Hood College

Targeted Mutagenesis of Human BRCA2 via Recombineering for Functional Evaluation of Unclassified Variants

Maximilian Riedel-Topper

CHEM 499: Departmental Honors

Dana Lawrence

April 29, 2019

Targeted Mutagenesis of Human BRCA2 via Recombineering for Functional Evaluation of Unclassified Variants

Introduction

Although the survival rate is steadily increasing, breast cancer remains the second leading cause of death in female cancer patients, behind lung cancer.[1] It is estimated that around 12.4 percent of women in the United States will develop breast cancer during their lives, so it remains a prevalent health concern.[2] Susceptibility to the cancer can be hereditary, so for women with a family history of the disease, this can be an additional cause of anxiety. Genetic predisposition to breast cancer is primarily caused by mutations to the *BRCA1* and *BRCA2* genes. Both of these genes encode proteins that are essential to DNA repair, although they perform different functions.[3] It was determined by a large clinical study that carriers of harmful *BRCA1* and *BRCA2* mutations had 72 percent and 69 percent chances of developing breast cancer, respectively.[4] Mutations in these genes have also been linked with an increased risk for other cancers, most notably ovarian. Therefore, genetic testing for mutations in these genes is a reasonable option for patients concerned with their genetic predisposition for cancer. Successful identification of a pathogenic mutation can help the prognosis of patients by leading to early screening and detection or utilization of treatment such as poly(ADP-ribose) polymerase (PARP) inhibitors.[5] Additionally, the benefits of genetic testing can extend to the counseling of family members, who may be carrying the same mutation.

However, genetic testing can often reveal a mutation in the *BRCA1* or *BRCA2* genes that is not necessarily pathogenic. Nonsense mutations prematurely introduce stop codons, which cause truncated proteins to be made during translation and, in most cases, lead to a loss of function. Single-nucleotide missense mutations, which cause a change in a single amino acid,

can be either neutral or pathogenic if protein function is impeded. Although mutations may be found through genetic screening, they can frequently be missense mutations with no known impact on BRCA1 or BRCA2 function and therefore unknown pathogenicity. These mutations are reported as variants of uncertain significance (VUS), and due to their ambiguous nature, there is interest in determining their effects on human cancer susceptibility. The uncertainty most notably impacts patients, since it can cause undue stress, medical intervention and expense if it is an ultimately benign variant.[6] VUSs are also quite common, especially in *BRCA2*, with only about 8 percent of missense mutations reported in the ClinVar database having defined pathogenic or neutral status.[7] More VUSs will be reported as genetic testing for cancer becomes more prevalent and as databases are filled. However, different methods of establishing the pathogenicity of variants have been developed and will be helpful in classifying these ambiguous mutations.

Determination of pathogenicity based on family history of cancer or patient variant data has been used to categorize some mutations but is not very feasible due to privacy limiting access to such data. There also exist sequence-based computational models that predict the classification of mutations based on known pathogenic missense mutations.[8] As in-silico methods are optimized, they can be very useful in predicting the clinical impact of *BRCA1* and *BRCA2* mutations. Nonetheless, neither of these methods can definitively prove what effect a variant will have a neutral or negative effect on protein function. Functional assays, which test for the ability of VUSs to perform the same function as the wild type gene, provide more conclusive results regarding potential pathogenicity. While all functional assays rely on introducing mutant genes into *BRCA*-deficient cells before testing for any changes in function, there are several different ways in which they can be performed.[9] However, established cell lines

with inactive *BRCA1* or *BRCA2* are inherently more susceptible to genome instability, which has led to the development of other methods in order to provide more reliable results. For example, a recent publication described a high-throughput, multiplex assay that screened for the ability of HeLa cells with *BRCA1* variants to perform homology-directed DNA repair.[10] In that study, the endogenous *BRCA1* was inactivated using small interfering RNAs (siRNAs) after the variants had been introduced.

The previously mentioned functional studies also rely on the insertion and transient expression of mutant genes as complementary DNA (cDNA) in a vector. While this is an effective means of testing many VUSs, the use of cDNA does not have the ability to produce results for mutations in introns or regulatory regions. cDNA expression also relies on vector promoters instead of the endogenous promoters contained in the full gene. Experiments using the full *BRCA* genes can be more difficult to perform, especially in the case of *BRCA2*, which is over 80 kilobases long. Dr. Shyam Sharan's group has developed a technique for generating point mutations in bacterial artificial chromosomes (BACs)[11,12] for use in a mouse embryonic stem (ES) cell functional assay.[13,14] BACs are vectors that are primarily grown in *Echerichia coli* and are able to hold much larger genetic material than plasmids. This allowed for the creation of a vector containing the full *BRCA2* gene, including both coding and non-coding regions such as introns and promoters, which removes many of the limitations of cDNA. Figure 1 contains a graphic depicting the important points of the mouse ES cell-based assay.

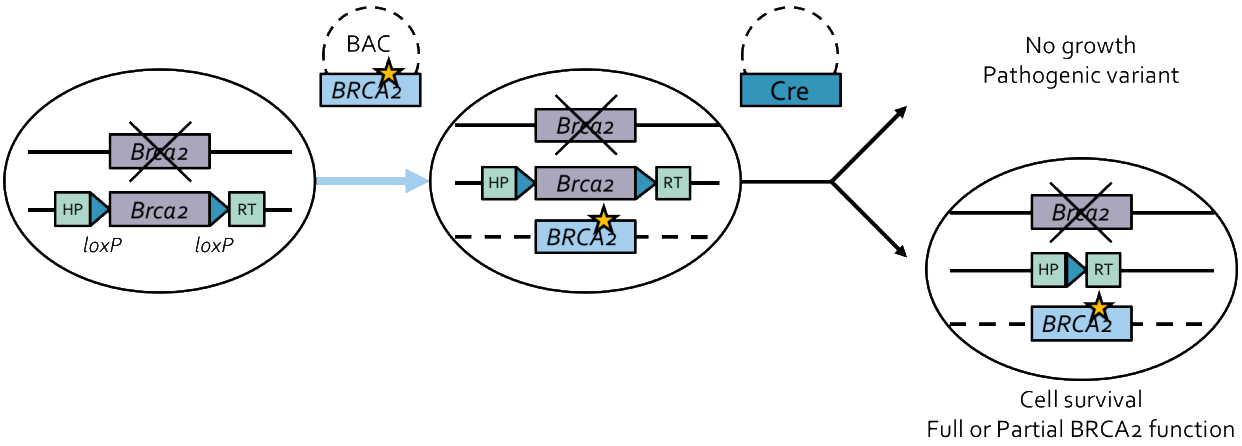


Figure 1. Schematic of the mouse embryonic stem cell-based functional assay developed by Shyam Sharan's group for evaluation of *BRCA2* variants.[14] The star in the figure represents a point mutation in the *BRCA2* gene contained in the BAC. The Cre-Lox recombination elements are shown in dark blue, and the two pieces of the *HPRT1* gene, that together confer the ability to metabolize purine, are shown in green.

