines to mimic human diseases (Doudna & Charpentier, 2014; Cox, Platt & Zhang, 2015). Some early historic examples of how genomic engineering has begun to crack open cancer include the revolutionary engineering of transgenic mice to express specific oncogenes in the 1980’s, and the successful regression of metastatic melanoma in patients following transduction of patient T cells to recognize the cancer cell antigens in 2006 (Morgan et al., 2006; Hanahan, Wagner,
Palmiter, 2007). In the last ten years, engineered nucleases have been revolutionary in honing the efficiency and accuracy of genome editing. Zinc-finger nucleases (ZFN) and transcription activator-like effector nucleases (TALEN) are synthetic proteins whose DNA binding domains enable them to cut DNA at specific loci (Gaj, Gersbach, & Barbas, 2013; Doudna & Charpentier, 2014). While these technologies have had moderate success in genome manipulation with a high specificity for the target sequence, they have several limitations (Doudna & Charpentier, 2014). Designing a ZFN domain that will bind sufficiently to the target sequence has proven challenging; in addition, the ZFN system can only bind to target sites every 50 or 200 base pairs (depending on the source of domain synthesis), so knock-in mutations at a particular site are not possible (FitzPatrick & McKay, 2015). The TALEN lentiviral vectors are also more toxic to cells once introduced, and are prone to rearrangements following cloning (Gaj, Gersbach, & Barbas, 2013). Both systems have shown significant off-target effects and must be custom-designed for each experiment (Gupta & Musunuru, 2014).

The latest in genomic engineering technology builds upon these previous two technologies, though its origin is ancient. The Clustered Regularly Interspersed Short Palindromic Repeat (more easily referred to as CRISPR-Cas9) system was discovered in E. coli in the late 1980’s by Japanese researchers (Ishino, Shinagawa, Makino, Amemura, & Nakamura, 1987), but its potential in genomic engineering was not realized until the mid 2000’s (Horvath & Barrangou, 2010; Doudna & Charpentier, 2014). Essentially, in its natural setting, the CRISPR-Cas9 system works as an immune memory for bacteria and archaea. Invading DNA material is recognized as foreign by the Cas complex, which then creates a new spacer “unit” for that particular foreign DNA and inserts it into the
existing CRISPR array. These arrays are transcribed by tracrRNA into small RNA molecules (crRNA) that serve to guide the Cas enzyme to invasive DNA that has been recognized for degradation (Horvath & Barrangou, 2010; Ran et al., 2013; Ma, Zhang, Huang, 2014). Once Cas9 has induced double-stranded DNA breaks (DSB), recombination is stimulated in the form of either the error-prone NHEJ pathways or high-fidelity homology-directed repair (HDR). There are several different CRISPR-Cas9 systems that operate on the same basic principles, with slight variation in their recognition and cleavage complexes. Type II CRISPR, depicted in Figure 1 (below), which only requires the single Cas9 enzyme, is the system commonly used for genomic engineering (Horvath & Barrangou, 2010; Doudna & Charpentier, 2014). It also combines the crRNA and tracrRNA into the simple guide (sg) RNA complex (Ran, Hsu, Wright, Arwala, Scott, & Zhang, 2014).

Figure 1: CRISPR Mechanism in the laboratory (from Charpentier & Doudna, 2013)

Without exogenous DNA to function as a repair template, the NHEJ repair pathway is stimulated. This results in DNA “lesions” in the form of insertion or deletion
mutations (commonly called indels). This is the simplest use of the CRISPR system within the laboratory environment. By simply targeting and cutting a coding exon (region of DNA that contains gene sequences that are transcribed into mRNA) to induce the NHEJ repair mechanism, these indels can introduce frameshift mutations into the DNA, which may results in altered or silenced protein production or stop codons, so transcription is inhibited. In the presence of a donor repair DNA template, such as a single-stranded DNA, the HDR pathway is stimulated. This method can be used to insert small mutations into the genome (Ran et al., 2013).

In a laboratory environment, the CRISPR system has been co-opted to perform precise DNA modifications at specific loci. In 2014, the Zhang lab of the Massachusetts Institute of Technology was able to clone both the Cas9 nuclease and the sgRNA containing a specific target sequence into a lentiviral vector, the LentiCRISPRv2 (Figure 2), for ease of delivery and manipulation (Sanjana, Shalem, & Zhang, 2014). The vector also contains a puromycin resistant gene to additionally identify cells that have undergone successful transfection and now carry the CRISPR genes. Specific target sequences are designed with several considerations. A 5’NGG protospacer adjacent motif (PAM) sequences must immediately precede the 20-nucleotide sgRNA sequence. This PAM sequence, which varies depending on the type of CRISPR system, flanks the 3’ end of the DNA target site and essentially controls the DNA target search function of Cas9 in recognizing self vs. non-self (Ran et al., 2013).
CRISPR has the potential to selectively and permanently delete genes from a cell’s genome by directly targeting DNA. It is a revolutionary technique for several reasons, most notably its high efficiency and specificity. It is also a relatively simple technique that can be performed in any laboratory setting. While other methods for gene silencing exist, like siRNA and shRNA, these are more complicated since they target RNA and often do not have long lasting or permanent effects. Compared to the earlier methods of genome engineering like ZFN’s and TALEN’s, CRISPR has several advantages. As described above, with the development of the lentiCRISPRv2, the system is easy to customize with only the insertion of a ~20 nucleotide oligo sequence. The Zhang laboratory at MIT has created a simple method for identifying possible guide RNA sequences upon the input of a DNA region of interest. The Cas9 protein consistently targets the 17th or 18th base pair of a DNA target sequence, while other systems target relatively non-specifically (Ran et al., 2013).

