Andreia Ringler Chande

Licenciatura em Bioquímica

Alternative Splicing Variants in Breast Cancer

Dissertação para obtenção do Grau de Mestre em Bioquímica para a Saúde

Orientador: Professora Susana Nunes da Silva, PhD, NOVA Medical School |

Faculdade de Ciências Médicas da Universidade NOVA de Lisboa

Dezembro, 2022

Andreia Ringler Chande

Licenciatura em Bioquímica

Alternative Splicing Variants in Breast Cancer

Dissertação para obtenção do Grau de Mestre em Bioquímica para a Saúde

Orientador: Professora Susana Nunes da Silva, PhD, NOVA Medical School |

Faculdade de Ciências Médicas da Universidade NOVA de Lisboa

Dezembro, 2022

Acknowledgments

É com grande sentimento de gratidão que finalizo esta fase importante da minha vida. Esta gratidão é dirigida a todos aqueles que de alguma forma contribuíram para o meu desenvolvimento e crescimento durante este percurso. Este foi sem dúvida um percurso marcante com obstáculos, mas especialmente com momentos e ensinamentos que vou guardar pelo resto da vida.

Em primeiro lugar gostaria de dirigir um profundo agradecimento à minha orientadora, Dr.ª Susana Nunes da Silva pela oportunidade que me deu, recebendo-me de braços abertos, especialmente num período em que não esperava. Por isso e por todo o apoio que me demostrou ao longo deste período, mostrando-se sempre disponível para me ouvir, deixo-lhe um "Muito Obrigado".

Ao Professor Sebastião Rodrigues que esteve sempre disponível para qualquer necessidade que tivesse, tanto a nível académico como contribuindo com ideias para o desenvolvimento deste projeto deixo-lhe um especial agradecimento. E como não poderia deixar de ser, gostaria também de agradecer ao Dr. Bruno Gomes que contribuiu igualmente para este projecto.

Ao Professor José Rueff, por me receber no grupo ToxOmics.

Um grande obrigado às minhas colegas de laboratório iniciais, Matilde, a minha "mini chefe", Isabel, Catarina e Rita que se tornaram amigas sendo que desde o início me acolheram e criaram um ambiente digno de se trabalhar. À Joana que fez parte de todo o meu percurso no laboratório e que me fez rir muito aliviando os dias mais pesados. À Diana, a "dinossauro" do nosso laboratório como ela própria se auto-intitulou, pelos conselhos e ajuda. À Jhenifer que voltou a fazer parte do meu percurso académico, mostrando preocupação e partilhado comigo muitos momentos que guardo. À Barbara Barradas que iniciou esta jornada comigo, partilhado tanto de alegrias como de momentos menos bons.

Aos meus mais recentes colegas, Mariana, Marta, Daniel e Carolina que espero ter podido apoiar, tendo mantido o ambiente fantástico para se trabalhar.

A todos os membros da ToxOmics, por todo o acompanhamento e ajuda.

Aos meus amigos pelos bons momentos, estando comigo desde sempre, especialmente á Mariana Laureano e a Jéssica Kamiki assim como a minha prima Liliana que passou pelo mesmo que eu estando sempre disponível.

Por último e muito importante, porque sem eles eu não estaria onde me encontro, à minha família, mãe, pai, e irmão, por todo o amor, apoio incondicional, sacrifício e confiança que sempre demostraram por mim e em mim.

Muito Obrigado a todos!

Abstract

Cancer is considered a genetic disease that incurs in uncontrolled and disproportional growth of the cells with distinct characterizing hallmarks. Breast Cancer (BC) is the most common type of cancer and a leading cause in death of women therefore remaining a target of study. Hereditary breast cancer (HBC) is tightly associated with mutations in tumor suppressor genes such as the *BRCA* genes. These are highly penetrant genes involved in the DNA damage response. *BRCA1* codes for a multifunctional protein that, when mutated with a pathogenic variant impairs many pathways in the cell. Genetic tests have increasingly been asked for these genes, many resulting in variants of unknown significance (VUS). VUS represent a concern as their biological impact is not known thus, their study is of vital interest. Several are the factors involved in their study. These include mRNA splicing assays, seen as sometimes these VUS lead to alternative splicing of the gene, thus identified as splice variants (SV). Our work is divided in two tasks, firstly our aim was to identify potential SV in different cell lines commonly used in BC research, using one non-tumorigenic (MCF-10A) for comparative purposes and three distinct tumorigenic cell lines (MCF-7, MDA-MB-231 and SK-BR-3). We performed mRNA splicing analysis for the exon 11 (highly alternative spliced region) of the *BRCA1* gene using PCR-based techniques. The results obtained for this task indicate that no SV was found in this exon for these *in vitro* models. In the second part of our study, we cultured an MCF-10A cell line previously cloned by our group with a VUS of the *BRCA1* gene located in the exon 11, using CRISPR-Cas9, and applied the same method used for mRNA splicing analysis of the other cells. First, we had to confirm that the cells indeed carried the VUS, which was confirmed establishing a new homozygous *in vitro* model, and only then applied the method, which, after analysis, apparently indicated that this is not a SV. For further confirmation off the results, we performed direct RNA sequencing by nanopore. Unfortunately, this data is still under analysis and the results are not presented here.

Keywords: BRCA1; Exon 11; Breast Cancer; Splice Variants; Alternative splicing; mRNA splicing analysis;

Resumo

O cancro é considerado uma doença genética que incorre num crescimento descontrolado e desproporcional de células que apresentam características distintas. O cancro da mama é o tipo de cancro mais comum e uma das principais causas de morte das mulheres, permanecendo por isso um alvo de estudo para a comunidade científica. O cancro hereditário da mama está frequentemente associado a mutações em genes supressores de tumores, tais como os genes *BRCA*. Estes são genes altamente penetrantes envolvidos nas vias de reparação do DNA. O gene *BRCA1* codifica para uma proteína multifuncional que, quando mutada com uma variante patogénica, prejudica muitas vias na célula. Testes genéticos têm sido cada vez mais solicitados para estes genes, muitos resultando em variantes de significado desconhecido (VUS). VUS representam uma preocupação, uma vez que o seu impacto biológico não é conhecido, sendo o seu estudo de interesse vital. Vários são os fatores envolvidos no seu estudo incluindo ensaios de análise do splicing pelo mRNA. Estes devem-se ao facto de que por vezes estas VUS despoletam mecanismos de splicing alternativo do gene, sendo por isso identificadas com variantes de splicing (SV). O nosso trabalho consistiu em duas partes. O primeiro objectivo foi identificar potenciais SV em diferentes linhas celulares muito utilizadas na investigação do cancro da mama, utilizando uma linha não tumoral (MCF-10A) para comparação, e três linhas tumorais distintas (MCF-7, MDA-MB-231 e SK-BR-3). Foi realizada a análise de splicing para o exão 11 (região alvo de muito splicing alternativo) do gene *BRCA1*, utilizando técnicas de PCR. Os resultados obtidos indicam que nenhuma SV se encontra presente neste exão para estas linhas. Na segunda parte do estudo, utilizou-se a linha celular (MCF-10A) anteriormente clonada, pelo nosso grupo, com uma VUS do gene *BRCA1* localizada no exão 11, utilizando CRISPR-Cas9, aplicando-se de seguida o mesmo método de análise de splicing. Confirmou-se que as células continham a VUS em homozigotia estabelecendo-se um novo modelo *in vitro*. Os resultados indicaram também que esta VUS não se trata de uma SV. Para confirmação destes resultados, procedemos à sequenciação direta do RNA por Nanopore. Infelizmente, estes dados estão ainda sob análise, não estando os resultados aqui apresentados.

Palavras-chave: BRCA1; Exão 11; Cancro da Mama; Variantes de Splicing; Splicing Alternativo; Análise de Splicing;

List of Abbreviations

ATM – Ataxia Telangiectasia Mutated
ATR – Ataxia Telangiectasia Mutated and Rad3-related protein
BARD1 – BRCA1-associated RING domain protein 1
BC – Breast Cancer
BRCA1 – Breast Cancer gene 1 early onset
BRCA2 – Breast Cancer gene 2 early onset
BRCT - BRCA1 C-Terminal
cDNA – Complementary DNA
DMEM – Dulbecco's Modified Eagle's Medium
DNA – Deoxyribonucleic Acid
DSB – Double-Strand Breaks
EGF – Epidermal Growth Factor
ER – Oestrogen Receptor alpha
ESE – Exonic Splicing Enhancers
ESS – Exonic Splicing Silencers
FBS – Fetal Bovine Serum
HBOC – Hereditary Breast and Ovarian Cancer
HBC – Hereditary Breast Cancer
HER2 – Human Epidermal Growth Factor Receptor-2
HR – Homologous Recombination
IR – Infrared Radiation
LOH – Loss of Heterzygosity
MRI - Magnetic resonance imaging
mRNA – Messenger RNA
PBS – Phosphate-buffered saline
PCR – Polymerase chain reaction
PR – Progesterone Receptor
Pre-mRNA – Precursor mRNA
RAD51 – DNA repair protein RAD51
RING – Really Interesting New Gene
RNA – Ribonucleic Acid
ROS – Reactive Oxygen Species b
SMT – Somatic Mutation Theory
SNV – Single Nucleotide Variant

SV – Splice Variant(s)

TNBC – Triple Negative Breast Cancer

TOFT - Tissue Organization Field Theory

TP53 – Tumor Protein 53

UV – Ultraviolet Light

VUS – Variants of Unknown Significance

WT – Wild Type

Table of Contents

List of Figures

List of Tables

1. Introduction

1.1. Breast Cancer

Cancer is a leading cause of death worldwide with a number of victims that tends to increase each year according to the World Health Organization (WHO[1]). In 2020 were estimated 18.1 million new cancer cases around the world (WCRFI, World Cancer Research Fund International[2]). Cancer is a multifactorial disease where the cell grows unregulated and disproportionately due to dysregulation of diverse pathways. This dysregulation will affect cell fundamental processes leading to changes known as the "hallmarks" of cancer. The most prominent of these "hallmarks" are resistance to cell death, higher replicative and invasiveness potential, and capacity to elude the immune surveillance (Kreeger and Lauffenburger, 2010). A major consequence of these features is that such cell will proliferate, and, in a secondary stage, the cancer will metastasize. Metastasis accounts for about 90% of cancer-associated mortality. It is a complex process that can happen due to, what may be, a single cancer cell that acquires the ability to migrate and invade other organs. It enters the bloodstream and disseminates to other parts of the body, forming secondary tumors (Chaffer and Weinberg, 2011).

There are two divergent theories that have been consolidating through the years on the attempt to explain the driving forces behind cancer development, the somatic mutation theory (SMT) and the tissue organization field theory (TOFT). The SMT masters that the driving cause of neoplastic lesions is set on a DNA level with a single somatic cell that accumulates several mutations in specific genes regulating in some degree those fundamental processes mentioned before. Contrary to this, TOFT enforces that tissue-level events triggered by carcinogenic agents are the cause, resulting in DNA mutations. Despite of this there are still a large number of clinical cases which gather biological facts that raise issues, putting both theories in doubt. However, what both theories are in accordance is that there is a genetic factor associated with cancer, therefore it can be considered a genetic disease, existing also physical, chemical, and biological external agents associated with it. These agents can arise due to risk behaviors or risk factors (Rosenfeld, 2013).

There are many types of cancer, but the most common form is breast cancer (BC) with a percentage of 12,5 % of all cancer cases in 2020 (WCRFI, World Cancer Research Fund International[3]). In woman the estimated prevalence of this type of

[1] https://www.who.int/news-room/fact-sheets/detail/cancer, last accessed on November 23rd, 2022.
[2, 3] Worldwide cancer data | World Cancer Research Fund International (wcrf.org), last accessed on November 23rd, 2022.

cancer is currently 1 in 8. Although there are men who also develop BC, the rates are much lower in this gender (Giaquinto *et al.*, 2022).

Breast cancer has a heterogeneous nature which can be clinically classified based on histological type, tumor grade, lymph node status and molecular characteristics, the last one being the protein expression status of oestrogen receptor alpha (ER), progesterone receptor (PR) and human epidermal growth factor receptor-2 (HER2). This characterization will allow to decide which of the subtypes of breast cancer the tumor belongs to, Luminal A (ER+/PR+, HER2-), Luminal B (ER+/PR+, HER2+/-), HER2-enriched (HER2 positive) or Basal-like (TNBC) which is the most aggressive but also the rarest subtype. It accounts for about 12% of all BCs, indicating a different etiology. The tumor can start in any part of the breast, but the majority of BC's begins in the ducts or lobules (Jiang *et al.*, 2016). Metastasis is what normally leads to death from BC (Scully *et al.*, 2012).

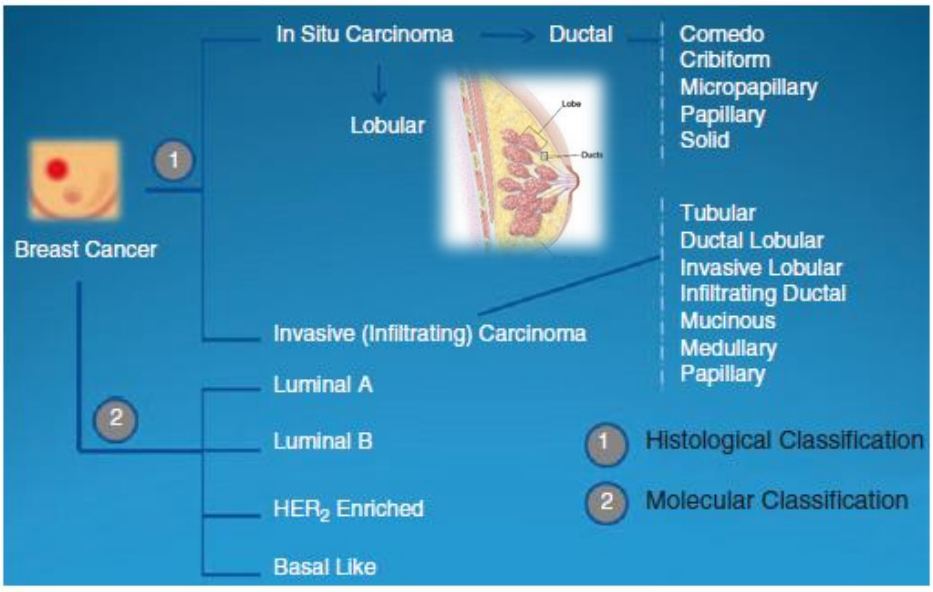


Figure 1.1 – Classification of BC subtypes. BC can be subtyped based in histological type (**1**) and in molecular characteristics depending on the expression status of ER, PR and HER2 (**2**). (Adapted from: Ahmad, 2013)

The main risk factors, unchangeable in an individual, associated with BC are age (≥ 55 years, mostly), gender, familial history, and genetics. Others are high breast density, obesity, and elevated alcohol consumption (American Cancer Society[4]).