The assay is based on the fact that mouse ES cells require the expression of functional BRCA2 protein to survive.[15] Mouse ES cells were designed with one inactive *Brca2* allele and one conditional *Brca2* allele, flanked by *loxP* sites, that could be removed using Cre-Lox recombination (Figure 1). In order to be able to select for successful removal of the conditional allele, two halves of the *HPRT1* mini gene were inserted on either side of the *Brca2*. Upon Cre-*loxP* recombination, the complete *HPRT1* gene allows the recombinant ES cells, which are lacking their own *Hprt* gene, to metabolize purine and thus grow in hypoxanthine-aminopterin-thymidine (HAT) media. Before the conditional allele is removed, a BAC containing the full human *BRCA2* gene with a variant of interest is transfected into the cells. Once it is established, via Western immunoblotting, that the human BRCA2 protein is expressing at levels similar to those of the endogenous mouse protein, the conditional allele is removed. The survival of the cells then indicates the functionality of the mutant BRCA2, as show in Figure 1. Variants that

result in viable ES cells are subsequently tested for sensitivity to DNA damaging agents which can distinguish between neutral or non-pathogenic variants and those that are partially functional.[14] This work focuses on the synthesis of point mutations in the BACs used in this assay.

The mutations were generated in BACs containing the entire *BRCA2* gene via recombineering. This is a method of genetic engineering that utilizes homologous recombination enzymes from bacteriophages to insert or alter DNA by replacing a region of the BAC with a DNA fragment that has ~50 base-long regions on either end that are homologous with the BAC. These fragments are most commonly generated by polymerase chain reaction and are referred to as targeting vectors due to the homologous regions they have that determine the location of recombination. Since the targeting of the BAC relies solely on homology, the recombination can be performed at any location. The actual enzymes that catalyze recombineering are introduced into the *E. coli* containing the BACs come from the bacteriophage λ Red system. The three most important of these are the Exo, Beta, and Gam proteins. Figure 2 shows a schematic of the recombineering process performed by these phage proteins.

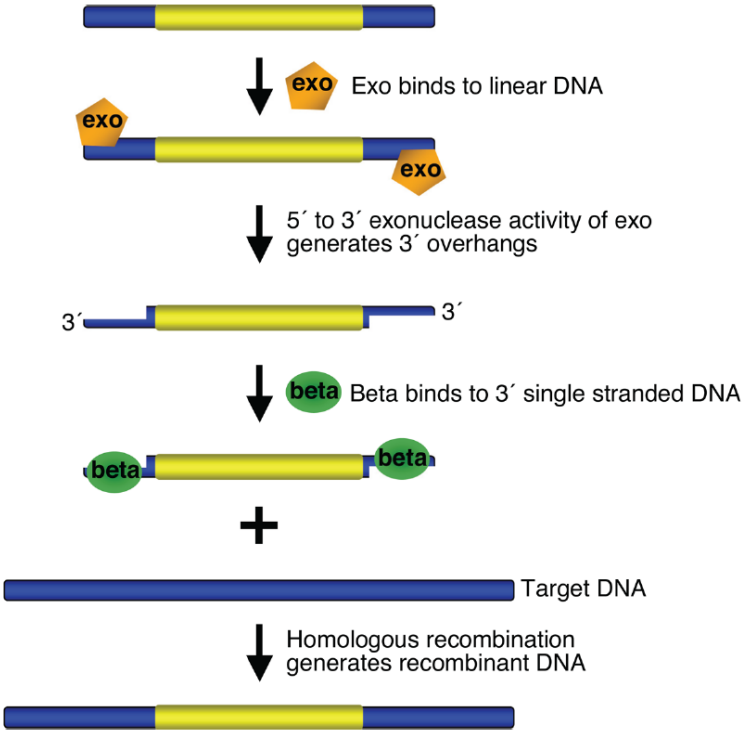


Figure 2. Schematic of the λ Red homologous recombination system used to introduce mutations into *BRCA2*. *Figure 1 in "Recombineering:a homologous recombination-based method of genetic engineering," by S. K. Sharan et al., 2009, Nature Protocols, vol. 4 no. 2, p. 206. Copyright 2009 Nature Publishing Group. Reprinted with permission from author.*

As shown in the figure above, after the targeting vector is introduced into the cell containing the λ Red system, Exo binds to the ends of the fragment. The proteins then cleave away some of the 5' end on either side of the fragment, generating 3' overhangs that will allow for recombination with the *BRCA2* gene. Beta binds the ssDNA generated by Exo, and subsequently inserts the fragment into the BAC by annealing it with homologous regions. Not shown in the figure is Gam, which facilitates this process by preventing the degradation of linear fragments by *E. coli* RecBCD nuclease. While these proteins are extremely useful in inserting the targeting vectors, prolonged activity of the λ Red proteins, especially Gam, will lead to cell death. Thus, the phage genes that encode these proteins are repressed by a temperature-sensitive repressor encoded by

the *cI857* gene. This repressor is functional at 32°C but inactive 42°C, which essentially allows expression of the recombination proteins by transiently shifting the *E. coli* culture to higher temperature.

For this work, *galK*-based seamless BAC recombineering was used to generate single-nucleotide missense mutations in exon 25 of the human *BRCA2* gene. The list of mutations was selected from the Breast Cancer Information Core (BIC, https://research.nhgri.nih.gov/bic/) database. The VarSome database and Clinvar were also consulted, as well as the International Agency for Research on Cancer (IARC) TP53 Database.[7,16,17] In addition, six different VUSs in *BRCA2* were found in patients with chordoma spinal cancer and sent to our lab for functional analysis. Therefore, these variants were also generated via BAC recombineering.

Methods

GalK-Based Seamless Recombineering Method

Recombineering was performed in the SW102 strain of *E. coli*, which is derived from the DY380 strain containing the repressed λ Red system.[18] The SW102 strain was generated by deleting the *galK* gene of the *gal* operon in DY380 cells, which is essential for *galK* positive-negative selection. Since cells containing the wild-type *BRCA2* BAC are essentially indistinguishable from the cells with *BRCA2* containing a single point mutation, this *galK*-based method is used to enable selection for pure mutant clones. The full procedure for this method is detailed in a paper by Biswas et al.[11] Figure 3 shows a diagram of how the λ Red BAC recombineering was used along with *galK* selection and counterselection to synthesize the modified *BRCA2* BACs.

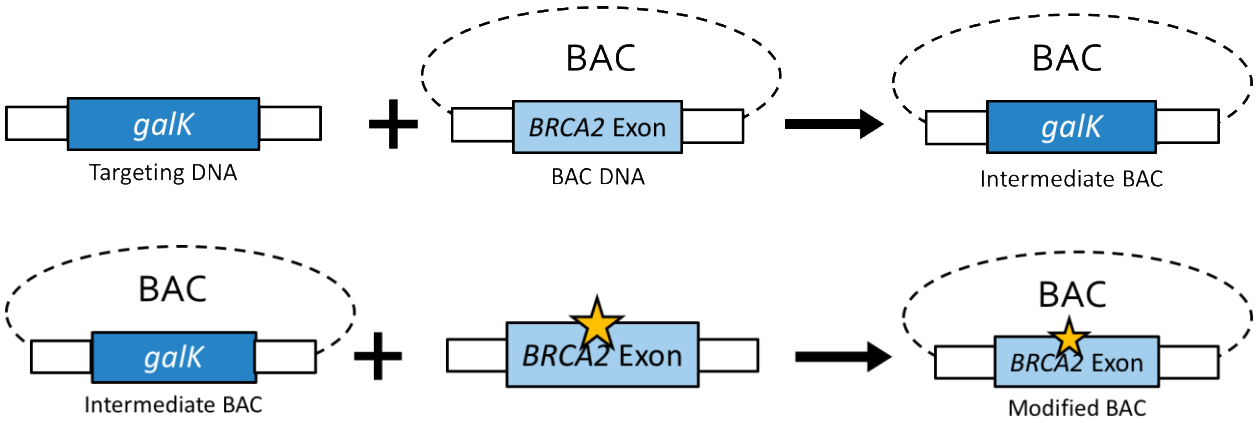


Figure 3. Schematic of recombineering steps performed in the *galK* positive/negative selection. The *BRCA2* exon with a star on it represents a targeting vector of the exon with a single base pair mutation. The white rectangles on either side of the *galK* gene insert and the targeting vector containing the mutation represent the homologous strands that target the wild-type *BRCA2* exon during recombineering.