The CRISPR system has been used successfully across a wide variety of different organisms, including plants, which could have great implications for GMO food production. In an in vivo setting, Wu et al., (2015) corrected a murine disease mutation for cataracts in later progeny by targeting the specific mutation locus in spermatogonial stem cell. The applications of CRISPR in cancer alone are seemingly endless. Findlay et
al., (2014) used the CRISPR/Cas-9 system to determine the result of every single nucleotide change in a 78 base pair BRCA1 exon, proving that the method can be used to examine the nature of mutations. Chen et al., (2014) adapted CRISPR for cell imaging purposes, and were able to visualize the movement of sgRNA targeted repeat elements in telomeres in real-time via an EGFP-tagged deactivated Cas9. Choi and Meyerson (2014) used CRISPR to generate models of several well-studied large chromosomal rearrangements that have been implicated in lung cancer. Similarly, Torres et al., (2014) have established several human cell lines with fusion genes commonly associated with acute myeloid leukemia. These groups and many others are able to investigate the function of single genes by creating isogenic cell lines, the so-called “patient in a test tube,” which can model the behavior of single mutations with a high degree of accuracy. This is a particularly important feat in this era of targeted cancer therapy. These mutational processes are best studied by using genomic and functional approaches.

Whole genome sequencing has proven a useful tool in determining cellular and molecular mechanisms of cancer, as well as finding predictors of treatment response (Nik-Zainal et al., 2012). This method is limited heavily by genetic variation between disease and “healthy” samples generated by single nucleotide polymorphisms (SNPs; variation in nucleotides between individuals at a particular locus), and a lack of genetically matched controls (Soldner, et al., 2011). The goal of this research project was to study the genomic and functional/phenotypic changes that accumulate in normal (heterozygous) cells of women with BRCA1 mutation on their way to tumorigenesis. Towards this goal, this project attempted to derive isogenic wildtype (BRCA1+/+; WT) and heterozygous (BRCA1+/−) BRCA1 cell strains to provide an important tool for a)
whole genome sequencing based approaches and b) BRCA1 functional studies to understand the contribution of BRCA1 mutation towards those early changes in the cells that lead to breast and ovarian cancer in women with BRCA1 mutations. The establishment of a reliable BRCA1 wt and heterozygous isogenic strain set would allow a detailed and direct comparison of genomic changes accumulated over time in wt vs heterozygous cells. Isogenic strain sets are critical for determining the phenotypic changes that heterozygous mutation bearing cells undergo compared to wt cells.

This work sought to take advantage of the high specificity, efficiency, and easy applicability of CRISPR/Cas9 system within human cell lines. Two general DNA repair pathways are activated by the double-stranded breaks resulting from the Cas9 cleavage: NHEJ (Non homologous end joining) and HDR (Homology directed repair). As discussed in the previous section, NHEJ is rapid but error-prone, leading to loss of nucleotides at the end of these breaks; in contrast, HDR is less efficient but highly accurate. NHEJ is more common in mammalian cells during the G1, S, and G2 phases of the cell cycles, while HDR occurs only during the end of S and the G2 phase, when DNA replication has been completed (Lin, Staahl, Alla, & Doudna, 2014). In CRISPR mediated genomic alterations, NHEJ is the repair mechanism of choice if frameshift mutations are desired; however, HDR is preferable when introducing specific nucleotide changes, as it is error free and will use the target sequence to introduce the mutation of choice. In the typical model of double-stranded break repair, exogenous/target DNA donates its homologous sequence for the repair of the break, essentially displacing the original strand (Chu et al., 2015; Morrical, 2015).
Recently, many groups have begun to focus on examining the mutational signatures of specific cancers, where different mutation inducers leave evidence of their effect via specific patterns of mutations (Alexandrov & Stratton, 2014). In their extensive profiling of 21 different breast cancers, Nik-Zainal et al., (2012a, 2012b) found that BRCA1 and BRCA2 cancers distinctly exhibit signatures indicative of a higher rate of indels. These heritable cancers show an increase in alternate methods of double-stranded break repair, such as the error prone NHEJ. It may be possible to use these mutational signatures as predictors of treatment response (Nik-Zainal et al., 2012a; 2012b). This study was carried out by comparing the tumor cells with the ‘normal’ non-tumor cells. Given that the non-tumor cells are BRCA1mut/+ (known to be haploinsufficient for stalled fork repair; Pathania et al., 2014), one wonders whether the genome of so-called ‘normal’ cells is already dotted with mutation changes. Such information will be critical in helping us understand some of the early genomic changes that occur in the cells of these mutation-bearing women. Such an analysis is not currently possible because of lack of isogenic +/+ cells to compare the heterozygous (BRCA1mut/+) cells against. Once established, these would allow investigation into whether introduction of 187delAG mutation in one BRCA1 allele exhibits a signature similar to that observed in BRCA1 tumor cells. This could indicate that this mutation is indeed directly responsible for starting a cascade of events that will eventually lead to tumorigenesis.

In their 2014 paper on BRCA1 haploinsufficiency, Pathania et al. outline assays for investigating several key phenotypic and functional aspects of a defective BRCA1 allele. These assays could be used to confirm whether the mutant isogenic strain exhibits haploinsufficiency with respect to stalled fork repair and inability to suppress fork
collapses, like the patient BRCA1mut/+ strains. The first and most obvious characterization is assessing the level of BRCA1 protein expression. Their results indicate that a haploinsufficient BRCA1 genotype should produce less protein than a wild-type BRCA1 genotype. As mentioned before, BRCA1mut/+ patient lines exhibit haploinsufficiency in at least two crucial DNA maintenance pathways—repairing stalled replication forks and suppressing replication stress.

The CRISPR system offers the opportunity for scientists to understand the effects of all kinds of mutations on the protein’s downstream pathways, gene expression, and protein translation, all of which is of vital importance in developing more effective targeted therapies and prophylactic measures against cancer and other diseases. Each new addition to the body of literature adds to the collective understanding of the immensely wide range of this deceptively simple tool. Finally, the novelty of using CRISPR to generate an isogenic line for this BRCA1 mutation cannot be overlooked, particularly because the mechanistic understanding of how mutations in this gene lead to tumorigenesis is still quite limited. There is vast potential for researchers to use this technique to produce isogenic lines of any previously characterized cancer mutation. With the advancement of sequencing technologies, the insights into tumorigenesis could be of enormous importance.
This study attempted to introduce a two base-pair mutation, 187delAG (BRCA1mut/+) into BRCA1 wild type (WT; mutation-free; BRCA1+/+) human mammary epithelial cells (HMECs), to generate isogenic WT and BRCA1mut/+ cell strains.