[4] https://www.cancer.org/cancer/breast-cancer/risk-and-prevention/breast-cancer-risk-factors-you-cannot-change.html, last accessed on November 24th, 2022.

Despite family history being a major risk for developing BC, the number of individuals with close relatives that suffer from the disease are significantly lower compared to the cases where the BC is considered sporadic and not associated with penetrant genes (Ahmad, 2013).

1.2. *BRCA1* gene and its role in BC

The genetic hereditary factor accounts for about 5 to 10% of all breast cancer cases with the majority of them being attributed to mutations in *BRCA1* (Breast Cancer 1, early onset) and *BRCA2* (Breast Cancer 2, early onset) genes passed through generation in an autosomal dominant manner (Ahmad, 2013; Jiang *et al.*, 2016). These are tumor suppressor genes, which once mutated may be associated with susceptibility to develop BC, therefore, identified as moderate-to-high penetrant genes (Jiang *et al.*, 2020). For the purpose of this project, we will focus on *BRCA1* gene.

BRCA1 was first mapped more than 30 years ago on chromosome 17q21.3 in a hereditary breast cancer study, and it was first cloned in 1994. *BRCA1* has a total of 24 exons, but there is some conflict about the number of coding exons it contains due to an historical misannotation of exon 4. Most authors refer it is composed by 22, others that it has 23 coding exons. For a matter of coherence, we here will admit it has 22 coding and 2 non-coding exons (Jong, de *et al.*, 2017; Ruiz de Garibay *et al.*, 2021).

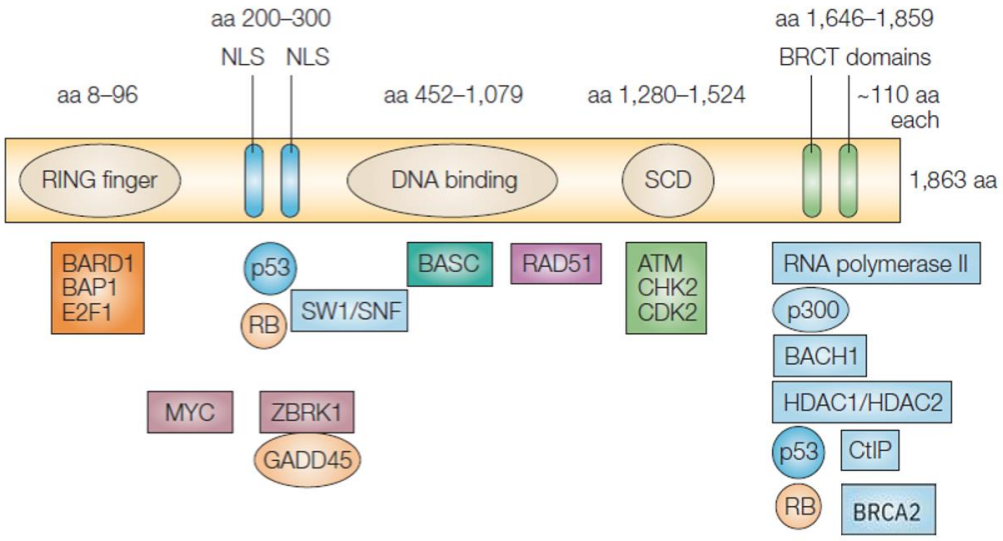


Figure 1.2 – BRCA1 functional domains and interactions. Representation of BRCA1, and its several functional domains that interact with a great range of proteins. The RING-finger domain binds to BARD1. The two regions at the C-terminus are the BRCT domains that can be found in proteins involved in DNA-repair pathways and interact with many proteins. They allow the interaction with BRCA2 protein. (Adapted from: Narod and Foulkes, 2004)

BRCA1 encodes a pleiotropic nuclear protein with 1863 aa., of 220kDa that possesses three well conserved structural and functional domains, among others. One RING (Really Interesting New Gene) finger is found at its N-terminus and two BRCT (BRCA1 C-Terminal) domains at its C-terminus (Figure 1.2). BRCA1 operates in both, cell-cycle-checkpoint activation (the G1–S and the G2–M transition) and DNA damage repair, playing a major role in this process, acting also in protein ubiquitination, transcriptional regulation and in chromatin remodeling. The role and the interactions of this protein, in order to maintain genomic stability, is highly complex and, until today, not completely clear but what is known is that BRCA1 protein interacts to BRCA2, TP53, BARD1 and RAD51, among many other proteins associated with the cell cycle and DNA damage response pathways being phosphorylated by ATM and other kinases (Figure 1.2). The lack of a functional produced protein will lead the cell to not undergo arrest in G2 phase of the cell cycle consequently affecting deeply the activation of DNA damage response to double-strand breaks (DSB) specially by homologous recombination (HR). DNA DSB are usually induced, either by external and internal hazards, such as radiation (IR, UV) or oxidative stress (ROS). Upon DNA damage, BRCA1 interacts with γH2AX (which is a histone variant phosphorylated in S_{139} either by ATM or ATR) in order to remodel chromatin structure acting as an intermediate of other proteins in the DNA damage site. (Godet and M. Gilkes, 2017; Krum *et al.*, 2010; Orr and Savage, 2015; Tarsounas and Sung, 2020; Wu, Lu and Yu, 2010).

The *BRCA1* gene is expressed in several tissues, such as breast and ovarian tissue. Its association with tumorigenesis sets on the role of this gene in those various cellular processes mentioned above. But as a tumor suppressor gene, *BRCA1* is functionally recessive, which means that both copies must be inactivated for protein loss of function. Many studies reenforce the role of this gene in BC and estimate that inherited mutations in *BRCA1* increase the risk of developing breast cancer by age 70 to 80%, and that patients that are carriers of these mutations usually develop cancer before the age of 50. Despite being well established that *BRCA1* plays a role in certain cancers it is still not completely understood the reason why mutations in this gene are mostly related to hereditary breast and ovarian cancer syndrome (HBOC). This could be simply because not so many studies related to this gene were undertaken for other types of cancer, however some relatively recent findings about how BRCA1 may repress the production of oestrogen metabolite may explain why *BRCA1* mutations carriers tend to develop cancer, specifically in hormonal tissues such as breast and ovaries. Since *BRCA1* is functional recessive, on individuals with inherited mutations in only one copy of this gene, the second copy must be somehow mutated

or lost during lifespan in order the cancer to develop. This may happen due to an acquired somatic mutation or due to what is called loss of heterozygosity (LOH), meaning the loss of the WT allele (Figure 1.3) (Curtit *et al.*, 2015; Godet and M. Gilkes, 2017; Jiang *et al.*, 2020; Orban and Olah, 2003; Orr and Savage, 2015).

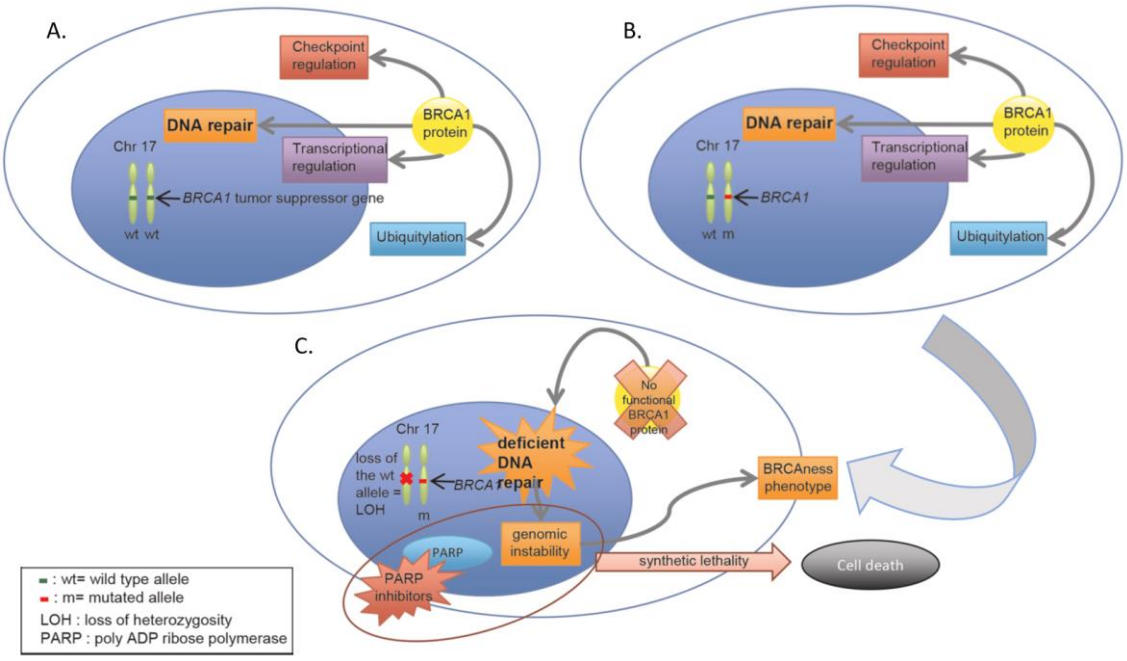


Figure 1.3 – Schematic representation of BRCA1 functions and alleles. A- BRCA1 fully functional with no mutation in any of the alleles and functions driscription. **B** – Germline mutation in one allele with the protein still functional. **C** – Loss of heterzygosity (LOH) due to loss of the second allele (WT), which will lead to genomic instability. (Adapted from: Curtit *et al.*, 2015).

BRCA1 mutations often appear in patients which the cancer was classified as TNBC. In accordance, studies have shown that women who carry these mutations are more likely to develop secondary tumors which may appear in the same breast or in the other. Histological characteristics of this cancers include high proliferation indices, invasive borders, and lymphocytic infiltrations. Interestingly, in sporadic BC patients with no mutations in *BRCA1*, its expression is often down-regulated, possibly as a result of promoter methylation or over-expression of ID4, a negative regulator of *BRCA1* expression. The poor prognosis normally attributed to certain patients with BC is mainly due to lack of a standard care approach as there are no targeted treatments currently available. On that note, many efforts are being made to change this (Ahmad, 2013; Orr and Savage, 2015).

1.3. Genetic testing and VUS classification in BC

Although breast cancer is the most common type of cancer, it is not the one that accounts for more deaths (WHO[5]). This happens because each passing year breast cancer cases can be detected earlier. Some HBCs (Hereditary breast cancers), can even be prevented if detected before it develops. What may help the early diagnostic and prevention, in these cases, is for individuals to assess their predisposition to this type of cancer.

Many efforts have been made to understand the human genetic and the underlying traits of the heritability phenomena. Assess the susceptibility of an individual to certain disease is now possible due to advanced scientific technologies (such as next generation sequencing, NGS) and genomic studies that enable the discovery of unique hereditary mutations allowing, nowadays, genetic testing on a clinical level (Godet and M. Gilkes, 2017).

Genetic testing has increasingly been urged by clinicians for clinical evaluation of patients. Also, due to the decline in costs of sequencing, the number of individuals performing it, has globally increased in an exponential way (Wright *et al.*, 2018). Genetic testing, also called genomic testing, consists in DNA screening for variants of specific genes associated with a certain inherited condition that is usually considered rare (NHS, National Health Service[6]; American Cancer Society[7]). Often, it provides diagnostic and prognostic information that can be used in management of a wide variety of health conditions, as with cancer (Burke, 2002).

The interest in genetic testing by oncologists has been gradually increasing in cancer management, even for patients with no family history that present young onset BC. Due to the association of certain types of BC with high frequency of *BRCA1* mutations, many clinicians ask for *BRCA1* genetic tests, or more precisely for multigene panel genetic tests that include the *BRCA1*. Possible results for a genetic test are, positive – as a variant considered pathogenic was found; negative – if no mutation was detected, or if the variant that was found has no clinical significance, meaning it was already characterized as non-pathogenic; or sometimes the result can be a variant of unknown significance (VUS).

VUS are variants of genes, that have not been characterized as their biological effect has not yet been studied. This may be due to their low frequency in population or due to other reasons of lack of association. After observing the effect of such variant,

[5] https://www.who.int/news-room/fact-sheets/detail/cancer, last accessed on November 23rd, 2022.
[6] https://www.nhs.uk/conditions/genetic-and-genomic-testing/, last accessed on November 25th,2022.
[7] https://www.cancer.org/healthy/cancer-causes/genetics/genetic-testing-for-cancer-risk/understanding-genetic-testing-for-cancer.html, last accessed on 25th, 2022.

which can be found in either a regulatory or coding region of the gene, it is then reclassified as pathogenic or benign. Because of the increased use of genetic testing in this area the number of VUS that have been discovered is increasing, as well as for the need for their characterization (Wright *et al.*, 2018) (Eccles *et al.*, 2015). Many factors such as the results of functional, biochemical and RNA splicing assays, clinical context and family history must be taken in consideration in order to classify VUS (Figure 1.4). Based on a multifactorial quantitative analysis model, VUS have been classified by the ENIGMA (Evidence-based Network for the Interpretation of Germline Mutant Alleles) consortium in B/LB (Benign/Likely Benign), VUS, or P/LP (Pathogenic /Likely Pathogenic). Despite all efforts, classifying VUS still presents a major challenge (Sullivan *et al.*, 2021).

Figure 1.4 – Factors involved in VUS classification. In order to be classificated VUS must be subject of a series of studies and the context in which it was found shuold be well considerated and understood. (Adapted from: http://epilepsygenetics.net/2016/08/11/nothings-for-sure-thats-for-sure-evaluating-variants-of-uncertain-significance/).

Most VUS found in genetic tests are classified as benign but, approximately one-fourth of them reclassify to pathogenic. The growing use of multigene panels has proven to be of clinical utility, however they have led to an increase of VUS discovery presenting difficulties for patients and clinicians in counseling and risk management (Chang *et al.*, 2019).

As verified, pathogenic variants of *BRCA1* cause hereditary breast and ovarian cancer syndrome (HBOC), which often presents poor prognosis. Thus, prevention strategies such as prophylactic mastectomy, surveillance by annual magnetic resonance imaging (MRI), and chemoprevention upon positive results on BRCA1 genetic tests may reduce breast cancer risk in these individuals. Therefore, genetic testing can play an indispensable role in preventing HBC, and also help on cancer therapy choice (Concolino *et al.*, 2019; Godet and M. Gilkes, 2017; Wangensteen *et al.*, 2019)

1.4. Splice Variants and Alternative Splicing in *BRCA1*

VUS are variants of specific genes associated with certain genetic diseases that may or may not be pathogenic. In order to reclassify each of them, several studies are performed. As mentioned, these studies include mRNA splicing analysis. This kind of analysis is important and mostly implicated when the variant in study may present characteristics of a splice variant (SV). SV are mostly identified when a VUS occurs within the intron-exon boundary region, in any of the conserved splicing elements, or when it creates new splice sites within the DNA sequence (exons and introns included), leading to aberrant or what is called alternative splicing. Its study can help figuring in which classification the VUS falls in (Bergsma *et al.*, 2018; J.Y., J. and R., 2014). Alternative splicing is used in higher eukaryotes to regulate gene expression and functional diversification of proteins in different tissues, being cell type specific. Alternative splicing of pre-mRNA provides different products allowing functional diversity of protein isoforms, all derived from a single coding sequence (Lixia *et al.*, 2007). Interestingly almost 95% of the human gene transcripts undergo alterative splicing (Camacho Londoño and Philipp, 2016).