The first part of Figure 3 shows the generation of an intermediate BAC by replacing a *BRCA2* exon with a *galK* gene cassette using homologous recombination. Targeting DNA was electroporated into *E.coli* cells, which containing the wild-type BAC, that had been immersed in a 42°C water bath for 15 minutes to induce the expression of λ Red recombinase proteins. The newly added *galK* gene restored the *galK*-negative cells' ability to metabolize galactose, thus the cells containing this intermediate BAC could be selected for by plating them on galactose minimal plates. This selected against the wild-type BACs remaining after recombineering. After the *galK* positive selection, mutated *BCRA2* exons were inserted into the BACs in place of the *galK* via recombineering. Once the electroporation was completed, the cells were grown on media to select against galactose metabolism so that only the cells containing mutant *BRCA2* would survive. The negative selection media contained deoxygalactose, which kills cells containing *galK* that attempt to metabolize it.

Prior to the work described in this paper, a library of *E. coli* clones was made with *BRCA2* BACs that had different exons replaced with galK cassettes (results not published). The eleventh exon of *BRCA2* is almost 5 kb in size. Therefore, it was divided into 10 regions, from exon 11a to exon 11j, when the library of *galK*-inserted BACs was generated. In this work, the *galK*-positive cell lines with exons 25, 11a, 11d, 11h, 14, and 15 replaced were used to generate point mutations in those exons. Polymerase chain reaction (PCR) was used to generate the targeting vectors containing the mutations. As this was a significant portion of work performed, this will be described in detail in the Results and Discussion section.

Polymerase Chain Reaction

PCR was used at several stages of the procedure, first to generate the targeting vectors, then to amplify the induced colonies to check for insertion of the vector, and ultimately to amplify the exons of interest before sequencing. All had the same standard reaction mixtures and PCR steps. All reactions used QuickTaq HS DyeMix (Toyobo) as the polymerase and forward and reverse primers at final concentrations of 20 nM. The colony PCR reactions were prepared with colonies picked into distilled water at final concentrations of 20%. The reactions were performed in T100 Thermal Cyclers (Bio-Rad) with the following steps: 95°C for 30 sec, 55°C for 30 sec, and 72°C for 1 min for 35 cycles.

The PCR products were loaded into 1% agarose gels and run in Tris-acetate EDTA buffer containing 0.00001% ethidium bromide using an electrophoresis apparatus set at 180V for 30 minutes. Purification of the products was performed using either the QIAquick PCR Purification Kit (Qiagen) for individual samples or the ZR-96 DNA Clean & Concentrator-5 (Zymo Research), and the purified samples were eluted in ultrapure distilled water.

SpeI Restriction Digestion

Before performing the restriction digestion, the *E. coli* clones containing the BACs with *BRCA2* variants were inoculated in a bacto-tryptone and yeast extract broth with 1/1000 volume of kanamycin antibiotic and grown overnight at 32°C. The parental clone containing the wild-type *BRCA2* BAC that was used to generate the mutations was used as a control. After the cells had grown, the QIAGEN Plasmid Mini Prep Kit was used to extract and purify the BAC DNA from each culture. The isolated DNA was resuspended in 30 μL of water. Digestion mixtures were set up as following: 3 μL NEBuffer 4 (New England BioLabs, Inc.), 3 μL 10x bovine serum albumin, 1 μL *Spe*I (New England BioLabs, Inc.), 3 μL water, and 20 μL BAC DNA. The DNA was digested for two hours at 37°C, before the samples were loaded into 1% agarose gels and run in Tris-acetate EDTA buffer containing 0.00001% ethidium bromide using an electrophoresis apparatus set at 30V overnight. Pictures of the gels were then taken using a UV transilluminator and camera box.

Sequencing Polymerase Chain Reaction

To prepare DNA for sequencing, the products of colony PCR using the appropriate primers was purified as described above. The sequencing PCR reaction mixtures were prepared using the BigDye kit (Thermo Fischer Scientific), 20 nM final concentration of a primer outside of the exon containing the mutation, and ~50 ng of DNA. The volume of purified DNA needed was calculated based on concentration estimations measured by a NanoDrop 1000 (Thermo Scientific) spectrometer. The reactions were performed in T100 Thermal Cyclers (Bio-Rad) with the following steps: 95°C for 30 sec, 55°C for 20 sec, and 60°C for 4 min for 26 cycles. The sequencing PCR products were purified using either Performa DTR Gel Filtration Cartridges (EdgeBio) for individual samples or Performa DTR Ultra 96-Well Plate Kit (EdgeBio), and all

products were eluted in 20 µL of ultrapure distilled water. The purified products were then sent to the Center for Cancer Research Sequencing Core Facility (CCR-SF) for sequencing.

Results and Discussion

Generation of Variant Targeting Vectors

In order to generate the *BRCA2* point mutations corresponding with the VUSs, linear DNA targeting vectors containing the mutations were made in two steps using PCR. Schematics of the two reactions are shown in Figures 5 and 6. The first step involved generating two halves of a targeting vector that were ultimately combined to extend the length of the *BRCA2* exon being replaced in recombination (Figure 4).

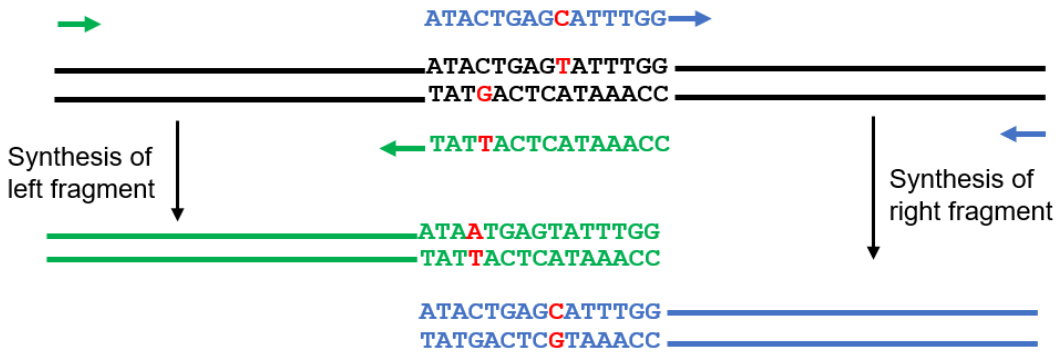


Figure 4. Schematic of the first step of polymerase chain reactions used to generate targeting vectors containing two distinct variants for recombination into *BRCA2*.

Primers were designed to be homologous with *BRCA2* at the site of the target mutation, with one single nucleotide change per primer to introduce the variant. These primers are represented in Figure 4 by the short blue and green sequences, while the black strands represent the wild type *BRCA2* gene. The bases shown in red represent the locations of the two mutations being made. The blue and green primers were utilized in two separate PCRs to generate fragments that overlapped by around 20 base pairs. Both reactions utilized primers from inside the *BRCA2* exon, paired with primers that were homologous with the intron adjacent to the exon of interest.