Guide RNA Design

A 220 nucleotide sequence of BRCA1 exon spanning the 187delAG mutation region was loaded into the CRISPR Design Tool designed by the Zhang Lab at MIT. This generated a list of thirteen potential guide sequences for the BRCA1 N-terminal region carrying 187delAG mutation (Figure 3). The WT and 187delAG sequences are shown in Table 1.

<table>
<thead>
<tr>
<th>BRCA1 WT</th>
<th>TGTCATTAATGCTATGCAGAAAATCTTTAGAGTGTCCC</th>
</tr>
</thead>
<tbody>
<tr>
<td>BRCA1 187delAG</td>
<td>TGTCATTAATGCTATGCAGAAAATCTTTAGTGTCGCC</td>
</tr>
</tbody>
</table>

In the 187delAG founder mutation, the second AG is deleted.
Target Sequence Cloning into LentiCRISPRv2

Oligos were ordered from Sigma Aldrich. ssDNA containing the BRCA1 187delAG mutant template was ordered from IDT. The oligo pairs were ligated into the LentiCRISPR v2 plasmid using the procedure detailed in Improved lentiviral vectors and genome-wide libraries for CRISPR Screening (Sanjana, Shalem, Zhang, 2014). Dr. Colin of the Livingston Lab at Dana-Farber Cancer Institute generously donated the LentiCRISPR v2 plasmid. The plasmid was digested BsmBI and gel purified using the QIAquick Gel Extraction Kit (Qiagen). Oligo pairs were annealed, and ligated to the LentiCRISPRv2 plasmid using MightyMix (Takara Bio), and then transformed into Stbl3 bacteria. Bacteria was spread onto LB+ ampicillin plates, and allowed to grow for 36 hours. Six colonies per oligo pair were selected to create a master clone plate. The sterile toothpicks used to pick up each colony were then inoculated overnight in LB+ ampicillin. After spinning down the bacterial culture, the pellet was processed using Qiagen Miniprep Kit (Qiagen) to purify DNA. The DNA was then sent for sequencing, and the correct clones (i.e. containing the original oligo sequence) were maxi-prepped and stored at 4C.

Transfection and cloning of HMECs

The single guide RNA (sgRNA) and single strand DNA (ssDNA) carrying the BRCA1 mutation were transfected into two different WT human mammary epithelial cell (HMEC) lines using a simple lipofectamine LTX (Invitrogen) transfection. These HMEC’s (CP29 and CP37) have been isolated and immortalized from human reduction mammoplasty tissues. Cells were grown in MEGM medium (Lonza) and dissociated with
0.05% Trypsin-EDTA (Gibco). Trypsin inhibitor (ThermoFisher) was used to stop dissociation. Following transfection, cells were grown to 75-90% confluency, and aliquots were frozen at -80°C for later use. A single-cell colony assay was performed in a microplate to isolate clones with the newly inserted mutation. This is a widely used dilution method to obtain only a single (attachment independent) cell per well in the microplate. Cell counts were obtained using a TC20 Automated Cell Counter (Bio-Rad). Cells were diluted via serial dilution in a 96 well plate (see Figure 3). 10 μL from the bottom row of the dilution plate were then aliquoted to each well of a new 96 well plate containing 100 μL of media, so the wells would only have approximately one cell per well. Colonies were visible after 4-6 days. Only wells that clearly had colonies originating from a single cell were marked. Once these colonies grew to confluence, cells were trypsinized and expanded to a 48 well plate. Cells were expanded to two sets of 6, 12, or 24 well plates as necessary, and cell health was monitored daily.
PCR and Sequencing

DNA was isolated from the clones using the DNeasy kit (Qiagen), and the desired sequence in these clones and in the control samples was amplified by Phusion Hi-Fidelity Polymerase (New England BioLabs) based PCR. The PCR was carried out by using primers designed by and obtained from the Pathania lab to amplify around the BRCA1 187delAG WT sequence. The amplified samples were sent to the Molecular Biology Core Facilities at Dana-Farber Cancer Institute for high-throughput DNA sequencing.
Restriction Digest Assay on DNA

For every experiment three controls were used: a negative control (water), a patient WT control, and a BRCA1 patient sample heterozygous for the 187delAG mutation. DNA from WT strain (derived from reduction mammoplasty) and 187delAG strain (derived from prophylactic mastectomy) was confirmed by sequencing. They served as controls to test the right clone. DNA was isolated from the clones using the DNeasy kit, and the desired sequence in these clones and in the control samples was amplified by Taq polymerase PCR. The PCR was carried out by using specific primers designed to insert a single nucleotide substitution in the 187delAG sequence to create a restriction enzyme site for enzymes BtsIMutI and TspRI (New England BioLabs; Table 2).

Table 2: Primers for amplification before restriction digest

<table>
<thead>
<tr>
<th>Forward Primer</th>
<th>Reverse Primer</th>
</tr>
</thead>
<tbody>
<tr>
<td>TTCGCGTTGAAGAAGTACAAAAATGTCATTAATGCTATGCGAGAAAATGTCAAG</td>
<td>TGCCCCAGTGCAGAACCAATCAAGACA</td>
</tr>
</tbody>
</table>

Primers designed to create restriction site (Sigma-Aldrich).

Restriction enzyme digest was carried out at 65 degrees for 15 minutes (TspRI) or 55 degrees for 1 hour (BtsIMutI). The digests were run on 4-12% TBE pre-cast gels at
200V. Negative control, WT and BRCA1 positive control samples were run concurrently with each test clone.

**Restriction Digest Assay on Cell Lysate**

This method involved using the cell lysate directly instead of purifying the DNA prior restriction digestion. Cells from the 96 well plates were expanded to two sets of 24 well plates; one set was lysed with DirectPCR Lysis Reagent (Viagen) by incubating the cells at 85°C for 45 minutes. Restriction digest and gels were performed as described previously.

**PCR with AGT Specific Primers**

This strategy used primers that were designed to only bind to 187delAG mutant sequences containing the AGT sequence (Table 3). Similar to the restriction digest primer design, these new primers were designed to amplify the nucleotide sequence homologous to the 187delAG mutation.