Genes normally express one major transcript which is called the canonical isoform containing all the canonical exons. When alternative splicing occurs, canonical splice sites can be skipped, and alternative splice sites can be used. Potential outcomes of alternative splicing include exon skipping and intron retention, as well as reading frame shifts that create premature termination codons that will lead to the decay of the transcript (Figure 1.5) (Bergsma *et al.*, 2018).

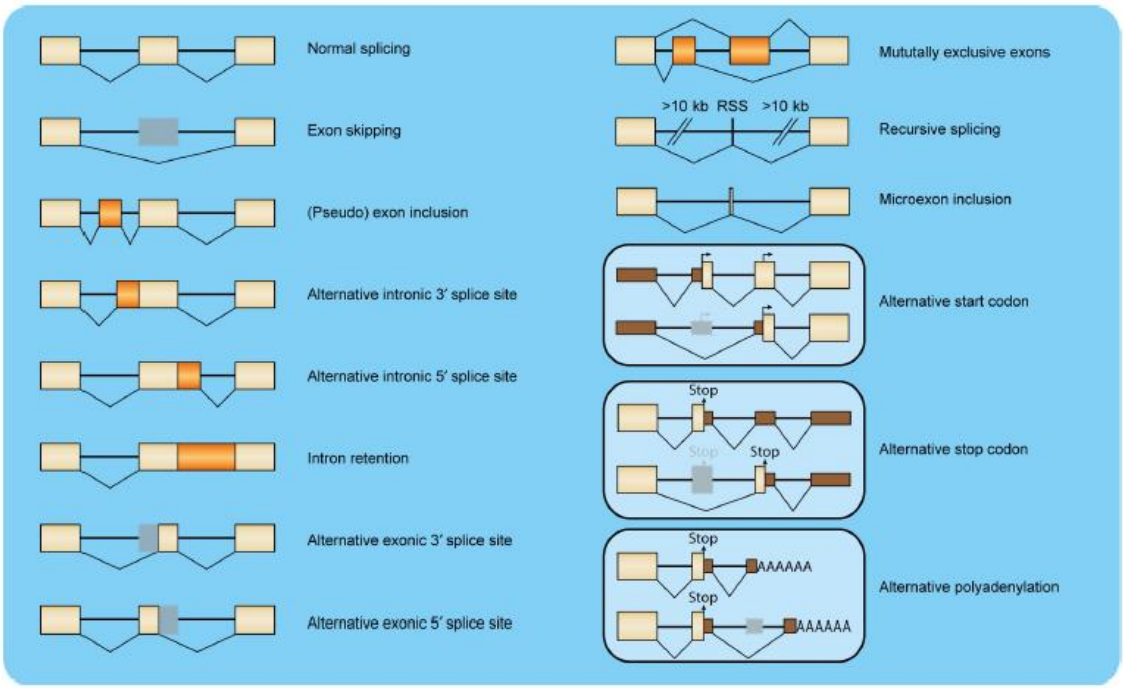


Figure 1.5 – Potential outcomes of alternative splicing. – Canonical exons are presented in beige, alternative exons and retained intronic sequences in orange and, partially skipped exons in gray. UTRs are in brown. (Bergsma *et al.*, 2018).

Alternative splicing can be considered one of the main mechanisms that contributes to biological diversity, being therefore tightly regulated. When this regulation is somehow disrupted, it can lead to diseases, as it happens on the presence of pathogenic SV (Camacho Londoño and Philipp, 2016).

To understand the alternative splice mechanism, and the different outcomes of this process, it is important understand how the splicing process works. Splicing is a very complex process regulated is several steps that uses a specific RNA/protein machinery called spliceosome. This complex will recognize conserved splice elements such as the 5' and 3' splice sites (also called doner and acceptor sites, respectively), and the branch site of the intron. On the presence of any variant on these splice elements or in the genes encoding for the subunits of the spliceosome, alternative splicing not expected can occur, showing the importance of SV detection. There are also other regulatory elements within the pre-mRNA sequence such as exonic splicing enhancers (ESEs), intronic splicing enhancers, exonic splicing silencers (ESSs) or intronic splicing silencers (ISSs) that, when mutated, can alter the normal splicing of the pre-mRNA (Bergsma *et al.*, 2018).

Therefore, alternative splicing can be highly relevant to disease, and for this matter, in tumorigenesis, also playing a role in therapy. It has been shown that changes in splicing can lead to loss of function of tumor suppressors or gain of function of proteins that will promote cancer development (Lixia *et al.*, 2007).

The expression of *BRCA1* tumor suppressor gene has been highly studied, once it appears to be differentially expressed in different tissues (Xiping *et al.*, 2017). This may happen due to alternative splicing occurring in such tissues. There is evidence of different transcripts being expressed from WT *BRCA1* suggesting that there is an impact of alternative splicing in *BRCA1* expression in hormonal tissues such as breast and ovarian tissues, what may explain why this gene is specifically associated with HBOC.

Exon 11 (10 in NCBI) is the major exon of *BRCA1*, with 3426 bp. Thus, many alternative splicing mechanisms occurring in this gene end up affecting specifically this exon. Three majors in-frame transcripts have been found for *BRCA1* gene, the full-length isoform (BRCA1 FL), the BRCA1- Δ11 isoform (skipping of the entire exon 11), and the BRCA1- Δ11q isoform (partial skipping of exon 11) (Figure 1.6). Different rates of expression of these transcripts were found in individuals with BC compared to healthy individuals. It was also reported that BC patients bearing exon 11 mutations have a worse overall survival compared to non-exon 11 mutation carriers (Ruiz de Garibay *et al.*, 2021; Yang *et al.*, 2019)

Figure 1.6 – The three main transcripts of BRCA1 with changes in exon 11. Alternative splicing events are shown on the left and their products on the right. (Adapted from: Yang *et al.*, 2019)

All this evidence taken together may indicate that BC cases associated with *BRCA1* mutations (especially in the 11 exon or its surroundings), or even the cases where alteration in *BRCA1* expression rates are observed, is caused by mutations that impact alternative splicing events, the so-called SV.

1.5. Aim and Clinical Relevance

With the growing global burden of this disease and its implications, prevention of cancer is one of the most significant public health challenges of the 21st century. As seen, early diagnosis, especially when the cancer has not yet metastasized, can be the key for a successful response to treatment of breast cancer. This research is one between many that intends to help on a clinical level the early diagnosis and even prevention.

Seen as SVs have shown to represent a challenge in genes associated with BC such as *BRCA1,* it is important to focus on research and methodologies that will allow their analysis and classification in an easier way.

With this general goal we here used three BC cell lines (MCF-7, MDA-MB-231 and SK-BR-3) and one non-tumorigenic breast cell line (MCF-10A) that are commonly used in BC research, focusing on screening their DNA for SVs that may exist in the 11th exon of the *BRCA1* gene, since it is a highly alternative spliced region of this gene, using for that one method based in PCR technics.

As a second step we intended firstly to confirm the presence of a VUS that was previously discovered and studied in our laboratory, in an MCF-10A cell line genome edited by CRISPR-Cas9. After confirmation of the point mutation in one clone, we submitted this new *in vitro* model to the same mRNA splicing analysis PCR-based to understand if it can be considered a SV or not.

2. Materials and Methods

Materials and Methods

2.1. Cell Culture

For the first task of this project, four different epithelial, human cell lines already established were used, one non-tumorigenic breast cell line, the MCF-10A, and three breast cancer lines, MCF-7, MDA-MB-231 and SK-BR-3. The characterization of these three lines follows in **Table. 2.1** (ATCC, 2012; Kao *et al.*, 2009; Krum *et al.*, 2010).

Table 2.1 - Caracterization of breast cancer cell lines.

Cell line ID	Subtype of BC (Molecular characterization)	Metastatic / Non-metastatic	Histology
MCF-7	Luminal A	Non-Metastatic	Metastatic adenocarcinoma
MDA-MB-231	TNBC	Metastatic	Metastatic Adenocarcinoma
SK-BR-3	HER2-enriched	Metastatic	Adenocarcinoma

2.1.1. MCF-10A cells culture

The MCF-10A cell line was cultured in Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham (Sigma-Aldrich #D8437) previously supplemented with 5% of Horse serum, Donor Herd (Sigma-Aldrich #H1270), 1% of Penicillin-Streptomycin (10,000 units/mL penicillin and 10 mg/mL streptomycin; Pen Strep, Gibco, Ca#15140122), Epidermal Growth Factor (EGF – Sigma Aldrich #E9644) in a final concentration of 20 ng/mL and Hydrocortisone (Sigma-Aldrich, #H0888) which was first diluted in ethanol 100% (1 mg/mL), with a final concentration of 0.5 µg/mL in the medium. At the time of culturing the cells, the medium was supplemented with insulin (Sigma-Aldrich #I9278) at the final concentration of 10 µg/mL, and choleric toxin (Sigma-Aldrich, #C8052) (1mg/mL) pre-diluted in ddH2O 1:100 with a final concentration of 0,1 µg/mL in the medium. The cells were incubated at 37°C with 5% CO_2 in a humidified chamber.

2.1.2. MCF-7 cells culture

MCF-7 cells were cultured in Low Glucose Dulbecco's Modified Eagle Medium (DMEM – Gibco #31885), previously supplemented with 10% Fetal Bovine Serum (FBS

– Sigma-Aldrich #F7524) and 1% of Penicillin-Streptomycin (10,000 units/mL penicillin and 10 mg/mL streptomycin; Pen Strep – Gibco, Ca#15140122). At the time of culturing, insulin (Sigma-Aldrich #I9278) was added to a final concentration of 10 µg/mL in the medium and then the cells were incubated at 37°C with 5% CO_2 in a humidified chamber.

2.1.3. MDA-MB-231 cell culture

MDA-MB-231 cells were cultured in Low Glucose Dulbecco's Modified Eagle Medium (DMEM – Gibco, Ca#31885), previously supplemented with 10% Fetal Bovine Serum (FBS – Sigma-Aldrich #F7524) and 1% Penicillin-Streptomycin (10,000 units/mL penicillin and 10 mg/mL streptomycin; Pen Strep – Gibco, Ca#15140122). At the time of culturing the cells, insulin (Sigma-Aldrich #I9278) was added to a final concentration of 10 µg/mL. in the medium and then were incubated at 37°C with 5% CO_2 in a humidified chamber.

2.1.4. SK-BR-3 cell culture

SK-BR-3 cells were cultured in McCoy's 5A Medium (Gibco, Ca#26600), previously supplemented with 10% Fetal Bovine Serum (FBS – Sigma-Aldrich #F7524) and 1% Penicillin-Streptomycin (10,000 units/mL penicillin and 10 mg/mL streptomycin; Pen Strep, Gibco, Ca#15140122). At the time of culturing the cells, insulin (Sigma-Aldrich #I9278) was added to a final concentration of 10 µg/mL. in the medium and then the cells were incubated at 37°C with 5% CO_2 in a humidified chamber.

2.1.5. MCF-10A VUS BRCA1 clone cell culture

For the second task of this project, an MCF-10A cell line previously cloned with a BRCA1 VUS (NM_007294.4:c.1067A>G; rs1799950: VUS identification in section 2.4.1) through genome editing using the CRISPR-Cas9 system performed by our group (Pedro, A., 2021), was cultured mostly following the guidelines used for the standard MCF-10A in section **2.1.1**. Initially, it was given a bust of additional 2,5 % of Horse serum, Donor Herd (Sigma-Aldrich #H1270) and the medium was only partially changed. The cells were also incubated at 37°C with 5% CO_2 in a humidified chamber.

All cell lines are adherent and were cultured in ventilated T -flasks until 70% to 90% of its confluence was achieved. Cells were then trypsinized to larger T-flasks with

10% trypsin solution from porcine pancreas (Sigma Aldrich #T4549) and 90% versene plus sodium bicarbonate (sodium bicarbonate 8.4% BRAUN #304494), centrifuge at 800 rpm for 5 min., and then resuspended in their respective supplemented medium and added to the T-flask.

2.2. DNA extraction and quantification

After cultured in T-flasks (T-175) until they reached 70 to 90% of its confluence, cells were again trypsinized and then centrifuge at 800 rpm for 5 min at room temperature. Cells were resuspended in 2 ml of PBS (Sigma Aldrich #P3813-10PAK) and counted, obtaining a cell suspension of at least 10^6 cells per mL. For DNA extraction of the cells, it was followed the Quick DNA Miniprep Plus Kit (Zymo Research #D4069) protocol according to the instructions. For the DNA elution step, it was followed the suggestion given and the samples were eluted twice. The quantification of DNA samples extracted was obtained by NanoDrop™ 2000 spectrophotometer (Thermo Fisher). DNA samples were stored at -20°C.

2.3. RNA extraction and quantification

For total RNA extraction, cells were submitted to the same initial steps as in DNA extraction in section 2.2. After resuspension in PBS and counted, cells were again centrifuge, now at 1000 rpm and resuspended in an appropriated volume of QIAzol Lysis Reagent (Qiagen #5346994) or TRIzol™ Reagent (Ambion #15596026) according to the Direct-zol™ RNA Miniprep Plus Kit (Zymo Research #R2071) quick protocol. The protocol was followed according to the manufacturer's instructions and the DNase I treatment was performed as recommended. All steps were held at room temperature and centrifugations were carried out at 16000 rcf for 30 sec. RNA was quantified by the NanoDrop™ 2000 spectrophotometer (Thermo Fisher) and then stored at - 80°C.

2.4. DNA sequencing to confirm point mutation of *BRCA1* VUS clones obtained

2.4.1. *BRCA1* VUS background

In a previously study, also developed in our group by Adubeiro, R. (2018), where two healthy individuals with family history were screened by NGS for a panel of genes, the same VUS was found (identified as NM_0072 94.3:c.106 7A>G at the time) in the

BRCA1 gene of both individuals. This is a missense variant seeing as it changes only one base in the sequence, replacing an adenine (A) by one guanine (G) on the forward sequence strand, and a thymine (T) by a cytosine (C) in the reverse sequence strand, turning a supposedly glutamine (Gln) into an arginine (Arg). This VUS was a target of study due to divergent classifications given by distinct *in silico* prediction tools- the *PolyPhen-2* report predicted it to be "probably damaging" (Lourenço, 2018; Pedro, 2021). Meanwhile its status in NCBI has been in constant update but the last evaluation classifies it as benign (Last identified as NM_007294.4(BRCA1):c.1067A>G (p.Gln356Arg); rs1799950).