The wild type BAC DNA was used in the reaction as the template for the primers to anneal to initially. PCR amplification generated linear DNA fragments containing mutations within the *BRCA2* exon and extending ~50-70 bases into the intron. In Figure 4, the PCR containing the green primer pair is shown synthesizing the "left" fragment, while the PCR using the blue primer pair makes the "right" fragment. Although this example shows how two different nearby variants can be generated simultaneously, single mutations were synthesized by using overlapping primers that both contained the same nucleotide change and were thus exact complements of each other. Before the second PCR step was performed, a *DpnI* restriction digestion was used to break up the wild-type BAC DNA template in order to prevent it from being amplified. *DpnI* is an enzyme that cuts in a restriction site that occurs frequently but only when the site is methylated. Since methylation only occurs inside cells, and not on DNA fragments synthesized using PCR, the BAC DNA is broken down in to many small fragments, while the mutation vector pieces are left intact. The digestion was performed before the purification of the DNA for the second step.

The second PCR reaction used in making the targeting vectors was to combine the two pieces of each targeting vector and amplify the full strands. Figure 5 contains a schematic of this PCR.

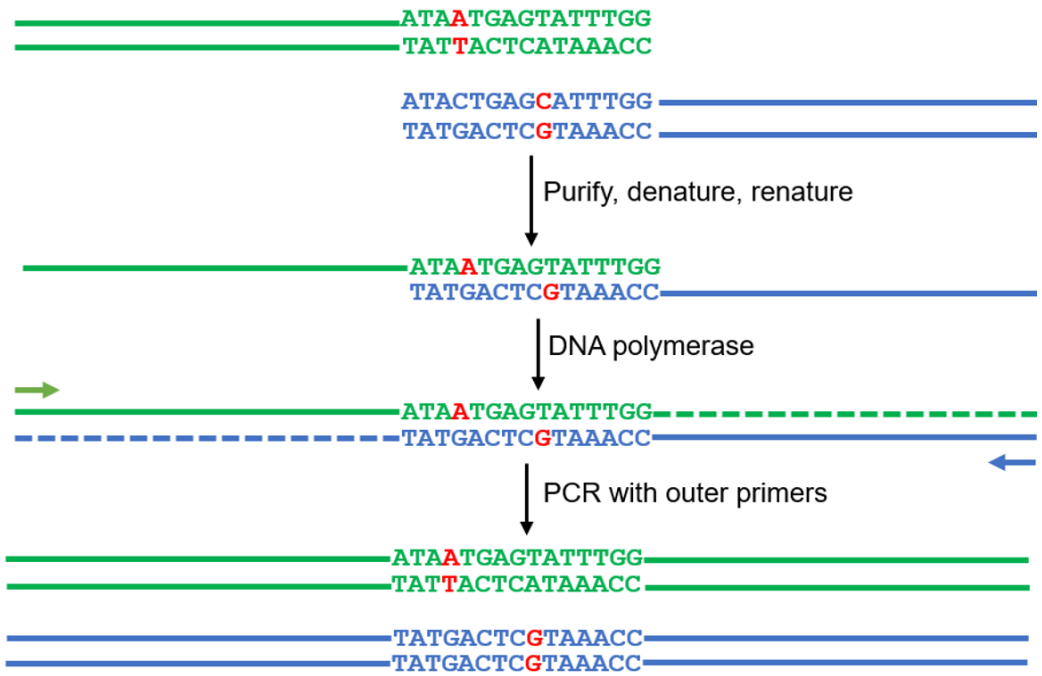
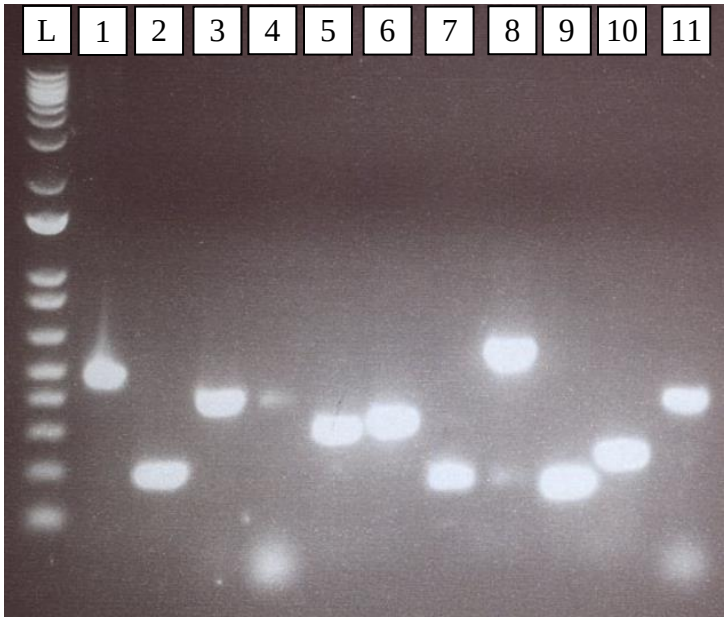


Figure 5. Schematic of the second step of polymerase chain reactions used to generate targeting vectors containing two distinct variants for recombination into *BRCA2*.

Overlapping DNA fragments containing the mutations made in the first step were purified and combined in PCR mixtures. The reactions began with denaturation and renaturation of the double stranded DNA. Once the overlapping strands annealed with each other, as shown in Figure 5, the Taq polymerase extended each strand using the other as a template. Then, the two complete strands, each containing a different mutation, were separated and subsequently amplified by the same intronic primers used in the first PCR. Although it was an intermediate product early in the reaction, double stranded DNA containing both mutations is not amplified by this reaction.

To confirm that the fragments of targeting vector were synthesized after the first step, DNA gel electrophoresis was performed on the PCR products. Figure 6 shows a picture of the agarose gel run to separate the products of the eleven PCRs corresponding to the six mutations from the chordoma patients.



Lane	Primer Pair
1	Exon 11d fwd/I1173F rev
2	Exon 11d rev/I1173F fwd
3	G715X fwd/Ex 11a rev
4	E714A rev/Ex 11 a fwd
5	Ex 11h fwd/K1690N rev
6	Ex 11h rev/K1690N fwd
7	Ex 14 fwd/T2237I rev
8	Ex 14 rev/T2237I fwd
9	S2522F fwd/Ex 15 rev
10	S2522F rev/Ex 15 fwd
11	E714A rev/Ex 11 a hit scr fwd

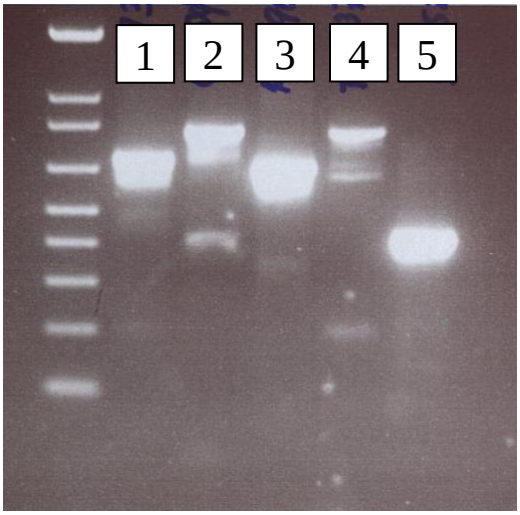
Figure 6. UV picture of agarose gel electrophoresis of the PCR products from the first reaction used to generate targeting vectors containing variants for recombination into *BRCA2*. The lanes are numbered, including one marked "L" for the 1 kb plus ladder, and the legend on the right lists the primer pairs used in each reaction.