<table>
<thead>
<tr>
<th>Primer</th>
<th>187delAG Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>187_CRISPR-F-1</td>
<td>5’TGTCATTAATGCTATGCAGAAAATCTTAGGTCCC3’</td>
</tr>
<tr>
<td>187_CRISPR-F-2</td>
<td>5’TGTCATTAATGCTATGCAGAAAATCTTAGGTCCC3’</td>
</tr>
<tr>
<td>187_CRISPR-F-2</td>
<td>5’TGTCATTAATGCTATGCAGAAAATCTTAGGTCCC3’</td>
</tr>
<tr>
<td>187_CRISPR-F-2</td>
<td>5’TGTCATTAATGCTATGCAGAAAATCTTAGGTCCC3’</td>
</tr>
</tbody>
</table>

BRCA1 187delAG sequences overlaid with sequences of AGT binding primers (in pink) (Sigma Aldrich)
Each primer was tested against the BRCA1 and WT controls to check efficiency. Transfection was repeated as before with all five oligos. Polyclonal cells were split and half were pelleted for DNA isolation. The new primer was tested against these cell populations. Polyclonal populations that amplified at 340 kb were passaged and also frozen for later use. A single-cell assay colony assay was performed on the appropriate cell populations. Cells were expanded to two sets of 6, 12, or 24 well plates, and PCR was performed using Taqman polymerase. Cell health was monitored closely throughout.

Figure 4 is a flowchart depicting the general outline of the methods followed in this work. This methodology may be applied for any work seeking to use CRISPR as a tool for inserting small nucleotide changes.

Figure 4: Flowchart of experimental procedures.
Chapter III
Results

This study endeavored to isolate an isogenic 187delAG BRCA1 mutation cell line from WT human mammary epithelial cell line derived from reduction mammoplasty.

Guide RNA and Initial Strategy

The CRISPR Design Tool generated a list of thirteen potential guide sequences appropriate for the target region. To make the project feasible given the time limitation, the six oligos with the highest score matches were selected. Table 4 contains the sequences of the six guide RNA’s and their match percentage.

Table 4: Guide RNA sequences

<table>
<thead>
<tr>
<th>#</th>
<th>Match Score</th>
<th>Sequence</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>82</td>
<td>CTTGTGCTGACTTACCAGAT</td>
</tr>
<tr>
<td>2</td>
<td>79</td>
<td>AAATCTTAGAGTGTCACATCT</td>
</tr>
<tr>
<td>3</td>
<td>71</td>
<td>CAGAATTGACCTTACATCT</td>
</tr>
<tr>
<td>4</td>
<td>68</td>
<td>AGAATTGACCTTACATATA</td>
</tr>
<tr>
<td>5</td>
<td>67</td>
<td>TCTTGTGCTGACTTACCAGA</td>
</tr>
<tr>
<td>6</td>
<td>57</td>
<td>TTTGCATAGGAGATAATCAT</td>
</tr>
</tbody>
</table>

Guide RNA sequences and match score as generated by the CRISPR design tool. Highlight in Guide # 2 contains the exact BRCA1 WT target sequence.
As Guide RNA #2 contains the BRCA1 WT sequence, the hypothesis was that this particular guide RNA would be the most successful guide RNA, particularly when compared to the positions of the other guide RNA’s, as shown in Figure 5.

Figure 5: BRCA1 WT sequence with guide RNA’s

Figure 5. Section of the 220 nucleotide BRCA1 WT sequence used to generate the guide RNA’s. The location of the 187delAG mutation is indicated by the central orange square.

The single guide RNA (sgRNA) and single strand DNA (ssDNA) carrying the BRCA1 mutation were transfected into two different WT human mammary epithelial cell
(HMEC) lines using lipofectamine LTX. Following transfection, a single-cell colony assay was performed in a microplate to isolate clones with the newly inserted mutation. Initial attempts at isolating an isogenic strain containing the 187delAG mutant consisted of amplifying the BRCA1 area of interest with primers designed by the Pathania lab, and sending those amplified samples for sequencing. None of the first 49 PCR-amplified samples sent contained the 187delAG mutant. In order to maximize time and financial resources, a more efficient strategy was developed to screen samples before sending them for sequencing.

Screening Strategy 1

Screening Strategy 1 used two non-specific restriction enzymes (TspRI and BtsIMutI) with similar target cut sequences were tested (Figure 6) that would theoretically be able to bind only the BRCA1 187delAG mutation site, and not the WT sequence, and would cut a 70 bp fragment if successful. Restriction digests were carried out according to the manufacturer’s specifications for each enzyme. During preliminary experiments with control WT and BRCA1 187delAG samples, the WT samples treated with the enzymes did not show a 70bp band, but the BRCA1 187delAG control sample did show a band where predicted (Figure 7). The initial specificity of these enzymes for the 187delAG sequence made them good candidates to use for the screen to isolate a CRISPR generated 187delAG mutant line. The BtsIMutI enzyme showed some faint background noise, while the TspRI enzyme yielded a cleaner band, so it was selected as the restriction enzyme for the screening protocol.
Figure 6: Left panel: Target sequences of restriction enzymes TspRI and BtsIMutI (N= A, C, G, or T; S= C or G). Right panel: Schematic of pre-restriction digest PCR amplification with single-nucleotide substitution primer, in red; underlined is the newly created restriction enzyme target site.

<table>
<thead>
<tr>
<th>TspRI: NNCASTGNN</th>
<th>xBRCA1 +/+: TCTTAGAGTG → TCTCAGAxGTG</th>
</tr>
</thead>
<tbody>
<tr>
<td>BtsIMutI: CAGTGN</td>
<td>BRCA1 mut/+: TCTTAGTG → TCTCAGTG</td>
</tr>
</tbody>
</table>

Figure 7. Comparison of TspRI and BtsIMutI enzymes on WT and BRCA1 control samples for Screening Strategy 1, with the 1Kb+ ladder (100-12, 000 bp).
PCR was done on 144 clones, and of those, 104 showed a 340 kb band (indicating the 187delAG fragment was amplified). The restriction analysis was carried out only with those clones that amplified the correct size fragment surrounding the 187delAG target sequence (about 340 kb) (Figure 8). 20 of those that showed a 70bp band were sent for sequencing, but none of these carried the 187delAG mutation. Furthermore, control samples in these digests were often inconsistent, with the BRCA1 behaving as a WT, or the WT showing the BRCA1 mutation-specific band at 70bp.