In order to perform functional studies with cells carrying this VUS, the CRISPR-Cas9 technology was used at the time to introduce the VUS of interest in the MCF-10A cell line creating a library of possible clones, being one of them successfully implemented, establishing a new *in vitro* model, the MCF10-A heterozygous clone of the VUS that was then submitted to the intended functional studies. The remaining possible clones were storage, and some were used in our current study.

2.4.2. Confirmation of point mutation by PCR and sequencing

In order to confirm the presence of the single nucleotide variant (SNV) *BRCA1* VUS (NM_007294.4:c.1067A>G) in the DNA samples of 4 possible clones (B, F, G and K), we sequenced the *BRCA1* VUS target region, firstly by designing the primer pair that covered the region where the VUS is found (489bp sequence within the exon 11). The primer design was performed with the help of the Oligo Analyzer Tool from Integrated DNA Technology (IDT) that gives information about the melting temperature for each primer (Table 2.2.).

Table 2.2 - Primers for BRCA1 VUS sequence and respective Tm °C.

Target region	Primers	Tm (°C)
BRCA 1 VUS (NM_007294.4:c.1067A>G)	**PFw:** 5'- AGCCTGGCTTAGCAAGGAGC -3'	60.1
	PRv: 5'- TGGGGAGGCTTGCCTTCTTC -3'	59.9

For amplification of the targeted regionof the VUS, we performed Polymerase Chain Reaction (PCR) using a working solution of 10 µM of the lyophilized primers

forward (PFw) and reverse (PRv) purchased from StabVida. For the master mix we used DreamTaq ™ Hot Start DNA Polymerase (Thermoscientific, #EP1702). The components and the corresponding amount for a total volume of 50 µL of the master mix used for each DNA sample follows in Table 2.3. For each master mix prepared it is used one negative control without DNA template.

Table 2.3 – Master mix components and correspondig amounts.

Component	Amounts
ddH_2O	up to 50 µL
10X DreamTaq Buffer[8]	1X
PFw	0,4µM
PRv	0,4µM
dNTPs (Bioline, #Bio-39029)	25 mM each
DreamTaq Hot Start DNA Polymerase	1.25U
DNA template (clones *B, F, G* and *K*)	~ 0,09 - 0,3 µg
Total Volume	50 µL

PCR was performed in a thermocycler (GeneAmp PCR System 9700, Applied Biosystems). The optimal conditions of temperature, time, and number of cycles for each step of the reaction used are presented in Table 2.4.

[8] The **10x DreamTaq Buffer** already contains $MgCl_2$, here on a final concentration of 2 mM.

Table 2.4 – Optimal cycling PCR conditions for amplification of *BRCA1* VUS target region.

Step	T (°C)	Time	No. cycles
Initial denaturation	95	2 min.	1
Denaturation	95	30 sec.	35
Annealing	60	30 sec.	
Extension	72	30 sec.	
Final Extension	72	7 min.	1
Stop reaction	4	---	---

PCR products were then run in an 2% agarose (Bioline #Bio41025) gel with the Hyperladder™ IV 100 bp (Bioline #Bio-33056), using 1X DNA Loading Buffer Blue (Bioline #Bio-3745) in 10ul of total volume with PCR product sample. The electrophoresis was performed at 110 V for 60 min. The gel was stained with a solution of 3x GelRed (GelRed 10000x Biotium #41003) for 20 minutes, and then the bands were observed in the Chemidoc™ Touch (Bio-Rad).

The remaining of the PCR products were purified using the GeneJet PCR Purification Kit (Thermo Fisher #K0701, #K0702) according to the manufacturer's instructions. Samples were quantified using NanoDrop™ 2000 spectrophotometer (Thermo Fisher) and then sequenced.

2.5. RNA splicing analysis by DNA and cDNA PCR

To detect the presence of SVs in the exon 11 of the *BRCA1* gene, total RNA extracted was converted into cDNA and it was performed PCR analysis with DNA samples and the corresponding cDNA, with primer pairs comprising the boundaries between the exon 11 and the two exons enclosing it.

2.5.1. Reverse transcriptase reaction

Total RNA extracted from the different cell lines was turned into cDNA by reverse transcriptase using the High-Capacity RNA-to-cDNA Kit (Applied Biosystems #4387406), following the respective protocol. For the first task of the project the

amount of RNA sample from the four established breast cell lines (MCF-10A, MCF-7, MDA-MB-231 and SK-BR-3) was adjust to ~1,4 µg per reaction with total volume of 20 µL of RT mix reaction mix, as indicated. For the second phase the amount of total RNA of the MCF-10A WT cells and the cells MCF-10A cloned with the BRCA 1 VUS (clone *F*), was adjust to 1,96 µg per reaction, also for the total volume of 20 µL of reaction mix. The reaction was carried out in a thermocycler (GeneAmp PCR System 9700, Applied Biosystems) by incubation at 37°C for 60 min. followed by heating to 95°C for 5 min. to stop the reaction. The cDNA was quantified by NanoDrop™ 2000 spectrophotometer (Thermo Fisher) and stored at - 20°C.

2.5.2. Primer's design and preparation

To identify possible alternative splicing events caused by the presence of SVs in the exon 11 of the *BRCA1* gene, we design three pairs of primers covering the region of the exon 11 and the entire intron sequences surrounding it, reaching the exons 10 and 12 of the gene. The primers were design with the help of the Oligo Analyzer Tool from Integrated DNA Technology (IDT) which gives us information about the melting temperature for each primer (Table 2.5).

Table 2.5 – Primer sequences for mRNA analysis.

Target region	Primer	Tm (°C)
Boundary exon 10 - exon 11 (P1)	**PFw 1:** 5'- CCTCAAGAACCAGGGATGAAATC -3'	57.5
	PRv 1: 5'- CGTCCAATACATCAGCTACTTTGGC -3'	57.9
Within exon 11 (P2)	**PFw 2:** 5'- GAGCCATGTGGCACAAATACTC -3'	56.4
	PRv 2: 5'- GAGGGGACGCTCTTGTATTATCTG -3'	56.7
Boundary exon 11 - exon 12 (P3)	**PFw 3:** 5'- GATTAGGGGTTTTGCAACCTGAGG -3'	58.1
	PRv 3: 5'- GTTTCACTCTCACACCCAGATGC -3'	58.0

Lyophilized primers were purchased from StabVida and were diluted in ddH_2O (stock solution of 100 µM) for a working solution of 10 µM.

2.5.3. PCR

For the PCR we used the same conditions for the three target sequence regions. It was performed using the thermocycler (GeneAmp PCR System 9700, Applied Biosystems) and the DreamTaq ™ Hot Start DNA Polymerase (Thermoscientific, #EP1702), with three reaction mix's (one for each pair of primers) (Table 2.6). The amounts of each component and the optimal conditions of temperature, time, and number of cycles are presented in Table 2.6 and Table 2.7 respectively.

Table 2.6 - Master mix for each pair of primers and corresponding amounts for each reaction.

Master mix components			**amounts**
Mix 1.	**Mix 2.**	**Mix 3.**	
ddH_2O			up to 50 µL
10X DreamTaq Buffer			1X
PFw 1	**PFw 2**	**PFw 3**	0,4µM
PRv 1	**PRv 2**	**PRv 3**	0,4µM
dNTPs (Bioline, #Bio-39029)			25mM
DreamTaq Hot Start DNA Polymerase			1,25 U
sample template			~ 0,1-0,6 µg
Total Volume			50 µL

As templates we used the cDNA and the corresponding DNA of each cell line for the different master mixes. For each master mix it was used a negative control with no sample template added.

Table 2.7 - Optimal cycling PCR conditions for cDNA and DNA amplification of target regions .

Step	T (°C)	Time	No. cycles
Initial denaturation	95	3 min.	1
Denaturation	95	30 sec.	35
Annealing	60	30 sec.	
Extension	72	1 min.	
Final Extension	72	7 min.	1
Stop reaction	4	---	---

In order to observe the different fragment sizes amplified, PCR products were applied in a 2% agarose (Bioline #Bio41025) gel with the Hyperladder™ I, 1Kb (Bioline #Bio-33053), using 2 µL of the 5x DNA Loading Buffer Blue (Bioline #Bio-3745) and 8 µL of PCR sample. The electrophoresis was run at 95 V for 90 min. Afterwards the gel was stained using a solution of 3x GelRed (GelRed 10000x Biotium #41003) for 20 minutes, and then the amplification bands were observed in the Chemidoc™ Touch (Bio-Rad).

2.6. SV detection by direct RNA sequencing with Nanopore

In order to confirm the results obtained by PCR regarding the presence of SVs in the *BRCA1* gene, direct RNA sequencing was performed by Nanopore MinION Mk1B using the Direct RNA Sequencing Kit (SQK-RNA002) and the Flow Cell Priming Kit (EXP-FLP002) from Oxford Nanopore Technologies and following the corresponding 2-steps protocol which includes the "Library preparation" and the "Priming and loading the SpotON flow cell" protocols according to the manufacture's instructions. except by an alteration made in the steps 7 and 8 of the "Library preparation" protocol. Briefly the library preparation protocol consisted in primer annealing and ligation followed by the reverse transcripatase reaction to synthesize the complementary strand of the RNA and finally the attachment of the sequencing adapters to the ends of the RNA-cDNA. For the reverse transcriptase reaction, the time of incubation (step 8) at 50°C was shortened from 50 min to 10 min. This was possible because in step 7 it was used the SuperScript IV instead of the one indicated in the protocol (SuperScript

III). For this procedure we used the total RNA previously extracted and quantified, adjusted to have 500 ng in 9 µL of volume. RNAClean XP beads (Beckman Coulter, #A63987) were used for purification during the library preparation protocol and It was used a SpotON flow cell, R9 Version (Oxford Nanopore # FLO-MIN106D) in the priming and loading.

In the second day (24h after the beginning of the first run) it was made an enrichment by following again the same protocols and applying the Flow cell Wash Kit (EXP-WSH004) protocol, running a second time the library in the same flow cell.

3. Results

Results

As mentioned before, this project consists in two main tasks: The identification of possible existent SVs in BC cell lines, comparing them to one representative of normal breast tissue (non-tumorigenic cell line); and the characterization of a VUS cloned in the MCF-10A cell line as a possible SV by distinct methodologies.

3.1. SV detection in established breast cell lines

For the first task, to DNA and RNA extraction followed by splicing analysis by PCR, were used the four already established cell lines, MCF-10A, MCF-7, MDA-MB-231 and SK-BR-3.

3.1.1. Cell culture

As referred, MCF-10A, MCF-7, MDA-MB-231 and SK-BR-3 are adherent cells lines commonly used in breast cancer research. The MCF-10A is an immortal cell line that resembles until certain point the normal tissue of the breast being the most common *in vitro* model used to this purpose (Puleo and Polyak, 2021). Meanwhile the other three are all tumorigenic and mimic distinct subtypes of BC. The MCF-7, which is the most commonly used BC cell line in this type of research, is considered a poorly aggressive and non-invasive cell line representative of luminal A subtype, with low metastatic potential (Comşa, Cîmpean and Raica, 2015). The MDA-MB-231 are a highly metastatic cell line representative of TNBC also frequently used in research (Ecacc, 2022). Finally, SK-BR-3 is another metastatic BC cell line classified as Her2-positive less frequently used than the above (ATCC, 2012).

The four cell lines were initially cultured as described in section 2.1, and their confluence was verified during the days after culturing. The four *in vitro* models differ in morphology and in growth rates as it was observed (Figure 3.1).

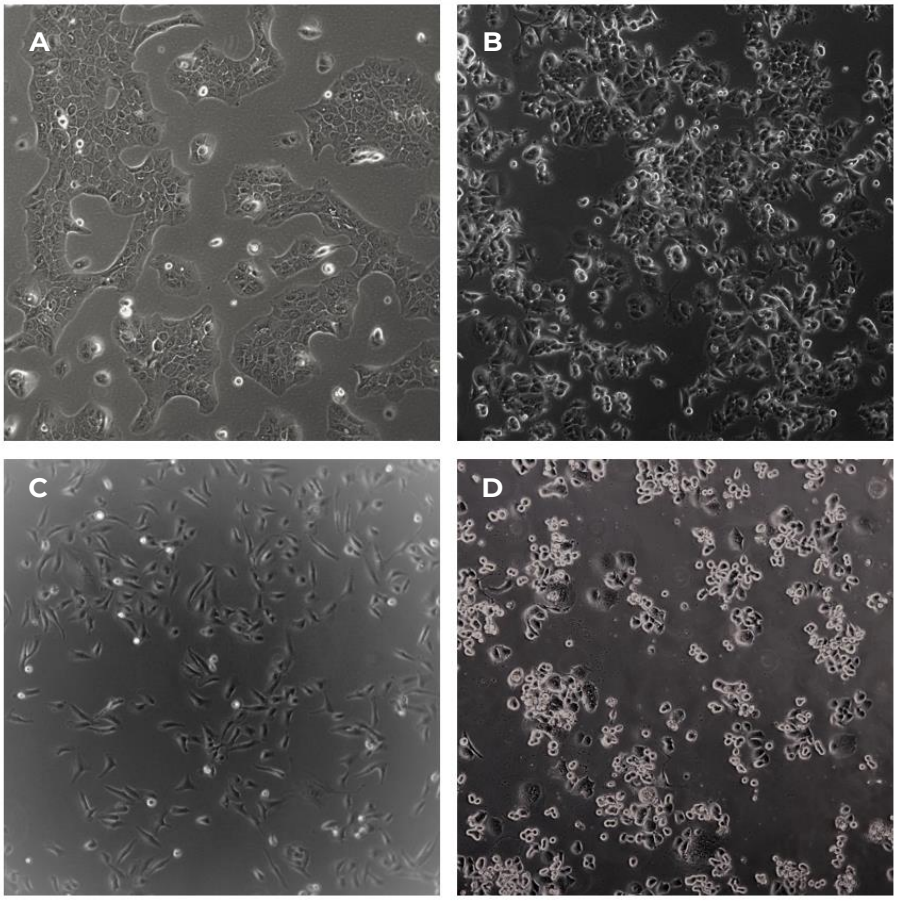


Figure 3.1 – Cell lines culture. **A-** MCF-10A cell line culture. **B-** MCF-7 cell line culture. **C-** MDA-MB-231 cell line culture. **D-** SK-BR-3 cell line culture. Pictures were taken with 100x ampliation.