As the primer pairs in Figure 6 indicate, each of the variants besides G715X and E714A were in different parts of the *BRCA2* gene, and thus could not be made together. For example, the S2522F mutation occurs in exon 15, so the mutation primer was paired with the corresponding external primer homologous with the region outside of the exon. Exon 11a, 11d, and 11h primers mentioned in Figure 6 all correspond with distinct regions of the eleventh exon of *BRCA2*.

The gel in Figure 6 has bright bands in almost every lane, which indicates successful PCR amplification of fragment. The bands are all under 500 to 600 bp long, which was expected. Since the wild type BAC containing *BRCA2* would not be amplified by the primers, it did not appear on the gel. Lane 4 had two weak bands, with a stronger, non-specific band near the bottom of the gel. That very short DNA fragment, along with the one indicated by the non-specific band in lane 11, is likely caused by primer dimers, which is the amplification of the

primers themselves. This can occur when the primers are not efficient at binding to the template and instead bind with other primers. Primer dimer formation hinders amplification of the target by using up PCR reagents. This can be seen in both the lane 4 and lane 11 bands, as their longer fragment bands are the weakest of all eleven reactions. However, the specific band in lane 11 was strong enough, so the fragment generated was used in the next PCR step along with the G715X fragment made in the third reaction. Lane 8 contains a very light second band, but the specific band was strong enough that the PCR product was used in the next step, unlike the PCR product from reaction mixture 4.

Agarose gel electrophoresis was also used on the products of the second PCR step to check for generation and amplification of the full targeting vectors. Figure 7 shows a picture of the gel with all five reaction mixtures used to generate the six mutations.



Lane	Variant
1	I1173F
2	G715X, E714A
3	K1690N
4	T22337I
5	S2522F

Figure 7. UV picture of agarose gel electrophoresis of the PCR products from the second reaction used to generate targeting vectors containing variants for recombination into *BRCA2*.

The lanes are numbered, and the legend on the right lists the mutation products from the first PCR used in each reaction. The furthest left lane contains the <1500 portion of 1 kb plus ladder. All five lanes had significant bands in them that indicate the combined, whole targeting vector. Lanes 1 and 3 show faint bands that are shorter than the specific bands, but lanes 2 and 4's extra bands appear brighter. These bands are most probably DNA fragments resulting from non-specific priming. The presence of non-specific bands ultimately highlights the need for gel elution of the targeting vector. Since the gel in Figure 7 shows that all five reactions generated longer DNA strands, all five were used in the gel elution. The expected lengths of the targeting vectors were determined from *BRCA2* sequence data. The bands of DNA matching the target lengths were cut out of the gel, purified, and electroporated into the *E. coli* cells for recombination.

DNA Gel Electrophoresis of Induced Cells

After the targeting vectors had been electroporated into the recombinase-induced cells, the cells were grown, and *galK* negative selection was performed. Colony PCR was then performed on the cells using the external primers corresponding to the exons containing the mutations. The purpose of this was to confirm the absence of the *galK* insert and to amplify the DNA for sequencing to confirm the presence of the correct mutation. Figure 8 displays the agarose gel of the PCR products from the colonies which the six chordoma patient variants were electroporated into.

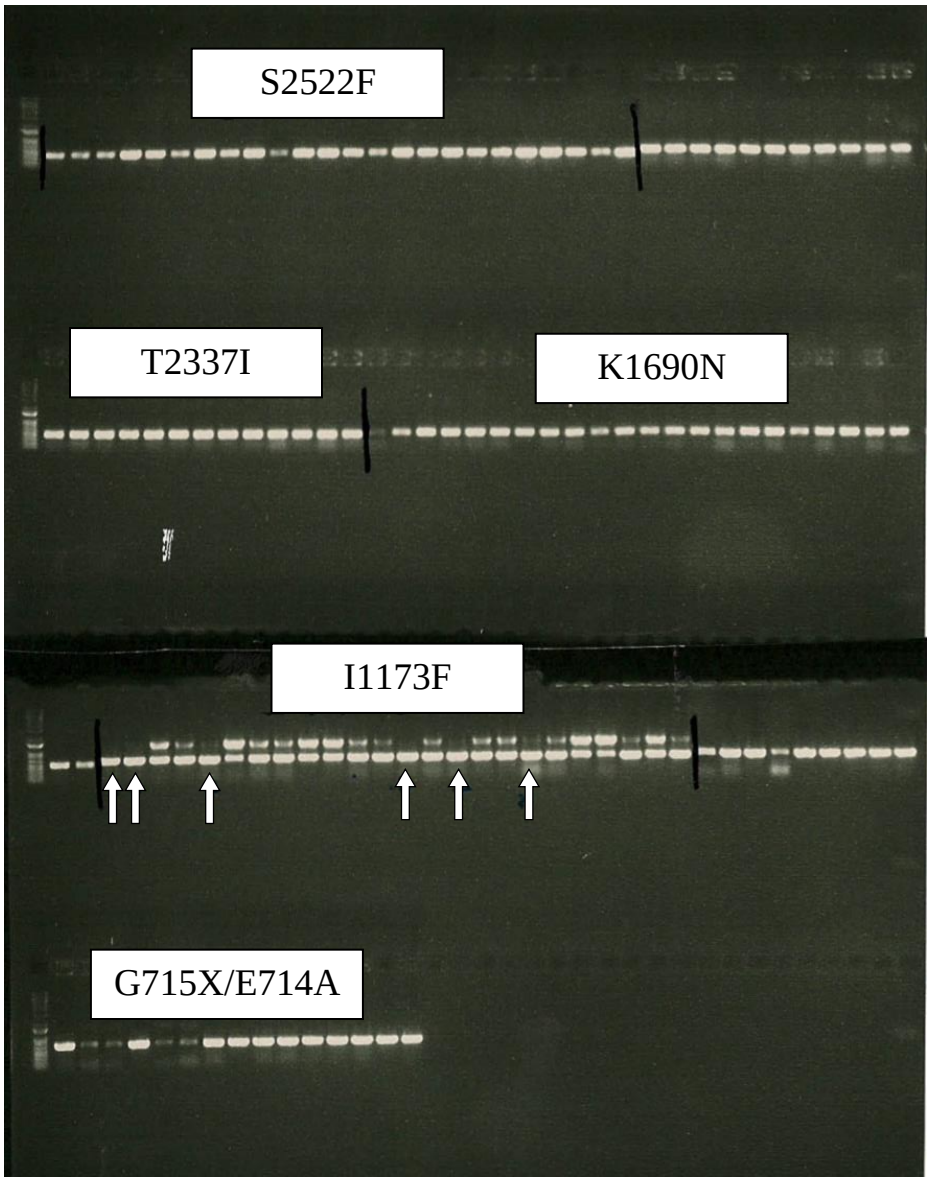


Figure 8. UV picture of agarose gel electrophoresis of the PCR amplification of electroporated cells containing *BRCA2* in BACs. Each group of 24 picked clones is separated and labeled with the variant that was introduced into the cells. Arrows indicate the I1173F cells that did not have an amplified *galK* cassette band.