Screening Strategy 2

Due to these inconsistencies, modifications were warranted. For Screening Strategy 2, cell lysate was used rather than DNA for the restriction digest, and various methods of PCR, primer combinations, and enzymes were experimented with to eliminate the inconsistencies we found during Strategy 1. There was success obtaining clear PCR results using the Phusion Hi-Fidelity Polymerase kit, but none of the samples demonstrated a second BRCA1 band at 70bp (Figure 8).
Screening Strategy 3

Because so much of the early work resulted in false positive or negative results, a different approach without restriction digest was used next. Four new primers were designed with sequences that would only bind to DNA carrying the 187delAG mutation (i.e. specific to the sequence AGT). These four primers were then tested against the WT and BRCA1 controls to identify the primer set that would best distinguish between WT and 187deAG mutation carrying sequence (Figure 9). Transfection was repeated as before with all five oligos. These polyclonal cells were split and half were pelleted for DNA isolation. The new primer was tested against these polyclonal cell populations, and
indicated the presence of the mutation via a clear additional band in the BRCA1 sample at 340 kb in cell populations transfected with oligos 2 and 3.

Figure 9: Screening Strategy 3 187_CRISPR-F primers

![Image of gel electrophoresis](image)

Figure 9. Comparison between WT and BRCA1 controls with four 187_CRISPR-F primers for Screening Strategy 3; the BRCA1 control with 187_CRISPR-F-1 demonstrates a strong band at 340 kb.

As shown in Figure 9, the 187_CRISPR-F-1 primer amplified DNA only from the 187delAG mutation carrying line and not from the WT line, implying that this primer was highly specific to the 187delAG mutant sequence, thus making it possible for us to distinguish between WT and 187delAG mutant DNA when screening the clones.

Sequencing would only be done as a final confirmation on clones that demonstrated the
BRCA1 band. Transfection was repeated as before with all five oligos. Polyclonal cells were split and half were pelleted for DNA isolation. The new primer was tested against these polyclonal cell populations. A different pattern in the polyclonal cells was expected from what was seen in the WT and BRCA1 controls and that was a faint band in the polyclonal cells since they are a mixed population, the majority of which is most likely WT. The resulting gel (Figure 10) indicated the presence of the mutation in cell populations transfected with oligos 2 and 3, clearly showing a strong band for the BRCA1 control, and fainter bands for the cells transfected with oligo 2 and 3.

Figure 10. PCR of 187_CRI5PR-F-1 on polyclonal cells

![Figure 10. PCR of 187_CRI5PR-F-1 on transfected populations of cell line CP29. Cells transfected with oligos 2 and 3 had bands indicative of the 187delAG mutation, as compared to the BRCA1 187delAG control](image)
This was in agreement with the initial hypothesis at the start of the project, as the guide RNA in oligo 2 contained exactly the BRCA1 WT sequence and so would be most successful at locating the target sequence. These four sets of polyclonal cells were passaged and also frozen for later use. It was postulated that after amplification with 187_CRISP-F-1, cell lysate samples containing the mutation would demonstrate a band at 340 kb, as the BRCA1 control sample did in Figure 9. Figure 11 is an ideal gel of Screening Strategy 3 results.

Figure 11: Predicted Screening Strategy 3 results

Figure 11. Schematic of desired PCR results from amplification with 187_CRISPR-F-1; 340 kb cut expected for BRCA1 sample
For Screening Strategy 3, gel columns contained two samples each to increase efficiency and output (Figure 12). If a band appeared, a repeat PCR separating the two samples was performed. In this stringent screening test no positive clones were found after screening two hundred clones. Given that this extensive screening of 200 clones did not show a single false positive clone suggests that this research, at the very least, has resulted in the design of a very stringent screening strategy that will be able to identify the right clone. Table 5 is a summary of the screening strategies used and the number of clones that were tested under each. A minimum of 200 clones was determined to be a satisfactory and feasible number to screen given time and monetary constraints and roughly 400 clones were screened by the termination of the project.
Figure 12: Screening Strategy 3 PCR performed with 187_CRISPR-F-1 on 48 clones (samples are run two per well)

Table 5: Summary table of screening strategies

<table>
<thead>
<tr>
<th>Strategy</th>
<th>Initial Sequencing</th>
<th>Strategy 1: DNA+PCR+Dig</th>
<th>Strategy 2: Lysis+PCR+Dig</th>
<th>Strategy 3: AGT Specific PCR</th>
</tr>
</thead>
<tbody>
<tr>
<td>False Positive</td>
<td>NA</td>
<td>20</td>
<td>NA</td>
<td>NA</td>
</tr>
<tr>
<td>Negative</td>
<td>49</td>
<td>84</td>
<td>51</td>
<td>200</td>
</tr>
<tr>
<td>Total</td>
<td>49</td>
<td>104</td>
<td>51</td>
<td>200</td>
</tr>
</tbody>
</table>

Table 5. Total n= 404; Screening Strategy 3, the most reliable PCR, had an n= 200.
Cell Health

During the course of maintaining the cell populations for lysis and passaging, cell health was monitored closely. The majority of single cell colonies were able to expand and be maintained normally. Several cell populations successfully formed single cell colonies in the 96-well plate setting, but had difficulty expanding at a normal growth rate after re-seeding in the larger well plates. Other cell populations often exhibited signs of adherence loss, but appeared healthy in the culture and did not exhibit the small, non-refractile morphology customary of dead cells. Because these cells were ostensibly healthy, we pelleted those cells and re-seeded them, and many were able to adhere to the new plate after this. Still other colonies were unable to attach after re-seeding to the larger well plates and exhibited an abnormal morphology and died. While these various cell responses to expansion post-transfection may be due to unintended variations in culture protocol, these different cell responses were found in each group of transfected and expanded cells, and thus may be intrinsic in each of the single cell colony populations.
The overall goal of this project was to isolate the 187delAG BRCA1 strain to generate BRCA1 187delAG +/- and mut/+ isogenic strains. Though we were not successful in getting the right clone, we were able to design a cheap and easily applicable strategy that could identify possible clones derived from a small nucleotide CRISPR alterations. This study demonstrated the use of the CRISPR-Cas9 system for introducing a two base pair deletion in primary HMECs. The size of the alteration precludes the use of conventional selection methods, so single cell colony screening was employed to isolate cells with the desired (187delAG) mutation. During the process of screening, we were able to design a stringent and reliable strategy; however, we were not successful in isolating a clone containing the desired mutation (n = ~400 clones were tested).