3.1.2. RNA splicing analysis by PCR

RNA splicing assays can be performed with the final purpose of evaluating the pathogenicity of variants, being one of the factors involved in reclassification of VUS. In this case, to identify possible alternative splicing events for the different cell lines, the intention was to perform RNA splicing analysis of the exon 11 of the *BRCA1* gene by comparing the size of the amplified DNA fragments with the corresponding fragments of cDNA using the primers presented in Table 2.5. Also, comparing the size of the same cDNA fragments between the different cell lines would allow us to observe changes that would indicate the presence of such events, which in perspective could be caused by the presence of a potential SV in the DNA sequence.

For this, the DNA of each cell line cultured was extracted, as well as their total RNA, this last being reverse transcribed to cDNA allowing us to analyze the mature mRNA from each of the cell lines on an indirect way.

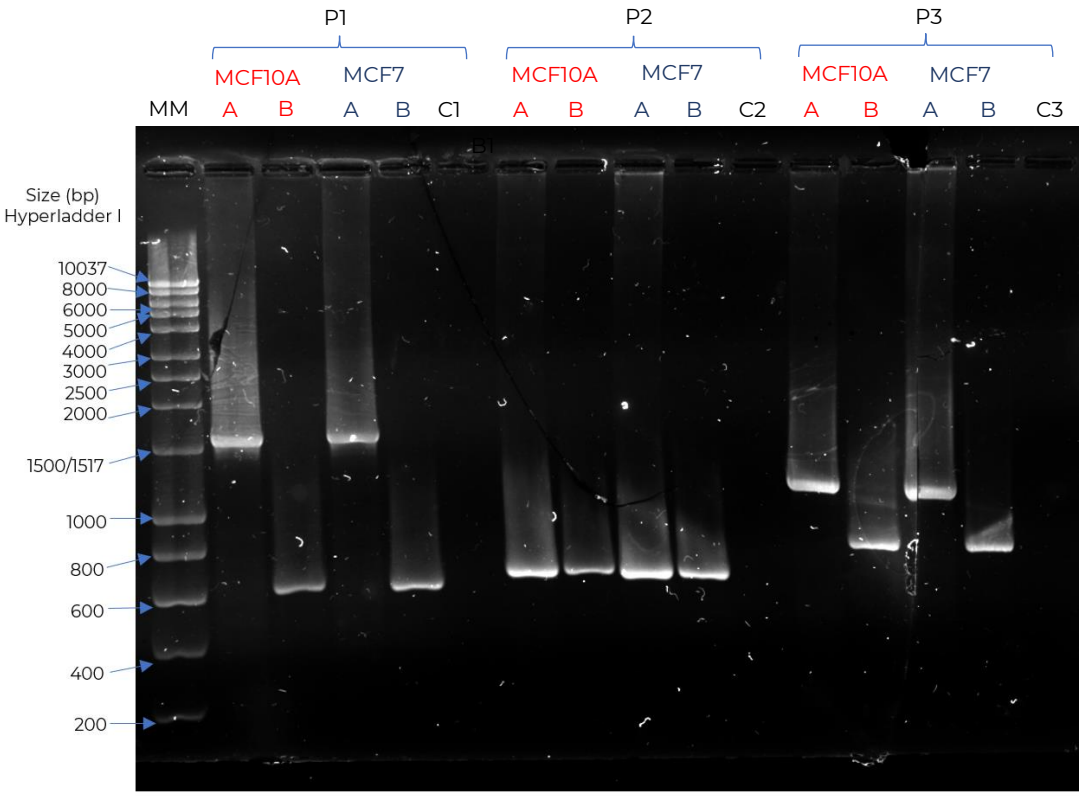


Figure 3.2 – RNA splice analysis of MCF-10A and MCF-7 cell lines. P1- fragments amplified with Mix 1. (Table 2.6); **P2**- fragments amplified with Mix 2. (Table 2.6); **P3**- fragments amplified with Mix 3. (Table 2.6). **A**- fragments amplified using the DNA extracted from the indicated cell line; **B**- fragments amplified using the cDNA obtained by reverse transcripatase of the RNA extracted from the indicated cell line. **C1**, **C2** and **C3** are the negative controls for each pair of primers.

As explained before, three pairs of primers were used in order to cover the majority of the exon 11 of *BRCA1* gene and its surroundings (Table 2.5). The sequences amplified with the different pair of primers for the DNA and for the cDNA differ due to splicing of the pre-mRNA (Figure 3.3).

Apparently, it is not observed any alteration on the size of the fragments of the cDNA between the different cell lines indicating no occurrence of alternative splicing events (Figure 3.2 and Figure 3.4).

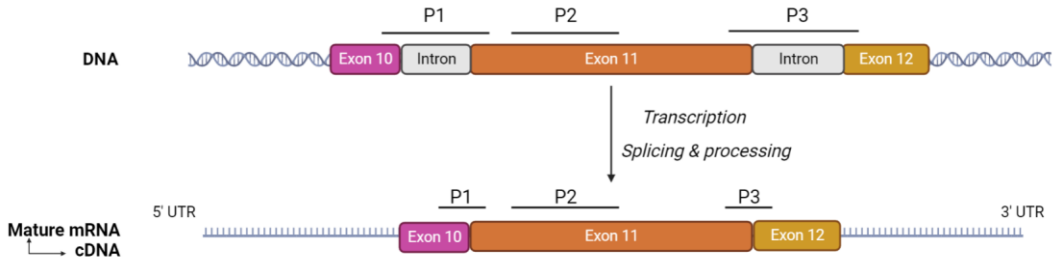


Figure 3.3 – Fragments applified by PCR with each pair of primers. Squematic representation of the applified fragments by PCR with each pair of primers for the **DNA** and for the **cDNA. P1 –** Represents the sequence covrage for the primer pair 1 (DNA- 1604 bp; cDNA- 619 bp); **P2 –** Sequence covrage for the the primer pair 2 (DNA- 682bp; cDNA 682 bp); **P3** – Sequence covrage for the primer pair 3 (DNA- 1210bp; cDNA- 850bp). (Created in: Biorender.com; bio).

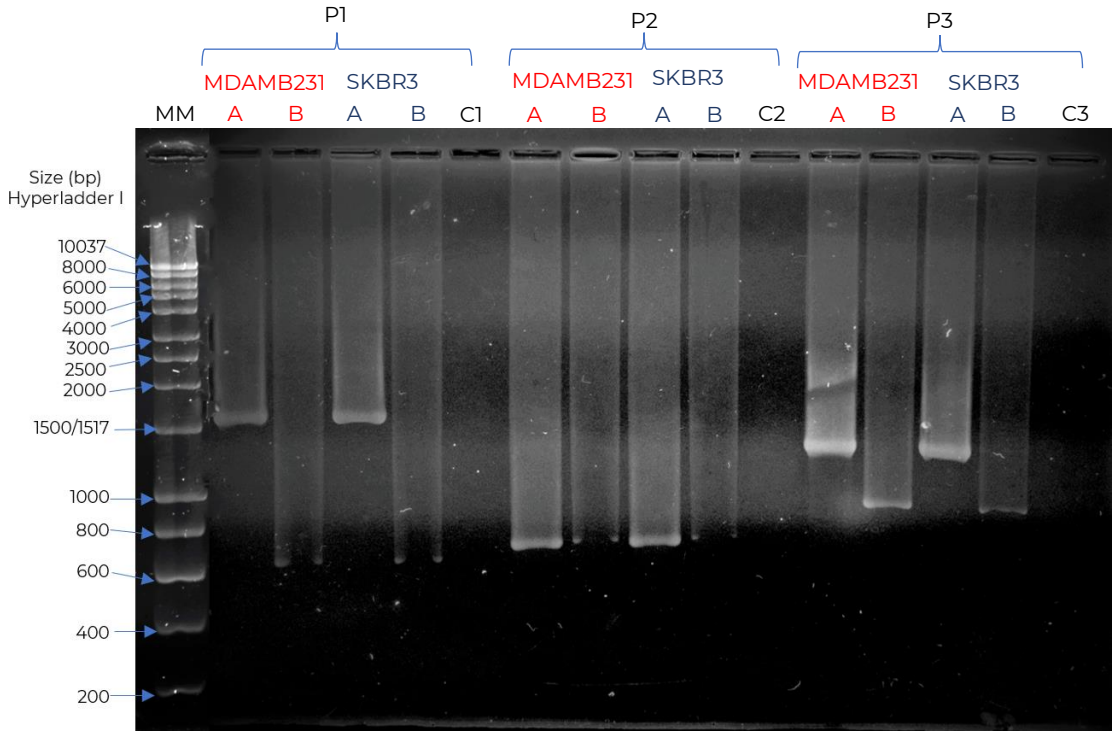


Figure 3.4 – RNA splice analysis of MDA-MB-231 and SK-BR-3 cell lines. P1- fragments amplified with Mix 1. (Table 2.6); **P2**- fragments amplified with Mix 2. (Table 2.6); **P3**- fragments amplified with Mix 3. (Table 2.6). **A**- fragments amplified using the DNA extracted from the indicated cell line; **B**- fragments amplified using the cDNA obtained by reverse transcripatase of the RNA extracted from the indicated cell line. **C1**, **C2** and **C3** are the negative controls for each pair of primers.

3.2. MCF-10A *BRCA1* VUS clone

For this task, 4 different attempts to introduce the *BRCA1* VUS (NM_007294.4:c.1067A>G) in the MCF -10A cell line with the CRISPR-Cas9 system for genome editing were, as mentioned, previously engaged by our group (Pedro, A., 2021). These possible clones were identified as B, F, G and K.

3.2.1. MCF-10A clone cell culture

With the purpose of establishing this *in vitro* model for the studies ahead, frozen cell suspensions were thawed and cultured as it was also proceeded for the WT MCF-10A cells. After culturing, the possible clone showed an initial delay in growth regarding to the WT MCF-10A. However, after a while the confluence of this culture spiked, achieving high confluence rates (Figure 3.5).

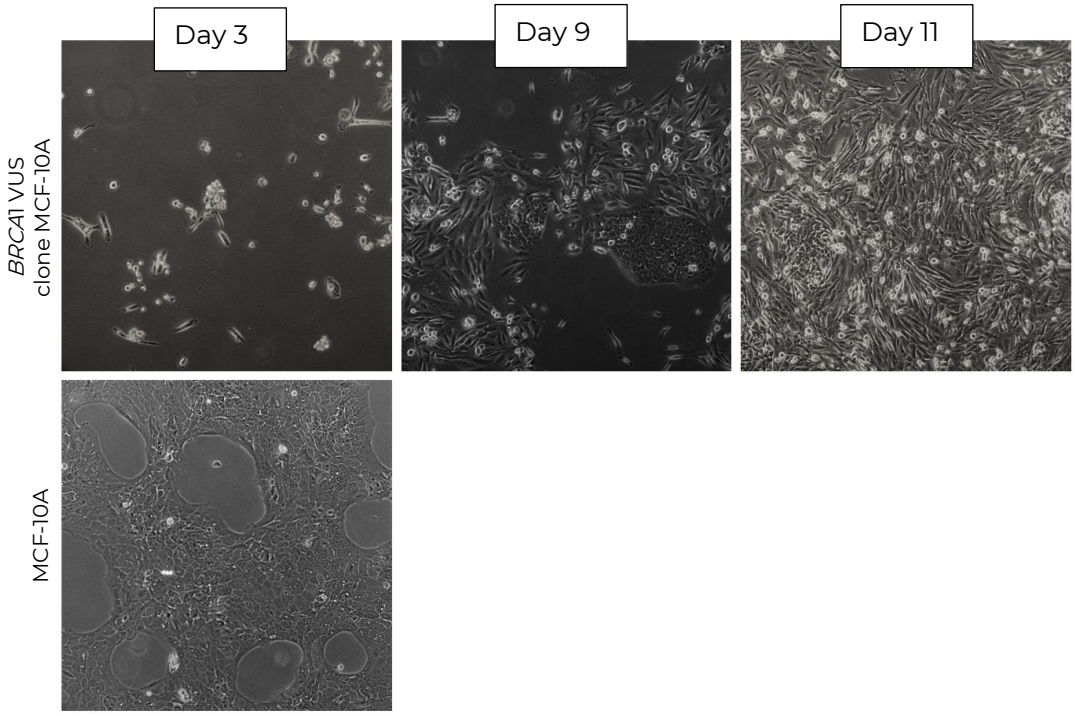


Figure 3.5 – WT MCF-10 and BRCA1 VUS clone culture. Pictures were taken in the indicated days after culture. On the first line we have the evolutional growth of the *BRCA1* VUS MCF-10A clone. On the second line we observe the same timeline for the MCF-10A culture. WT MCF-10A was tripsinized in day 5. Pictures were taken with 100x ampliation.

3.2.2. Confirmation of point mutation by PCR and Sequencing

In order to confirm that the genome edition was successful, the DNA of the 4 possible clones provided (B, F, G and K), as well as the DNA from WT MCF-10A cells, was extracted, and the target region of the *BRCA1* VUS was amplified by PCR. The amplified fragments were then observed on an agarose gel (Figure 3.5). Upon confirmation of the amplification of the VUS target region the PCR samples were sent to be sequenced. The following figures show the results for each step here described (Figure 3.6 and Figure 3.7).

The purified samples of the four possible clones were sent to be sequenced obtaining the results presented in the annexes. From the four, only the clone *F* was actually successfully cloned. The results show that this in *vitro model* harbors the mutation in both copies of the DNA, possessing the introduced VUS in homozygosity (Figure 3.7).

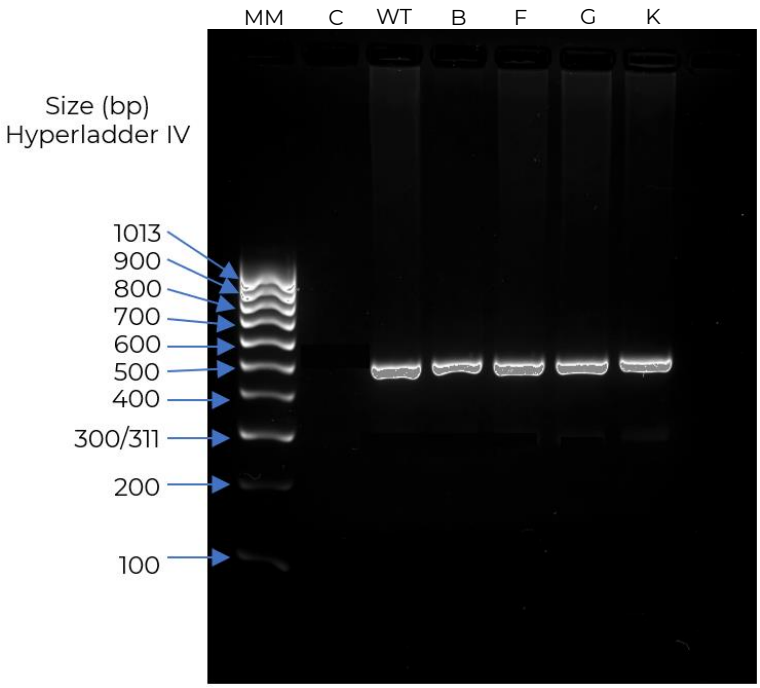


Figure 3.6 – BRCA1 VUS target region amplification. MM- molecular marker; **C**- negative control of the PCR (no DNA);**WT**- MCF-10A WT DNA; **B**- "clone" B; **F**- "clone" F; **G**- "clone" G; **K**- "clone" K.