Although the DNA in the gel is not separated enough to show the ladders properly, the rest of Figure 8 provides enough information to interpret the results. All of the S2522F, T2337I, and K1690N clones, along with the majority of the G715X/E714A clones, show one bright band and

one shorter, lighter band, which is likely from primer dimerization, as described earlier. The lack of a longer band above any of the bright bands indicates that the targeting vector containing the variant was successfully incorporated, replacing the longer *galK* insert. In the case of the I1173F clones, most showed the presence of both the incorporated mutation vector and the *galK* cassette. Since one cell could not contain a BAC with both in the same location, this indicates that the colony picked did not represent a pure clone. This was likely caused by how crowded the plate was with *E. coli* colonies. The crowding could also account for the presence of *galK*-positive cells growing on media chosen to negatively select against those cells. If the plate were crowded enough, cells growing on top of other cells would not necessarily metabolize the media and thus avoid selection. This can be remedied by re-growing and re-streaking plates with less of the cells in question. However, there were six clones in Figure 8, indicated by the white arrows, which were free from contamination with *galK*-positive cells. This was enough to perform the next confirmation steps needed before saving stocks of pure cultures with mutant *BRCA2* for future use. The clones confirmed to be *galK*-negative were plated again to grow more colonies for BAC restriction digestion and mutation sequencing.

SpeI Restriction Digestion

While the colony PCR helped to confirm that the target site was successfully replaced, in order to check the integrity of the entire BAC containing *BRCA2*, a restriction digestion was performed. When the λ Red phage recombinase system of the *galK*-positive cells is induced, the proteins that are expressed can make the cells more prone to DNA rearrangement. Ideally, the amount of protein expressed while the repressor was deactivated would be just enough for recombination without causing any other rearrangement. However, since there is no mechanism actively preventing this occurrence, a digestion must be performed to check the integrity of the

BAC. The *galK*-negative clones were sequenced to check for the presence of the correct point mutations, and then four clones of each mutation were grown for the restriction digestion. Mini BAC preparations (mini-preps) were performed to extract the BAC from each. The *Spe*I restriction enzyme was used to cut up the BACs and compare their fragment lengths with a *Spe*I-digested *BRCA2* BAC control. A picture of the agarose gel showing the restriction digestion results for four of BACs containing the chordoma patient variants is shown in Figure 9.

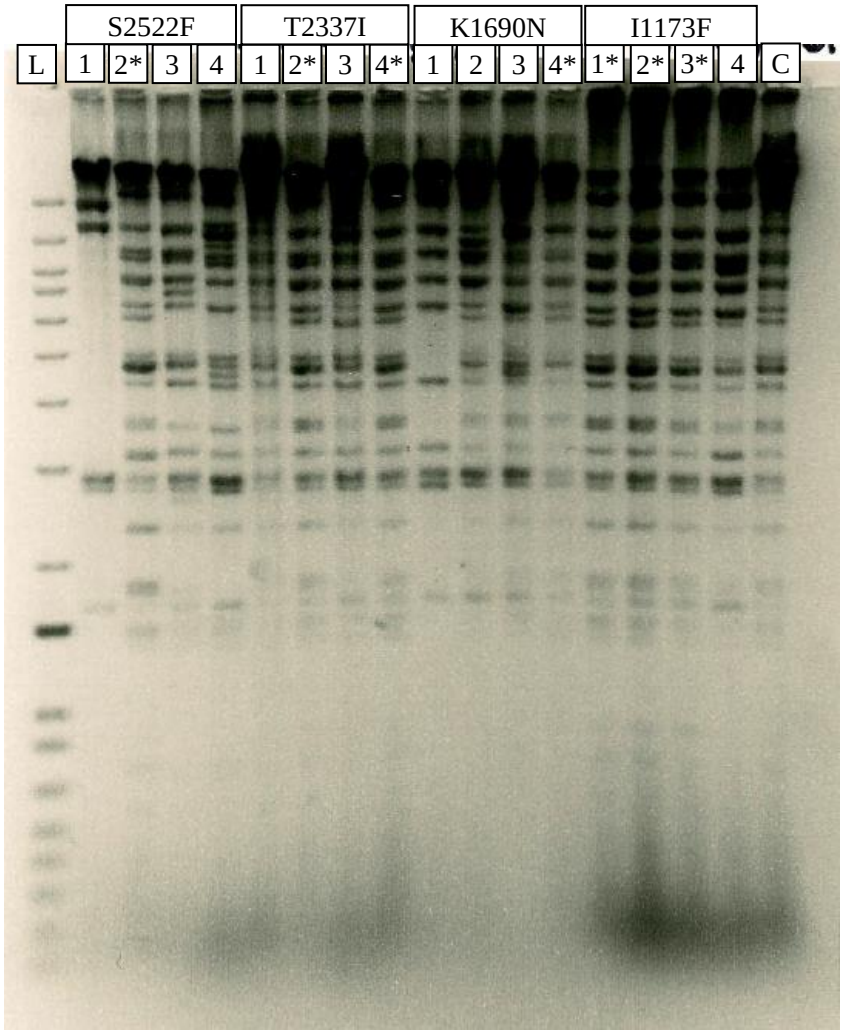


Figure 9. UV picture of the agarose gel electrophoresis of *Spe*I restriction digestions of BACs with *BRCA2* variants introduced via recombineering. All lanes are labelled, including four clones each of four different mutations. The "L" lane indicates 1 kb plus ladder, and the "C" lane

contains the restriction digestion product of a *BRCA2* control BAC. Starred numbers indicate clones whose restriction pattern matched the control.

The dark, indistinct bands at the top of the gel in Figure 9 could indicate that the digestion was incomplete. However, this lack of resolution was more likely caused by the fact that high molecular weight bands do not separate well. The bands were separated well enough to distinguish between BACs that were entire and matched the wild-type control, and ones that did not. To avoid the process of discussing every similar and dissimilar band, the clones that matched the control were starred in Figure 9. Some of the clones that did not match were clearly missing several fragments, such as the first S2522F clone and the first K1690N clone. It is likely that those BACs had fragments that were not cut during the digest, but since those clones also had bands of abnormal length compared to the control, they were dismissed regardless. The other clones that did not match had more subtle inconsistencies such as one or two extra bands, or even simply two bands that were of different lengths that those of the control. All of these irregularities indicate BACs that have been rearranged so that they no longer retain their natural restriction digestion pattern. BACs that had not been rearranged were sequenced to confirm that they still contained the mutation of interest.

Sequencing for Mutation Reconfirmation

The final step in confirming the generation of each variant was to sequence the recombined BACs and compare the exons of interest with the known sequence of *BRCA2*. Since the entire gene was far too large to sequence, primers that were a few hundred base pairs or less away from the mutations were used to perform sequencing PCRs (seqPCRs) on the BACs. The PCR products were then sent to CCR-SF to generate sequencing results. Figure 10 shows a match result generated from one piece of sequencing data entered into the National Center for

Biotechnology Information (NCBI) standard Nucleotide Basic Local Alignment Search Tool (BLAST).[19] NCBI BLAST searches a database of known human genomic and transcript DNA sequences in order to find the closest matches with a given sequence.