A crucial next step in enhancing our understanding of genetic diseases through deep DNA sequencing is developing ways to study isogenic samples wherein the disease-causing mutation is the only variable between the WT and mutant cell line. Cell lines with minimal genetic variability between them, termed isogenic, are important tools in cancer genetics. The mechanistic nature of BRCA1 tumorigenesis is still widely unknown, and currently there are no isogenic pairs of BRCA1 heterozygous and WT lines. Such isogenic strains, with no variability between the wild type and mutant strain except the mutation of interest, would provide an important tool for a) whole genome
sequencing based approaches and b) BRCA1 functional studies to understand the contribution of BRCA1 mutation towards those early changes in the cells that lead to breast and ovarian cancer in women with BRCA1 mutations. Using the various PCR protocols described in the methods and results sections, roughly 350 clones had been screened for the 187delAG mutation by the time the project was terminated due to time and financial constraints. Though this study was unsuccessful in isolating a clone with the desired 187delAG mutation, we have gained many new and interesting insights into deriving CRISPR generated isogenic WT and mutant cell lines wherein the mutation is introduced in only one allele. These insights and other valuable resources generated during this study are detailed below:

1. Many isogenic WT lines were generated, which may provide immediate use in future studies, though that was out of the timeframe for the current study. Though the generation of the isogenic WT lines is most important in the context of a mutant line, a haplotype analysis (looking at closely linked sections of DNA) using these WT lines against a BRCA1 siRNA knockout may yield information on whether the SNP distribution is altered with the loss of BRCA1; i.e., if the SNPs/mutations upon DNA damage remain the same across for both WT and siRNA knockout, or if more mutations are generated in the knockout compared to the WT.

2. We were successful in designing an effective set of primer that consistently amplified the BRCA1 187delAG mutation against and not the WT in our final screening strategy. As discussed during the initiation of this project, there were several factors that could limit the generation of the proposed stable cell lines.
The first two screening methods, using a restriction enzyme digest, was designed in such a way that the enzyme would recognize only the mutant strain due to a single-base pair insertion following PCR. The restriction enzymes used to target the specific sequence of the 187delAG mutation were not highly specific; that is, their target sequence contains several unspecified nucleotides, which opens up the possibility of multiple different target sites. This may be the reason the digest often yielded many bands and false positives. Additionally, the control samples in these digests were inconsistent, with the BRCA1 often behaving as a WT, or the WT showing the BRCA1 mutation-specific band at 70bp. Thus, we hesitate to call those samples that yielded band “false positives,” given these experimental inconsistencies; however, they were loosely regarded as such at the time. The inconsistency of this two methods and skepticism due to false positives warranted a new design, hence our final primer design specific for the AGT mutation sequence. Screening Strategy 3 is a robust assay using primers specifically designed to only bind to the mutant sequence. The controls demonstrated consistent and expected behavior throughout testing, which gave us confidence in our clone results.

3. While this project was unsuccessful in isolating both the isogenic WT and mutant lines, we have gained a better understanding of what different approaches could have helped us in isolating the clone more efficiently. The next few points will discuss how our current method may be improved. For example, the specificity of the sgRNA’s is an important consideration. While off-target effects have been mitigated in the newer versions of the CRISPR vector, there are several steps one
can take to ensure high accuracy target sequence recognition. Shortening the target sequence by 3-4 nucleotides can eliminate excess binding affinity and make the system more sensitive to mismatches (Zheng et al., 2014). The use of five different guide RNA sequences surrounding the 187delAG locus was to ensure that at least one successfully transfected line carrying the mutation. In addition, by providing ssDNA sequence as the target (insert) vector, the sgRNA won’t target the ssDNA sequence (normally targets dsDNA due to the two active sites in the Cas9 complex). Since the ssDNA sequence is essentially the same as the WT sequence but for the AG deletion, there is a chance that the sgRNA may also cleave that; however the dsDNA is WT (TAGAGT), and the ssDNA contains the mutation (TAGT), so the guide RNA (contains TAGAGT) will only recognize the dsDNA. As previously mentioned, Guide # 2 included the exact BRCA1 WT sequence for our target area, so we suspected that would be the most successful oligo, and indeed, we did see evidence of mutant cells in the polyclonal PCR done in Screening Strategy 3. This again gives us confidence that our final screening strategy would be able to identify the mutant clone from that population. Hsu et al. (2013) describe methods for minimizing off-target activity by carefully titrating the amounts of Cas9 and sgRNA introduced into the cell. They obtained the highest efficiency by decreasing the amounts of transfected DNA.

4. HMEC’s are typically more challenging cells to perform transfection on, as efficiencies tend to be low with classical methods and result in loss of function or high cell mortality (Gresch & Altrogge, 2012). This project used an enhanced transfection reagent (Lipofectamine LTX) that is generally gentler on the cells
and more efficient than typical transfection agents such as Lipofectamine 2000, 3000, or Fugene. However, it is still a form of viral transfection, so it may also be reasonable to investigate whether electroporation is a more efficient method of transfection. Electroporation, where a high intensity electric field disrupts the cell membrane briefly, allowing exogenous molecules to move easily into the cell, has been shown to have limited effects on cell physiology and a higher efficiency of transfection. Nucleofection®, a method of electroporation developed by Amaxa™ for cells that are difficult to transfect such as primary cells, transports DNA rapidly into the nucleus using cell-type specific solution and their specific combination of electrical conditions (short, high bursts of voltage), which minimizes damage to cell function and membranes; this method has been shown to be effective in HMECs as well as other primary cells (Distler et al., 2005; Iverson et al., 2005; Zeitelhofer et al., 2007; Mao et al., 2009; Woolery et al., 2015). All of these methods of transfection are very dependent on the type of cell line, confluency of cells at transfection, and health of the cells.