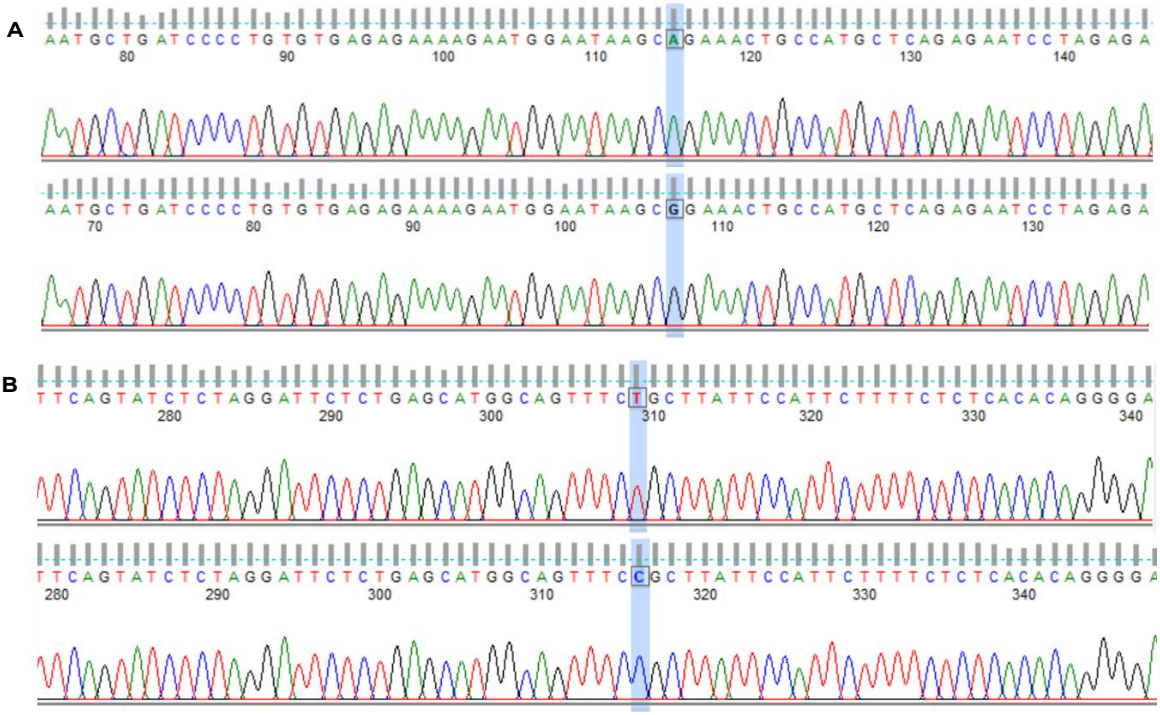


Figure 3.7 – Confirmation of BRCA1 VUS introduction in the MCF-10A cell line. A- Sequencing results for the forward sequence strand. **B**- Results for de reverse sequence strand. In A and B on the top are

presented the results of the WT MCF-10A and on the bottom are presented the results for the clone *F*.

3.2.3. Clone *F* RNA splicing analysis with PCR

In resemblance of what was done for the established cell lines, the same procedure for RNA splicing analysis was applied to this *in vitro* model that was successfully cloned with the VUS in study.

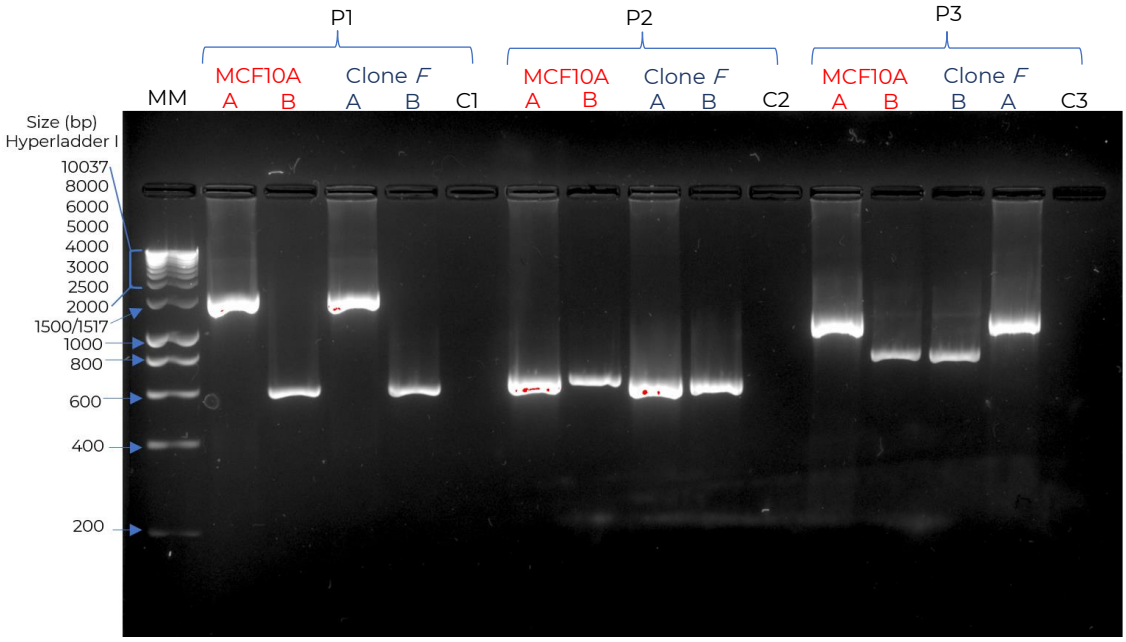


Figure 3.8 - RNA splice analysis of the clone F of the BRCA1 VUS compared to MCF-10A cell line. P1- fragments amplified with Mix 1. (Table 2.6); **P2**- fragments amplified with Mix 2. (Table 2.6); **P3**- fragments amplified with Mix 3. (Table 2.6). **A**- fragments amplified using the DNA extracted from the indicated cell line; **B**- fragments amplified using the cDNA obtained by reverse transcriptase of the RNA extracted from the indicated cell line. **C1**, **C2** and **C3** are the negative controls for each pair of primers.

Once again it was not observed any changes in the fragments size between the WT MCF-10A and the clone *F* indicating that normal splicing as occurred in both cell lines. As such, these results indicate that this particular variant is not a SV (Figure3.8).

3.3. Nanopore direct RNA Sequencing

For further confirmation of the RNA splicing assays applied to the VUS in study, we used a high-throughput technique that would allow us to screen even more accurately for SV and alternative splicing events happening in this gene. Here we used the direct RNA sequencing technique using the Minion Nanopore.

Nanopore direct RNA sequencing results, for both the clone *F* and the WT MCF-10A RNA, are still under analyze.

4. Discussion and Conclusions

Discussion and Conclusions

Cancer is a major health problem and worldwide incidence has been growing consistently.

Breast cancer, which is the type of cancer that this study focuses on, is currently the most common type of cancer, thus being target of many studies. In BC research, the choice of which model to use to a certain investigative work depends greatly on the focus of said research. In order to make that choice, it is important to know which models we have at our disposal and what characterizes them.

In vitro models can represent a step forward to a better insight of BC pathogenesis and for potential therapies through reliable assays. There are some drawbacks comparing 2D to 3D in vitro models such the lack of accuracy in recapitulate the BC microenvironment. However, this type of models detains many advantages comparing to human samples, allowing us to perform functional assays without the burden of all the time-consuming procedures and invasiveness involved in getting a sample from individuals. Additionally, these models allow us to replicate studies with much less obstacles. Thus, one of the proposes of our study is to address this issue by using distinct cell lines widely used in BC (Guiro and Arinzeh, 2015; Pedro, 2021) which gave us diversification in terms of BC subtypes.

The main goal of this project is focused on variants of unknown significance (VUS), more precisely splice variants (SV) found in highly penetrant genes in HBC, such as the *BRCA 1* gene.

HBC englobes a considerable part of all BC cases (5 to 10%) affecting, globally, many families. Most of these cases are due to inherited mutations in the *BRCA* genes. Patients that are *BRCA1* variant carriers tend to show poor prognosis and early onset of the disease (Pouptsis *et al.*, 2020). The study of these variants is important, not only to help the early diagnostic and prevention, but also to understand the mechanisms behind the development of the cancer and find the right therapies.

BRCA1 is a multifunctional protein involved in many pathways that ensure genomic stability to the cell. It is involved in DNA damage response as it appears to act as a sensor of DNA DSB in repair by homologous recombination (HR) and in cell cycle-checkpoint. Although it is proven by diverse studies and reports that the loss of a functional protein of this gene will lead to genome instability, the exact role it plays is still poorly understood, due to its complexity (Wu, Lu and Yu, 2010). In the BC susceptibility genetic tests that have been increasingly asked for *BRCA1* gene, as well as for others penetrant genes (multi gene panels), many variants have been found, with a considerable number of them being classified as VUS since their impact has

yet to be evaluated. These variants represent a source of concern, thus being important to understand their biological impact (Gustavsson, Galvis and Juth, 2020). Since some of these VUS are SV, they may alter the structure, the function and/or even the target localization of the gene product by affecting the splicing process of the pre-mRNA, consequently affecting the diverse pathways where the protein detains some functional role. On a clinical level, this variants are important but, usually more difficult to evaluate, especially if the proper functional assays were not performed since no alterations are immediately displayed as for other variants (Lindor *et al.*, 2013).

The outcome of the changes caused by a SV in the mRNA are the known alternative splicing events, which create a different isoform of the final transcript. It is important to highlight that alternative splicing doesn't only occur in the presence of a SV, being one of the factors that helps gene expression regulation in the different tissues. Therefore, such events may occur independently of a SV presence.

Here we propose the identification of this specific type of variants in the exon 11 of the *BRCA1* gene by PCR-based techniques and with the high-throughput technique of direct RNA sequencing by nanopore. The choice of this region of the *BRCA1* gene is because, exon 11 is described in the literature as a highly alternative spliced region of the *BRCA1* gene, probably because of the length of this exon. Also, it should be highlighted that it is in this exon that the coding sequences to the protein signal for nuclear localization are found (Ruiz de Garibay *et al.*, 2021). If this sequence is disrupted many processes should be affected causing the genomic instability mentioned before.

Therefore, we firstly used the MCF-10A cell line as a representative of the normal breast tissue for comparative purposes as it is a non-tumorigenic *in vitro* model, while maintaining the general characteristics of the remaining tumorigenic cell lines. The three established tumorigenic cell lines used were the MCF-7, the MDA-MB-231 and the SK-BR-3. In resemblance to the MCF-10A, they are all adherent, immortal, epithelial human breast cell lines, but each one is classified with a distinct BC subtype in terms of molecular characteristics (Puleo and Polyak, 2021).

Using these *in vitro* models, we started by culturing them, obtaining cultures with different morphologies and confluence rates (Figure 3.1). The MCF-10A culture, being a non-tumorigenic cell line, clearly demonstrated a slower confluence rate than the others. These confluence rates are affected by diverse factors seen as in their absence they alter. The cells reported to have higher metastatic potential, the MDA-MB-231 and the SK-BR-3, achieved the intended confluence more quickly than the MCF-7 although, this last also displayed a high confluence rate.

In order to screen the DNA of these cell lines for SV, the DNA and RNA were extracted from the same culture of each cell line, and PCR primers were designed to identify the presence of alternative splice events when observed in the gel.

The different pairs of primers were carefully chosen for that, the sequence coverage delimited by them, included the exon-intron boundary, before and after the exon 11 (Figure 3.3) (primers exact localization in the annexes). SV are more frequently found in the intronic region of the gene near the exon-intron boundary, although there are already evidences that suggest that mutations in exonic regions where ESEs and ESSs are found, can also incur in alternative splicing. Therefore, it was important in this project to focus in all these regions of the gene (Bergsma *et al.*, 2018; Riehn *et al.*, 2011).

In the presence of a SV within the target region covered by the pairs of primers the splicing would have been abnormal and alternative splicing events such as exon skipping, total or partial intron retention, creation of new splice sites among others (Figure 1.5), would occur and be evidenced by a change in the size of the amplified fragment of the cDNA. However, we could not be sure that an event like that would be caused by a SV since alternative splicing also occurs with a high frequency on a spontaneously mode in order to regulate gene expression in diverse tissues (Lixia *et al.*, 2007).

Despite the predictive tools used in the primer design (IDT- Oligo Analyzer Tool) indicate gaps in the melting temperatures (Tm°C) predicted for each primer pair during optimization of PCR, it was concluded that the primers would amplify all using the same optimal conditions (Table 2.7).

Indeed, no differences were found in the size of the cDNA fragments amplified for all the different cell lines (Figure 3.2 and Figure3.4), therefore indicating that non alternative splicing event occurred, allowing us to incur that there are no SV in either side of the exon 11, and also that these *in vitro* models do not undergo spontaneous alternative splicing in this region of the gene. Also, this results confirm previously reported data for these cell lines, which stated that no *BRCA1* variants were found (Dai *et al.*, 2017). Although highly improbably, what may pass unnoticed in this assay is the occurrence of partial “substitution” of the sequence (intronic retention and partial exon skipping happening at the same time basically) since it was not in the primer annealing region, as the fragments size would not be apparently altered. In order to obtain the best possible results for this method it is important for the PCR product concentration for the cDNA not to be high since, small changes in the final transcript could also be unnoticed and not perceptible in the gel.

Further in our study we focused on a specific VUS that had been previously studied by our group, present in exon 11 of the *BRCA1* gene. As explained this VUS was found in two healthy individuals with family history when screened for a multigene panel of penetrant genes. The importance of multigene panels is because they give a larger view of interactions and associations between major and minor penetrant genes that would not be seen if it was a genetic test for a single gene only. In the study performed they took interest in this specific VUS because in some *in silico* predictive tools it was classified as probably damaging, going against other reports that classified it as benign or likely benign. This is a missense SNV that changes the amino acid glutamine by one arginine in the position 356 of the protein (Lourenço, 2018).