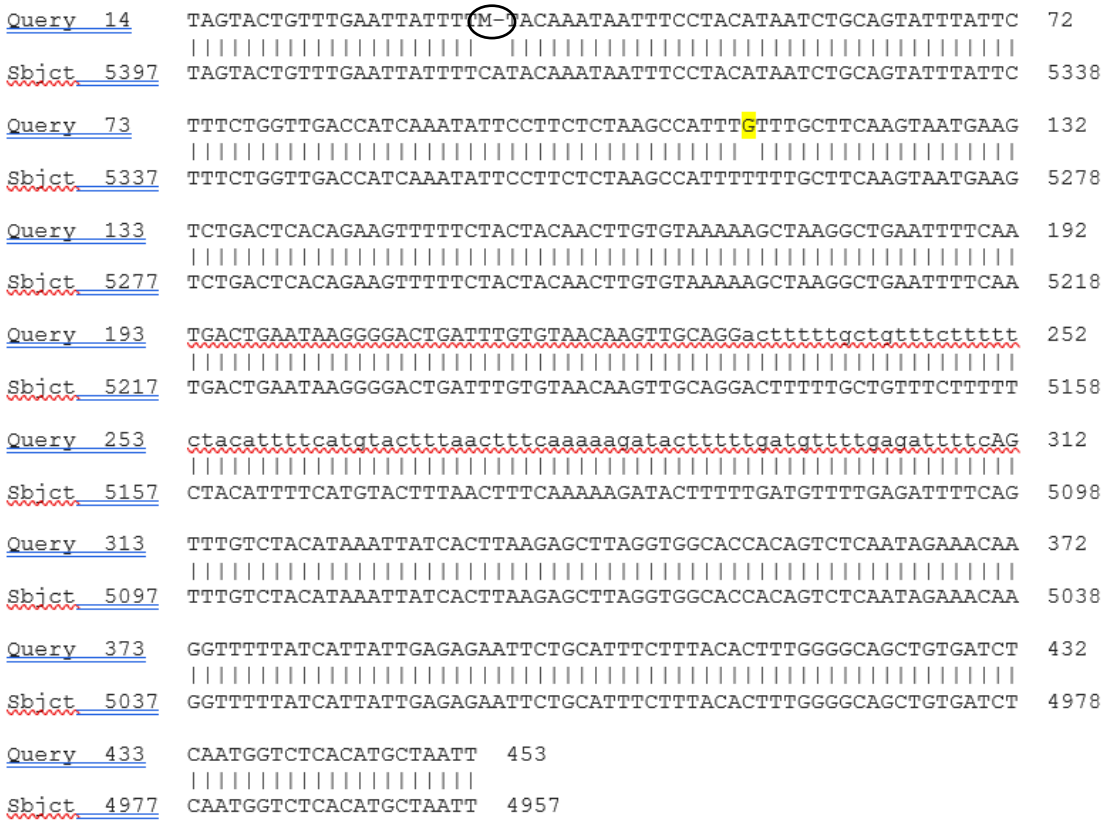


Figure 10. Sequence data of a K1690N clone matched with "Homo sapiens BRCA2, DNA repair associated (BRCA2), mRNA" sequence by the NCBI Standard Nucleotide BLAST.[19] The confirmed point mutation is highlighted in yellow, and other mismatches are circled in black.

Figure 10 shows a T>G point mutation in *BRCA2*. The nucleotide change for K1690N is recorded as an A>C mutation. This seeming discrepancy occurs because the sequencing was performed using a reverse primer, leading to the amplification of the complement strand. In Figure 8, this can be seen, since the "query" is matched with the wild-type *BRCA2* "subject" in reverse. In other words, the T>G reading from the sequence data indicates the correct mutation

that was desired. The presence of this base pair change and no other mutation confirms that the variant was successfully created in the *BRCA2* BAC. There is one other location, early in the sequence, where the amplified sample sequence does not match the known gene. However, this is an artifact caused by sequencing machine error. It has been established that the first 30-40 bases are not reliable. Thus, this inconsistency was not considered to be relevant to the integrity of the BAC. Figure 11 shows an image of part of the trace, also received from sequencing, near the location of the mutation.

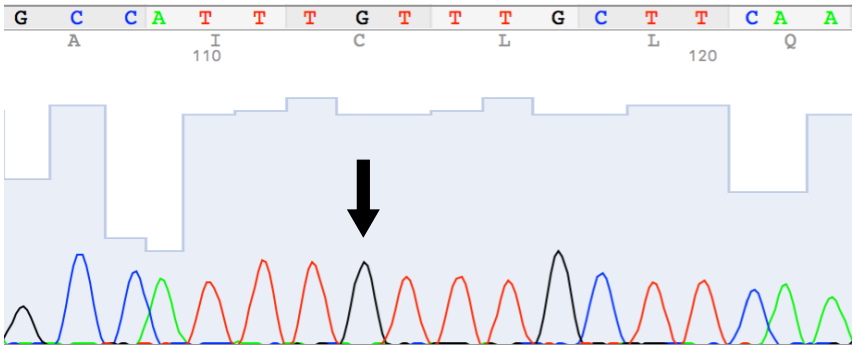


Figure 11. Graphic of sequencing trace data for clone 4 of the K1690N variant created by 4Peaks sequence trace file viewer program (Nucleobytes, Amsterdam). The T>G point mutation is indicated by a black arrow.

The sequencing machine did not fully separate the C and A bases seen on both sides of the figure. Besides this exception, the sequencing trace shows well resolved peaks with minimal noise. The base at the point of the mutation is definitively a guanine residue, confirming the synthesis of the correct mutation.

Ultimately, once the BAC was confirmed to have not undergone rearrangement during recombination and confirmed to house the *BRCA2* mutation of interest, a culture of the *E. coli* clone containing the mutant BAC was grown and a glycerol stock was saved. BAC preparations were used to extract the DNA for transfection into the mouse ES cells in preparation for

functional analysis. Table 1 is a list of all the variants that were successfully generated and confirmed by the process described in this paper.

Table 1. Summary of successfully synthesized *BRCA2* VUSs in Exon 25 and from chordoma patients

AA Change	Nucleotide Change
A3088V	c.9263C>T
Y3092C	c.9275A>G
Y3092S	c.9275A>C
Y3092X	c.9276T>G
D3095E	c.9285C>G
E3096X	c.9286G>T
Y3098X	c.9294C>G
L3101R	c.9302T>G
I3103V	c.9307A>G
I3103M	c.9309A>G
W3106X	c.9317G>A
K3115R	c.9344A>G
M3118T	c.9353T>C
L3119X	c.9356T>G
L3125H	c.9374T>A
L3125L	c.9375C>G
R3128X	c.9382C>T
P3129A	c.9385C>G
G3134V	c.9401G>T
D3142G	c.9425A>G
P3150L	c.9449C>T
E3152K	c.9454G>A
E3152G	c.9455A>G
G3153A	c.9458G>C
F3159L	c.9477C>A
K3161X	c.9481A>T
G715X	c.2143G>T
I1173F	c.3157A>T
K1690N	c.5070A>C
T2337I	c.7010C>T
S2522F	c.7565C>T

Table 1 lists the variants by their amino acid change names, which were used in this paper, along with their nucleotide change names based on the Human Genome Variation Society's recommendations.[20] There were 15 variants attempted that were not successfully confirmed by

sequencing or *Spe*I digestion. These mutations, including the E714A variant from the chordoma patient data, were not listed as they will be attempted again in the future. The lack of success in obtaining a viable E714A clone may cause doubt in the method of generating two nearby mutations at once. However, while in this particular instance, one of the mutations could not be isolated, all 26 of the successful exon 25 variants were generated using this method. Therefore, while there is some inefficiency, it is still a viable method of generating *BRCA2* variants in BACs.

Moving forward, the variants listed in Table 1 will be tested using the mouse ES cell functional assay.[13] The library of BACs with *galK* cassette insertions in place of *BRCA2* exons streamlines the processes of generating point mutations by providing BACs that are ready for selection after performing recombination with a variant-carrying targeting vector containing the entire exon originally replaced by *galK*. A series of steps has been established to confirm the integrity of the BACs and the presence of single-nucleotide mutations to the *BRCA2* contained in the BACs. This method allows for clinical variants to be generated at a request for functional analysis, as demonstrated with the chordoma patient variants. Along with the thoroughness of the ES cell functional assay, this has great potential utility in classifying the pathogenicity of *BRCA2* VUSs, which will provide valuable information for patients concerned with their susceptibility to breast cancer as well as other forms of cancer.