It may also be worthwhile for researchers who continue this project to optimize the mode of transfection with only the CRISPR plasmid initially before adding the ssDNA. There are other CRISPR plasmids available commercially that incorporate a green fluorescent protein (GFP) tag (23 alone on the Addgene catalog for mammalian CRISPR plasmids) so as to more easily assess which cells are have incorporated the transfected DNA. GFP is a commonly used reporter protein that does not interfere with any biological functioning and can be easily co-transfected along with target DNA. Its expression in post-transfection cells
indicated that those cells now carry both the gene of interest and the gene for GFP production. Bell et al. (2014) describe their next-generation sequencing methodology for identifying CRISPR-induced mutations, where sequence-specific “barcodes” are incorporated into PCR primers and amplified in the PCR product during each successive cycle, thereby identifying the clones with the desired mutation. For their studies they used FACS analysis to initially sort the cells that were successfully transfected via a GFP tag. While this strategy would not have been useful in isolating a two-base-pair insertion, using a GFP tag with the CRISPR plasmid may have been helpful for us in identifying cells that had undergone transfection successfully. Using our final stringent screening strategy with a more robust transfection efficiency assay, it should not be necessary to screen more than 200 clones to isolate a mutant line.

As mentioned in the results section briefly, the morphology of the cells once they were expanded from the 96-well plate often varied even among sets of cells that were trypsinized and re-plated within minutes of each other. While this may of course be due to handling or a decrease in viability following transfection, some of the cell responses are interesting enough to warrant some exploration. Several cell populations were unable to sustain growth following expansion and rapidly died. Since we did confirm presence of the 187delAG mutation in polyclonal cell populations that were transfected with oligos 2 and 3, perhaps these failing populations of cells are homozygous for the mutation, and quickly lose any normal cell function. Alternatively, if the guide RNA’s targeted additional sites, these cells may have other mutations we are not screening for that
may inhibit normal growth. However, the most likely explanation for these non-viable cells is transfection toxicity. Transfection reagents can affect DNA damage repair and induce an oxidative stress response (Khodthong, Ismaili, & Juckem, 2012).

The other interesting cell response were the cells that seemed to lose their adherence but retained a healthy appearance. Cell adhesion is regulated by many different transmembrane glycoproteins, including selectins, integrins, and syndecans (Theocharis et al., 2015). Each of the four syndecan proteins have been implicated in a variety of different cancers including lung, pancreatic, colon, and breast cancer (Lim, Multhaupt, & Couchman, 2015). Many recent studies have begun to investigate this relationship between breast cancer cells and abnormal proteoglycan regulation. High expression of syndecans 1 and 2 are frequently found in triple negative breast cancers and epithelial ovarian cancers, and syndecan-2 is correlated closely with metastasis (Connor et al., 2012; Ibrahim et al, 2013). Syndecan-2 acts as a co-receptor for key growth factors such as VEGF and TGFB, and promotes angiogenesis and motility (Lim, Multhaupt, & Couchman, 2015). While most of the cells with adhesion loss eventually regained this property after pelleting and re-seeding and exhibited the WT sequence upon screening, this may be an interesting avenue for future researchers to investigate.

In their work on platinum resistance in ovarian cancer, Connor et al., (2012) established possible tumor-promoting links between BRCA1 regulation, an anti-apoptotic protein (DcR3), and Syndecan-2.
5. There are two common methods used for detecting mutations and estimating mutation hit rate with CRISPR-inserted mutations, the Surveyor Assay and the T7E1 assay. Both anneal PCR amplicons of two different DNA sources to create heteroduplex DNA. The nuclease then cleaves the DNA at mismatches, and the sizes of the resulting products indicate the location of the mutation. The T7E1 assay has been shown to be simpler and more sensitive than the Surveyor assay for CRISPR mutation detection, and yields cleaner electrophoresis results (Vouillot, Thelie, Pollet, 2015). This assay would also be an easy way of assessing transfection efficiency. It is possible that using either of these two methods may have improved the efficiency of identifying populations of cells with the 187delAG mutation, by creating a heteroduplex with the patient BRCA1 strain used as a control during these experiments. The final PCR screening method did involve a somewhat similar approach, where unsorted cell lysates were amplified only if they contained an AGT sequence. However, the issue of isolating the single cells carrying the desired mutations remains even with one of these more efficient assays to identify cell populations where the mutation is present.

6. CRISPR works well with large deletions or frameshift mutations, where the gene function is disrupted or disabled. These mutations generally do not need highly specific location sites and the repair mechanism does not matter as much. These types of knockout mutations are also more conducive to a co-expression of Cas9 and GFP. A GFP reporter incorporated into a CRISPR-Cas9 system ensures easy cell-sorting and detection of transfection efficiency Because our goal was to insert
a two base-pair mutation into only one allele, thus creating a heterozygous line, our work was more challenging. If the mutation is inserted into both alleles, this results in cell death. Several papers discuss the difficulty of isolating a rare mutant cell after nuclease treatment and suggest alternate methods to alleviate the tedious screening process. Miyaoka et al. (2014), in their paper describe a method using droplet digital PCR (ddPCR) for quickly isolating single-base pair mutations post nuclease treatment in induced pluripotent stem cells (iPS), state that such small mutations occur at frequencies below 1%, and identifying single cells carrying these mutations is incredibly laborious without the use of selection markers. Their method combined a simple TaqMan PCR with the ddPCR technique. After subdividing iPS cells into 96 well plates, portions of each well were used for ddPCR analysis and the rest were frozen. This method uses a water-oil emulsion to fraction a sample into many thousand droplets, each of which is then PCR amplified with fluorophores or probes in the digital PCR system. It is possible to identify with precision those wells that contain a target sequence. Once a well has been identified to contain the desired sequence, it is possible to then enrich the remaining cells for the mutation by continuous subdivision. Their rates of mutation frequency increased from 0.023% to 6.8%. This method would work extremely well in streamlining identification of cell populations carrying the 187delAG mutation, and should be seriously considered for future work with this project. The Zhang lab CRISPR information page on the Addgene site does reference the strategy we used, where single cell colonies are generated and screened to find single-allelic mutation cells, as the majority of the cells in a
polyclonal population will only have one allele affected. Thus, with higher transfection efficiency and our stringent PCR screening strategy, we predict that screening another 200 or so clones would isolate our mutant.