For our current study we initially cultured MCF-10A cells that were previously genome edited by the CRISPR-Cas9 system in order to introduce this VUS. To assess the success of this transformation we cultured them in parallel with MCF-10A WT cells as shown in Figure 3.5. As stated, initially the potential clone displayed a slow growth, and small changes in morphology. At day 9 it was possible to distinguish two distinct cell morphologies. Some still displayed a similar morphology to the WT MCF-10A culture, while others clearly differ from those. After the initial delay the cells display much higher confluence rates with distinct morphology from the MCF-10A WT cell line, indicating perhaps, already in early stages, the success of the cloning.

To confirm the presence of the point mutation we amplified the region of interest (where the VUS was found) and then we sent the samples to be sequenced. The results presented in Figure 3.7 confirm the presence of the mutation for the clone F only, being this the one that was cultured, and the cells are displayed in Figure 3.5. Meanwhile the other clones did not show any alteration on the sequence comparing to the WT MCF-10A (results in the annexes). The obtained results for both copies also showed that the VUS was present in homozygosity as the mutation was found in both strands of the DNA. It is important to highlight this fact once functional studies had already been applied to one of the clones obtained by CRISPR-Cas9 but that one, contrary to this was only mutated in one of the DNA strands. It would be of interest to apply the same functional studies previously done and applied for the heterozygous clone, now to this new homozygous *in vitro* model helping us to further characterize it.

After once again extracted the DNA and the RNA from the same culture, the RNA was again reverse transcribed to cDNA and the same analysis made for the different *in vitro* models was now applied to the cell line with this VUS. The results were once again null as no alternative splicing events have occurred, therefore indicating that this is most probably not a SV.

All this information taken together allowed us to retrieve new information concluding that splicing assays are without doubt important to understand the mechanisms behind tumorigeneses specially in highly penetrant genes such as *BRCA1* and that this kind of studies may have an impact on clinical management of BC patients. It also it gave us perspectives for further studies.

4.1. Future Perspectives

To aggregate credibility to this assay, it would also be of interest to apply this same PCR based method to already classified SVs and to other cell lines reported to have variants in this gene.

On that note, further studies of more accurate techniques such as the nanopore direct RNA sequencing, which is a high-throughput method, although with some reported margin of error, could be of use in this type of RNA splicing analysis, especially because limitations such as only having a partial transcript view would be overcome. It would give us a major panorama of the all transcript of the *BRCA1* for example (Jong, de *et al.*, 2017).

As explained before the first technique here described for splicing analysis allows to evaluate possible alterations in the RNA transcript of the *BRCA1* gene, meanwhile, as the opportunity was presented, we were able to engage the direct RNA sequencing method by the MinION nanopore for the *BRCA1* clone RNA, that should allow us to identify full-length splice isoforms. In this assay an enrichment method was applied in the second day after the first run. Chang. *et al* stated that this enrichment could be a potential source of artifacts in the results due to the high number of manipulations it implies. Also, it has been reported that these types of experiments are error prone on top of the introduced human error (Chang, 2018). As that was only recently performed, data is still being analyzed.

5. References

References

AHMAD, Aamir - **Breast Cancer Metastasis and Drug Resistance: Challanges and Progress**. ISBN 978-3-030-20301-6. .

ATCC - **BREAST CANCER P53 HOTSPOT MUTATION CELL PANEL** . 2012. Available in WWW:<URL:https://www.atcc.org/-/media/product-assets/documents/panels/cell-biology/breast-cancer-p53-hotspot-mutation-cell panel.pdf?rev=5f3759c84c934c509cde17c2b190ca15>.

BERGSMA, Atze J. *et al.* - Alternative Splicing in Genetic Diseases: Improved Diagnosis and Novel Treatment Options. In **International Review of Cell and Molecular Biology** [Online]. 1. ed. [S.l.] : Elsevier Inc., 2018 Available at WWW:<URL:http://dx.doi.org/10.1016/bs.ircmb.2017.07.008>.v. 335. p. 85–141.

BURKE, Wylie - Genetic Testing. **New England Journal of Medicine**. 347:23 (2002) 1867–1875. doi: 10.1056/NEJMoa012113.

CAMACHO LONDOÑO, Julia; PHILIPP, Stephan E. - A reliable method for quantification of splice variants using RT-qPCR. **BMC Molecular Biology**. . ISSN 14712199. 17:1 (2016) 1–12. doi: 10.1186/s12867-016-0060-1.

CHAFFER, Christine L.; WEINBERG, Robert A. - A perspective on cancer cell metastasis. **Science**. . ISSN 00368075. 331:6024 (2011) 1559–1564. doi: 10.1126/science.1203543.

CHANG, Jenny *et al.* - Clinical Management of Patients at Risk for Hereditary Breast Cancer with Variants of Uncertain Significance in the Era of Multigene Panel Testing. **Annals of Surgical Oncology**. . ISSN 15344681. 26:10 (2019) 3389–3396. doi: 10.1245/s10434-019-07595-2.

CHANG, Youen J. J. - Targeted Enrichment and Nanopore Sequencing of RNA Transcripts from Breast Cancer Risk Genes. October (2018).

COMŞA, Şerban; CÎMPEAN, Anca Maria; RAICA, Marius - The Story of MCF-7 Breast Cancer Cell Line: 40 years of Experience in Research. **Anticancer research**. . ISSN 0250-7005. 35:6 (2015) 3147—3154.

CONCOLINO, Paola *et al.* - BRCA1 and BRCA2 testing through next generation sequencing in a small cohort of italian breast/ovarian cancer patients: Novel pathogenic and unknown clinical significance variants. **International Journal of Molecular Sciences**. . ISSN 14220067. 20:14 (2019) 1–16. doi: 10.3390/ijms20143442.

CURTIT, Elsa *et al.* - First description of a sporadic breast cancer in a woman with BRCA1 germline mutation. **Oncotarget**. . ISSN 19492553. 6:34 (2015) 35616–35624. doi: 10.18632/oncotarget.5348.

DAI, Xiaofeng *et al.* - Breast cancer cell line classification and Its relevance with breast tumor subtyping. **Journal of Cancer**. . ISSN 18379664. 8:16 (2017) 3131–3141. doi: 10.7150/jca.18457.

ECACC - Cell line profile. **ECACC, European Collection of Authenticated Cell Cultures**. 4:92020424 (2022) 1–2.

ECCLES, B. K. *et al.* - Understanding of BRCA VUS genetic results by breast cancer specialists. **BMC Cancer**. . ISSN 14712407. 15:1 (2015) 1–9. doi: 10.1186/s12885-015-1934-1.

GIAQUINTO, Angela N. *et al.* - Breast Cancer Statistics, 2022. **CA: A Cancer Journal for Clinicians**. . ISSN 0007-9235. 72:6 (2022) 524–541. doi: 10.3322/caac.21754.

GODET, Inês; M. GILKES, Daniele - BRCA1 and BRCA2 mutations and treatment strategies for breast cancer. **Integrative Cancer Science and Therapeutics**. . ISSN 2056-4546. 4:1 (2017) 1–7. doi: 10.15761/icst.1000228.

GUIRO, Khadidiatou; ARINZEH, Treena L. - Bioengineering Models for Breast Cancer Research. **Breast Cancer: Basic and Clinical Research**. . ISSN 11782234. 9s2:2015) 57–70. doi: 10.4137/BCBCR.S29424.

GUSTAVSSON, Erik; GALVIS, Giovanni; JUTH, Niklas - Genetic testing for breast cancer risk, from BRCA1/2 to a seven gene panel: an ethical analysis. **BMC Medical Ethics**. . ISSN 14726939. 21:1 (2020) 1–8. doi: 10.1186/s12910-020-00545-8.

J.Y., Cheon; J., Mozersky; R., Cook-Deegan - Variants of uncertain significance in BRCA: A harbinger of ethical and policy issues to come? **Genome Medicine**. . ISSN 1756-994X. 6:12 (2014) 1–10.

JIANG, Guanglong *et al.* - Comprehensive comparison of molecular portraits between cell lines and tumors in breast cancer. **BMC Genomics**. . ISSN 14712164. 17:Suppl 7 (2016). doi: 10.1186/s12864-016-2911-z.

JIANG, Yue *et al.* - Identification of Recurrent Variants in BRCA1 and BRCA2 across Multiple Cancers in the Chinese Population. **BioMed Research International**. . ISSN 23146141. 2020:2020). doi: 10.1155/2020/6739823.

JONG, Lucy C. DE *et al.* - Nanopore sequencing of full-length BRCA1 mRNA transcripts reveals co-occurrence of known exon skipping events. **Breast Cancer Research**. . ISSN 1465542X. 19:1 (2017) 1–9. doi: 10.1186/s13058-017-0919-1.

KAO, Jessica *et al.* - Molecular profiling of breast cancer cell lines defines relevant tumor models and provides a resource for cancer gene discovery. **PLoS ONE**. . ISSN 19326203. 4:7 (2009). doi: 10.1371/journal.pone.0006146.

KREEGER, Pamela K.; LAUFFENBURGER, Douglas A. - Cancer systems biology : a network modeling perspective. 31:1 (2010) 2–8. doi: 10.1093/carcin/bgp261.

KRUM, Susan A. *et al.* - BRCA1 forms a functional complex with γ-H2AX as a late response to genotoxic stress. **Journal of Nucleic Acids**. . ISSN 20900201. 2010:2010). doi: 10.4061/2010/801594.

LINDOR, Noralane M. *et al.* - BRCA1/2 Sequence Variants of Uncertain Significance: A Primer for Providers to Assist in Discussions and in Medical Management. **The Oncologist**. . ISSN 1083-7159. 18:5 (2013) 518–524. doi: 10.1634/theoncologist.2012-0452.

LIXIA, Miao *et al.* - Alternative splicing of breast cancer associated gene BRCA1 from breast cancer cell line. **Journal of Biochemistry and Molecular Biology**. . ISSN 02191024. 40:1 (2007) 15–21. doi: 10.5483/bmbrep.2007.40.1.015.

LOURENÇO, Rita Adubeiro - **Rita Adubeiro Lourenço Functional Characterization of Variants of Unknown Significance in Familial Breast Cancer**. [S.l.] : Universidade Nova de Lisboa, 2018

NAROD, Steven A.; FOULKES, William D. - BRCA1 and BRCA2: 1994 and beyond. **Nature Reviews Cancer**. . ISSN 1474175X. 4:9 (2004) 665–676. doi: 10.1038/nrc1431.

ORBAN, T. I.; OLAH, E. - Emerging roles of BRCA1 alternative splicing. **Journal of Clinical Pathology - Molecular Pathology**. . ISSN 13668714. 56:4 (2003) 191–197. doi: 10.1136/mp.56.4.191.

ORR, Katy S.; SAVAGE, Kienan I. - The BRCA1 and BRCA2 Breast and Ovarian Cancer Susceptibility Genes — Implications for DNA Damage Response, DNA Repair and Cancer Therapy. In CHEN, CLARK C. (Ed.) - **Advances in DNA Repair** [Online]. Rijeka : IntechOpen, 2015 Available at WWW:<URL:https://doi.org/10.5772/59996>.

PEDRO, Ana Rita Serra - **Genome editing to be applied in functional characterization of genetic variants in familial breast cancer patients**. [S.l.] : Universidade Nova de Lisboa, 2021

POUPTSIS, Athanasios *et al.* - Surgical and Systemic Treatment of Hereditary Breast Cancer: A Mini-Review With a Focus on BRCA1 and BRCA2 Mutations. **Frontiers in Oncology**. . ISSN 2234943X. 10:October (2020) 1–7. doi: 10.3389/fonc.2020.553080.

PULEO, Julieann; POLYAK, Kornelia - The MCF10 model of breast tumor progression. **Cancer Research**. . ISSN 15387445. 81:16 (2021) 4183–4185. doi: 10.1158/0008-5472.CAN-21-1939.

RIEHN, Mathias *et al.* - A BACH2-BCL2L1 Fusion Gene Resulting from a Lymphoma Cell Line BLUE-1. **Cancer**. 396:January (2011) 389–396. doi: 10.1002/gcc.

ROSENFELD, Simon - Are the somatic mutation and tissue organization field theories of carcinogenesis incompatible? **Cancer Informatics**. . ISSN 11769351. 12:2013) 221–229. doi: 10.4137/CIn.s13013.

RUIZ DE GARIBAY, Gorka *et al.* - Altered regulation of BRCA1 exon 11 splicing is associated with breast cancer risk in carriers of BRCA1 pathogenic variants. **Human mutation**. United States. . ISSN 1098-1004 (Electronic). 42:11 (2021) 1488–1502. doi: 10.1002/humu.24276.

SCULLY, Olivia Jane *et al.* - Breast Cancer Metastasis. **Cancer Genomics - Proteomics**. 9:5 (2012) 311 LP-320.

SULLIVAN, Teresa *et al.* - Epidemiological and ES cell-based functional evaluation of BRCA2 variants identified in families with breast cancer. **Human Mutation**. . ISSN 10981004. 42:2 (2021) 200–212. doi: 10.1002/humu.24154.

TARSOUNAS, Madalena; SUNG, Patrick - The antitumorigenic roles of BRCA1–BARD1 in DNA repair and replication. **Nature Reviews Molecular Cell Biology**. . ISSN 14710080. 21:5 (2020) 284–299. doi: 10.1038/s41580-020-0218-z.

WANGENSTEEN, Teresia *et al.* - Diagnostic mRNA splicing assay for variants in BRCA1 and BRCA2 identified two novel pathogenic splicing aberrations. 9:2019) 1–8.

WRIGHT, Moriah *et al.* - The American Journal of Surgery Factors predicting reclassi fi cation of variants of unknown signi fi cance. 216:2018) 1148–1154.

WU, Jiaxue; LU, Lin Yu; YU, Xiaochun - The role of BRCA1 in DNA damage response. **Protein and Cell**. . ISSN 16748018. 1:2 (2010) 117–123. doi: 10.1007/s13238-010-0010-5.

XIPING, Zhang *et al.* - A summary of relationships between alternative splicing and breast cancer. **Oncotarget**. . ISSN 19492553. 8:31 (2017) 51986–51993. doi: 10.18632/oncotarget.17727.

YANG, Quan *et al.* - Aberrant alternative splicing in breast cancer. **Journal of Molecular Cell Biology**. . ISSN 17594685. 11:10 (2019) 920–929. doi: 10.1093/jmcb/mjz033.