References

(1) *American Cancer Society: Cancer Facts and Figures 2019*; American Cancer Society: Atlanta, GA, 2019; p 8.
(2) Noone, A.; Howlander, N.; Krapcho, M.; Miller, D.; Brest, A.; Yu, M.; Ruhl, J.; Tatalovich, Z.; Mariotto, A.; Lewis, D.; et al. *SEER Cancer Statistics Review, 1975-2015*; National Cancer Institute: Bethesda, MD, 2018.
(3) Roy, R.; Chun, J.; Powell, S. N. BRCA1 and BRCA2: Different Roles in a Common Pathway of Genome Protection. *Nat. Rev. Cancer* **2011**, *12* (1), 68–78. https://doi.org/10.1038/nrc3181.

(4) Kuchenbaecker, K. B.; Hopper, J. L.; Barnes, D. R.; Phillips, K.-A.; Mooij, T. M.; Roos-Blom, M.-J.; Jervis, S.; van Leeuwen, F. E.; Milne, R. L.; Andrieu, N.; et al. Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. *JAMA* **2017**, *317* (23), 2402–2416. https://doi.org/10.1001/jama.2017.7112.

(5) Bryant, H. E.; Schultz, N.; Thomas, H. D.; Parker, K. M.; Flower, D.; Lopez, E.; Kyle, S.; Meuth, M.; Curtin, N. J.; Helleday, T. Specific Killing of BRCA2-Deficient Tumours with Inhibitors of Poly(ADP-Ribose) Polymerase. *Nature* **2005**, *434* (7035), 913–917. https://doi.org/10.1038/nature03443.

(6) Welsh, J. L.; Hoskin, T. L.; Day, C. N.; Thomas, A. S.; Cogswell, J. A.; Couch, F. J.; Boughey, J. C. Clinical Decision-Making in Patients with Variant of Uncertain Significance in BRCA1 or BRCA2 Genes. *Ann. Surg. Oncol.* **2017**, *24* (10), 3067–3072. https://doi.org/10.1245/s10434-017-5959-3.

(7) Landrum, M. J.; Lee, J. M.; Benson, M.; Brown, G. R.; Chao, C.; Chitipiralla, S.; Gu, B.; Hart, J.; Hoffman, D.; Jang, W.; et al. ClinVar: Improving Access to Variant Interpretations and Supporting Evidence. *Nucleic Acids Res.* **2018**, *46* (D1), D1062–D1067. https://doi.org/10.1093/nar/gkx1153.

(8) Hart, S. N.; Hoskin, T.; Shimelis, H.; Moore, R. M.; Feng, B.; Thomas, A.; Lindor, N. M.; Polley, E. C.; Goldgar, D. E.; Iversen, E.; et al. Comprehensive Annotation of BRCA1 and BRCA2 Missense Variants by Functionally Validated Sequence-Based Computational Prediction Models. *Genet. Med. Off. J. Am. Coll. Med. Genet.* **2019**, *21* (1), 71–80. https://doi.org/10.1038/s41436-018-0018-4.

(9) Carvalho, M. A.; Couch, F. J.; Monteiro, A. N. A. Functional Assays for BRCA1 and BRCA2. *Int. J. Biochem. Cell Biol.* **2007**, *39* (2), 298–310. https://doi.org/10.1016/j.biocel.2006.08.002.

(10) Starita, L. M.; Islam, M. M.; Banerjee, T.; Adamovich, A. I.; Gullingsrud, J.; Fields, S.; Shendure, J.; Parvin, J. D. A Multiplex Homology-Directed DNA Repair Assay Reveals the Impact of More Than 1,000 BRCA1 Missense Substitution Variants on Protein Function. *Am. J. Hum. Genet.* **2018**, *103* (4), 498–508. https://doi.org/10.1016/j.ajhg.2018.07.016.

(11) Biswas, K.; Stauffer, S.; Sharan, S. K. Using Recombineering to Generate Point Mutations:GalK-Based Positive-Negative Selection Method. *Methods Mol. Biol. Clifton NJ* **2012**, *852*, 121–131. https://doi.org/10.1007/978-1-61779-564-0_10.

(12) Sharan, S. K.; Thomason, L. C.; Kuznetsov, S. G.; Court, D. L. Recombineering: A Homologous Recombination-Based Method of Genetic Engineering. *Nat. Protoc.* **2009**, *4* (2), 206–223. https://doi.org/10.1038/nprot.2008.227.

(13) Kuznetsov, S. G.; Chang, S.; Sharan, S. K. Functional Analysis of Human BRCA2 Variants Using a Mouse Embryonic Stem Cell-Based Assay. *Methods Mol. Biol. Clifton NJ* **2010**, *653*, 259–280. https://doi.org/10.1007/978-1-60761-759-4_16.

(14) Kuznetsov, S. G.; Liu, P.; Sharan, S. K. Mouse ES-Cell-Based Functional Assay to Evaluate Mutations in BRCA2. *Nat. Med.* **2008**, *14* (8), 875–881. https://doi.org/10.1038/nm.1719.

(15) Sharan, S. K.; Morimatsu, M.; Albrecht, U.; Lim, D.-S.; Regel, E.; Dinh, C.; Sands, A.; Eichele, G.; Hasty, P.; Bradley, A. Embryonic Lethality and Radiation Hypersensitivity Mediated by Rad51 in Mice Lacking Brca2. *Nature* **1997**, *386* (6627), 804–810. https://doi.org/10.1038/386804a0.

(16) Bouaoun, L.; Sonkin, D.; Ardin, M.; Hollstein, M.; Byrnes, G.; Zavadil, J.; Olivier, M. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. *Hum. Mutat.* **2016**, *37* (9), 865–876. https://doi.org/10.1002/humu.23035.
(17) Kopanos, C.; Tsiolkas, V.; Kouris, A.; Chapple, C. E.; Albarca Aguilera, M.; Meyer, R.; Massouras, A. VarSome: The Human Genomic Variant Search Engine. *Bioinformatics*. https://doi.org/10.1093/bioinformatics/bty897.
(18) Lee, E. C.; Yu, D.; Martinez de Velasco, J.; Tessarollo, L.; Swing, D. A.; Court, D. L.; Jenkins, N. A.; Copeland, N. G. A Highly Efficient Escherichia Coli-Based Chromosome Engineering System Adapted for Recombinogenic Targeting and Subcloning of BAC DNA. *Genomics* **2001**, *73* (1), 56–65. https://doi.org/10.1006/geno.2000.6451.
(19) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. Basic Local Alignment Search Tool. *J. Mol. Biol.* **1990**, *215* (3), 403–410. https://doi.org/10.1016/S0022-2836(05)80360-2.
(20) Dunnen, J. T. den; Dalgleish, R.; Maglott, D. R.; Hart, R. K.; Greenblatt, M. S.; McGowan-Jordan, J.; Roux, A.-F.; Smith, T.; Antonarakis, S. E.; Taschner, P. E. M. HGVS Recommendations for the Description of Sequence Variants: 2016 Update. *Hum. Mutat.* **2016**, *37* (6), 564–569. https://doi.org/10.1002/humu.22981.