7. Many recent papers in CRISPR literature discuss the challenge of increasing HDR over NHEJ when inducing specific sequence alterations. Because HDR occurs only at certain times in the cell cycle in addition to occurring concurrently with NHEJ, the frequency of HDR repair is vastly lower than NHEJ. Thus, this poses inherent difficulties in using CRISPR to insert specific sequence alterations. With the applications of CRISPR rapidly expanding, many researchers are attempting to find ways to avoid NHEJ and boost HDR. Richardson et al., (2016), in their in-depth analysis of Cas9 behavior, found that Cas9 releases the 3’ end of the non-target strand preferentially from the target strand. In the process, the non-target strand is additionally available for annealing to exogenous DNA. The group was able to increase HDR significantly (approximately 1%) by introducing ssDNA containing the sequence of the non-target strand, and using a catalytically inactive form of Cas9 (dCas9). Another group used the technique of cell cycle synchronization to deliver Cas9 and DNA templates at the optimal time in the cell cycle. They were able to demonstrate a 38% efficiency of HDR, and also determined that 60 nucleotides was the optimal length for the ssDNA (Lin, Staahl, Alla, Doudna, 2014). Mayurama et al., (2015), blocked initiation of NHEJ by inhibiting a key ligase in the pathway. They were able to increase rates of HDR in both murine cell lines and zygotes. Chu et al. (2015) report similar results using a variety of NHEJ inhibitors. The methods discussed in these papers are easily
applicable and would be intelligent additions to future work developing the 187delAG isogenic strains.

Thus, while financial and time constraints precluded the generation of the 187delAG +/+ and mut/+ isogenic strains, important conclusions can still be drawn from the work to better guide future researchers in related projects. A stringent and easy PCR screening strategy was developed that can be applied to other similar projects where small nucleotide insertions are made to a specific point in a genetic sequence. Strategies for improving CRISPR transfection and targeting efficiency such as enhancing the high fidelity repair mechanism by blocking the error-prone repair pathway and employing alternate methods of transfection like nucleofection, as well as a GFP tag and the T7E1 assay, are strongly recommended for future researchers in the continuation of this project and of other genetic manipulations with small nucleotide insertions. Given the importance of increasing HDR for specific mutation alterations cited in many recent CRISPR papers, this seems like a logical and simple starting point. Using the target oligos already generated, NHEJ inhibitors and cell cycle synchronization during electroporation would likely increase rates of HDR. Additionally, employing the Surveyor or T7E1 assays would more quickly identify populations of cells containing the desired mutations. It is entirely conceivable that future researchers easily could use the strategy developed during this project or use alterations suggested herein to isolate a BRCA1 187delAG isogenic strain from these samples, especially since the PCR on the polyclonal cell samples seems to indicate the presence of the mutant strain. The establishment of a reliable BRCA1 WT and heterozygous isogenic strain set would allow a detailed and direct comparison of genomic changes accumulated over time in WT vs
heterozygous cells. This strategy would be particularly a particularly significant contribution to the burgeoning field of CRISPR applications, as it is still challenging to manipulate precise single nucelotides.

These isogenic strain sets will also enable investigation into the phenotypic changes that the heterozygous mutation bearing cells undergo compared to wt cells. In their 2014 paper, Pathania et al. determined that the heterozygous cells from BRCA1 mutation bearing women (BRCA1 mut/) HMECs (human mammary epithelial cells) were defective in their ability to carry out stalled replication fork repair, resulting in increased fork collapse and higher genomic instability. Their results indicate that while BRCA1 mut/+ HMECs function normally in some respects, such as centrosome maintenance, regulation of cell cycle checkpoints, and homology-directed repair of double-stranded breaks, these cells are haploinsufficient for repairing stalled replication forks and exhibit signs of replication stress. These crucial BRCA1 pathways are tumor promoting when impaired. Research on whether synthesized BRCA1 cell lines are susceptible to the same damage-induced replication stress as exhibited by the patient HMEC will help to determine whether the 187delAG mutation acts like a switch that confers not only a mutant sequence, but a mutant phenotype as well.

The establishment of these isogenic lines would have immediate applications, most importantly in investigating how this BRCA1 founder mutation affects genomic mutational landscape and basic cell functions in the mutation bearing cells. Whole genome sequencing of isogenic BRCA1+/+ and mut/+ lines will certainly be able to provide answers about a) what genes are up-regulated or down-regulated as a result of the mutation and b) what kind of genomic changes are accumulated over time in the mut/+
lines compared to +/+ cells. This will most certainly shed light on the kind of genomic changes that accumulate in mut/+ cells that contribute towards tumorigenesis.

Cancer remains an incredibly complex and baffling disease. Genetic diseases like heritable breast and ovarian cancer are particularly challenging, as they exhibit no early phenotype, and the pathway from heterozygous mutant cells to disease is essentially unknown. To understand how ‘normal’ heterozygous cells transform to tumor cells, a better understanding of the genomic and functional changes that the mutation carrying cells undergo on their way to tumorigenesis is warranted.


with corresponding alteration in the expression of BRCA1. *BMC Cancer*, 12 (176)


Petrucelli, N., Daly, M., & Feldman, G. (2010.) Hereditary breast and ovarian cancer due to mutations in BRCA1 and BRCA2. Genetics in Medicine, 12 , 245-259.


http://dx.doi.org/10.1016/j.maturitas.2015.01.015.