6. Annexes

Homo sapiens BRCA1 DNA repair associated (BRCA1), transcript variant 1, mRNA (CDS)

NCBI Reference Sequence: NM_007294.4

GenBank Graphics

>NM_007294.4:114-5705 Homo sapiens BRCA1 DNA repair associated (BRCA1), transcript variant 1, mRNA

ATGGATTTATCTGCTCTTCGCGTTGAAGAAGTACAAAATGTCATTAATGCTATGCAGAAAATCTTAGAGT
GTCCCATCTGTCTGGAGTTGATCAAGGAACCTGTCTCCACAAAGTGTGACCACATATTTTGCAAATTTTG
CATGCTGAAACTTCTCAACCAGAAGAAAGGGCCTTCACAGTGTCCTTTATGTAAGAATGATATAACCAAA
AGGAGCCTACAAGAAAGTACGAGATTTAGTCAACTTGTTGAAGAGCTATTGAAAATCATTTGTGCTTTTC
AGCTTGACACAGGTTTGGAGTATGCAAACAGCTATAATTTTGCAAAAAAGGAAAATAACTCTCCTGAACA
TCTAAAAGATGAAGTTTCTATCATCCAAAGTATGGGCTACAGAAACCGTGCCAAAAGACTTCTACAGAGT
GAACCCGAAAATCCTTCCTTGCAGGAAACCAGTCTCAGTGTCCAACTCTCTAACCTTGGAACTGTGAGAA
CTCTGAGGACAAAGCAGCGGATACAACCTCAAAAGACGTCTGTCTACATTGAATTGGGATCTGATTCTTC
TGAAGATACCGTTAATAAGGCAACTTATTGCAGTGTGGGAGATCAAGAATTGTTACAAATCACCCCTCAA
GGAACCAGGGATGAAATCAGTTTGGATTCTGCAAAAAAGGCTGCTTGTGAATTTTCTGAGACGGATGTAA
CAAATACTGAACATCATCAACCCAGTAATAATGATTTGAACACCACTGAGAAGCGTGCAGCTGAGAGGCA
TCCAGAAAAGTATCAGGGTAGTTCTGTTTCAAACTTGCATGTGGAGCCATGTGGCACAAATACTCATGCC
AGCTCATTACAGCATGAGAACAGCAGTTTATTACTCACTAAAGACAGAATGAATGTAGAAAAGGCTGAAT
TCTGTAATAAAAGCAAACAGCCTGGCTTAGCAAGGAGCCAACATAACAGATGGGCTGGAAGTAAGGAAAC
ATGTAATGATAGGCGGACTCCCAGCACAGAAAAAAAGGTAGATCTGAATGCTGATCCCCTGTGTGAGAGA
AAAGAATGGAATAAGCAGAAACTGCCATGCTCAGAGAATCCTAGAGATACTGAAGATGTTCCTTGGATAA
CACTAAATAGCAGCATTCAGAAAGTTAATGAGTGGTTTTCCAGAAGTGATGAACTGTTAGGTTCTGATGA
CTCACATGATGGGGAGTCTGAATCAAATGCCAAAGTAGCTGATGTATTGGACGTTCTAAATGAGGTAGAT
GAATATTCTGGTTCTTCAGAGAAAATAGACTTACTGGCCAGTGATCCTCATGAGGCTTTAATATGTAAAA
GTGAAAGAGTTCACTCCAAATCAGTAGAGAGTAATATTGAAGACAAAATATTTGGGAAAACCTATCGGAA
GAAGGCAAGCCTCCCCAACTTAAGCCATGTAACTGAAAATCTAATTATAGGAGCATTTGTTACTGAGCCA
CAGATAATACAAGAGCGTCCCCTCACAAATAAATTAAAGCGTAAAAGGAGACCTACATCAGGCCTTCATC
CTGAGGATTTTATCAAGAAAGCAGATTTGGCAGTTCAAAAGACTCCTGAAATGATAAATCAGGGAACTAA
CCAAACGGAGCAGAATGGTCAAGTGATGAATATTACTAATAGTGGTCATGAGAATAAAACAAAAGGTGAT
TCTATTCAGAATGAGAAAAATCCTAACCCAATAGAATCACTCGAAAAAGAATCTGCTTTCAAAACGAAAG
CTGAACCTATAAGCAGCAGTATAAGCAATATGGAACTCGAATTAAATATCCACAATTCAAAAGCACCTAA
AAAGAATAGGCTGAGGAGGAAGTCTTCTACCAGGCATATTCATGCGCTTGAACTAGTAGTCAGTAGAAAT
CTAAGCCCACCTAATTGTACTGAATTGCAAATTGATAGTTGTTCTAGCAGTGAAGAGATAAAGAAAAAAA
AGTACAACCAAATGCCAGTCAGGCACAGCAGAAACCTACAACTCATGGAAGGTAAAGAACCTGCAACTGG
AGCCAAGAAGAGTAACAAGCCAAATGAACAGACAAGTAAAAGACATGACAGCGATACTTTCCCAGAGCTG
AAGTTAACAAATGCACCTGGTTCTTTTACTAAGTGTTCAAATACCAGTGAACTTAAAGAATTTGTCAATC
CTAGCCTTCCAAGAGAAGAAAAAGAAGAGAAACTAGAAACAGTTAAAGTGTCTAATAATGCTGAAGACCC
CAAAGATCTCATGTTAAGTGGAGAAAGGGTTTTGCAAACTGAAAGATCTGTAGAGAGTAGCAGTATTTCA
TTGGTACCTGGTACTGATTATGGCACTCAGGAAAGTATCTCGTTACTGGAAGTTAGCACTCTAGGGAAGG
CAAAAACAGAACCAAATAAATGTGTGAGTCAGTGTGCAGCATTTGAAAACCCCAAGGGACTAATTCATGG

TTGTTCCAAAGATAATAGAAATGACACAGAAGGCTTTAAGTATCCATTGGGACATGAAGTTAACCACAGT
CGGGAAACAAGCATAGAAATGGAAGAAAGTGAACTTGATGCTCAGTATTTGCAGAATACATTCAAGGTTT
CAAAGCGCCAGTCATTTGCTCCGTTTTCAAATCCAGGAAATGCAGAAGAGGAATGTGCAACATTCTCTGC
CCACTCTGGGTCCTTAAAGAAACAAAGTCCAAAAGTCACTTTTGAATGTGAACAAAAGGAAGAAAATCAA
GGAAAGAATGAGTCTAATATCAAGCCTGTACAGACAGTTAATATCACTGCAGGCTTTCCTGTGGTTGGTC
AGAAAGATAAGCCAGTTGATAATGCCAAATGTAGTATCAAAGGAGGCTCTAGGTTTTGTCTATCATCTCA
GTTCAGAGGCAACGAAACTGGACTCATTACTCCAAATAAACATGGACTTTTACAAAACCCATATCGTATA
CCACCACTTTTTCCCATCAAGTCATTTGTTAAAACTAAATGTAAGAAAAATCTGCTAGAGGAAAACTTTG
AGGAACATTCAATGTCACCTGAAAGAGAAATGGGAAATGAGAACATTCCAAGTACAGTGAGCACAATTAG
CCGTAATAACATTAGAGAAAATGTTTTTAAAGAAGCCAGCTCAAGCAATATTAATGAAGTAGGTTCCAGT
ACTAATGAAGTGGGCTCCAGTATTAATGAAATAGGTTCCAGTGATGAAAACATTCAAGCAGAACTAGGTA
GAAACAGAGGGCCAAAATTGAATGCTATGCTTAGATTAGGGGTTTTGCAACCTGAGGTCTATAAACAAAG
TCTTCCTGGAAGTAATTGTAAGCATCCTGAAATAAAAAAGCAAGAATATGAAGAAGTAGTTCAGACTGTT
AATACAGATTTCTCTCCATATCTGATTTCAGATAACTTAGAACAGCCTATGGGAAGTAGTCATGCATCTC
AGGTTTGTTCTGAGACACCTGATGACCTGTTAGATGATGGTGAAATAAAGGAAGATACTAGTTTTGCTGA
AAATGACATTAAGGAAAGTTCTGCTGTTTTTAGCAAAAGCGTCCAGAAAGGAGAGCTTAGCAGGAGTCCT
AGCCCTTTCACCCATACACATTTGGCTCAGGGTTACCGAAGAGGGGCCAAGAAATTAGAGTCCTCAGAAG
AGAACTTATCTAGTGAGGATGAAGAGCTTCCCTGCTTCCAACACTTGTTATTTGGTAAAGTAAACAATAT
ACCTTCTCAGTCTACTAGGCATAGCACCGTTGCTACCGAGTGTCTGTCTAAGAACACAGAGGAGAATTTA
TTATCATTGAAGAATAGCTTAAATGACTGCAGTAACCAGGTAATATTGGCAAAGGCATCTCAGGAACATC
ACCTTAGTGAGGAAACAAAATGTTCTGCTAGCTTGTTTTCTTCACAGTGCAGTGAATTGGAAGACTTGAC
TGCAAATACAAACACCCAGGATCCTTTCTTGATTGGTTCTTCCAAACAAATGAGGCATCAGTCTGAAAGC
CAGGGAGTTGGTCTGAGTGACAAGGAATTGGTTTCAGATGATGAAGAAAGAGGAACGGGCTTGGAAGAAA
ATAATCAAGAAGAGCAAAGCATGGATTCAAACTTAGGTGAAGCAGCATCTGGGTGTGAGAGTGAAACAAG
CGTCTCTGAAGACTGCTCAGGGCTATCCTCTCAGAGTGACATTTTAACCACTCAGCAGAGGGATACCATG
CAACATAACCTGATAAAGCTCCAGCAGGAAATGGCTGAACTAGAAGCTGTGTTAGAACAGCATGGGAGCC
AGCCTTCTAACAGCTACCCTTCCATCATAAGTGACTCTTCTGCCCTTGAGGACCTGCGAAATCCAGAACA
AAGCACATCAGAAAAAGCAGTATTAACTTCACAGAAAAGTAGTGAATACCCTATAAGCCAGAATCCAGAA
GGCCTTTCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCAGTAAAAATAAAGAACCAGGAGTGG
AAAGGTCATCCCCTTCTAAATGCCCATCATTAGATGATAGGTGGTACATGCACAGTTGCTCTGGGAGTCT
TCAGAATAGAAACTACCCATCTCAAGAGGAGCTCATTAAGGTTGTTGATGTGGAGGAGCAACAGCTGGAA
GAGTCTGGGCCACACGATTTGACGGAAACATCTTACTTGCCAAGGCAAGATCTAG

ABC – Exon 11

ABC – Exon 10 e 12

ABC – primer pair 1 - sequence coverage of 619bp

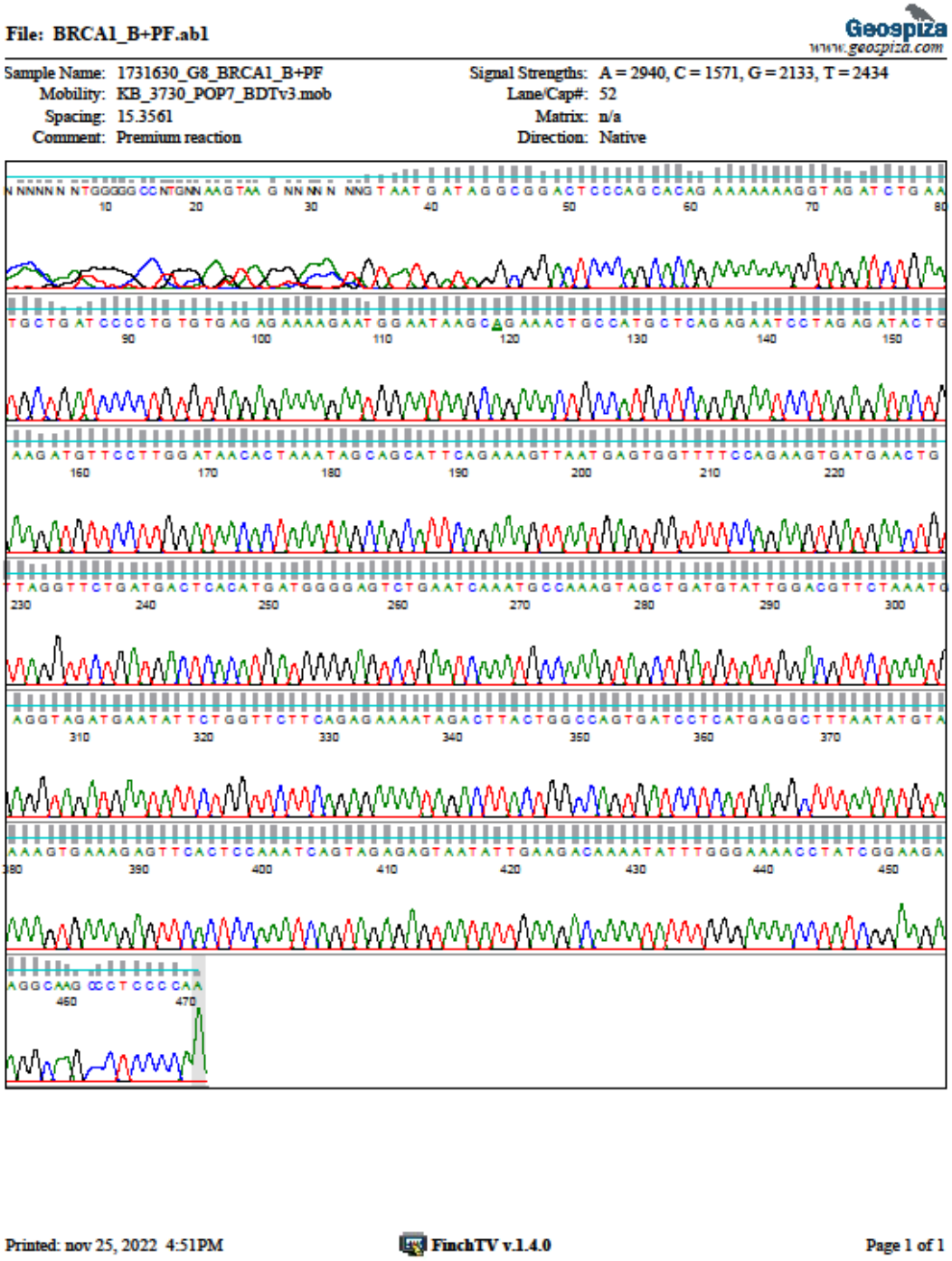
ABC – primer pair 2 –sequence coverage of 682bp

ABC – primer pair 3 –sequence coverage of 850bp

BRCA 1 gene, Exon 11 – Primers for VUS target region (489bp)

```
      ctgcttg tgaattttct gagacggatg taacaaatac tgaacatcat caacccagta
 841 ataatgattt gaacaccact gagaagcgtg cagctgagag gcatccagaa aagtatcagg
 901 gtagttctgt ttcaaacttg catgtggagc catgtggcac aaatactcat gccagctcat
 961 tacagcatga gaacagcagt ttattactca ctaaagacag aatgaatgta gaaaaggctg
1021 aattctgtaa taaaagcaaa cagcctggct tagcaaggag ccaacataac agatgggctg
1081 gaagtaagga aacatgtaat gataggcgga ctcccagcac agaaaaaaag gtagatctga
1141 atgctgatcc cctgtgtgag agaaaagaat ggaataagca gaaactgcca tgctcagaga
1201 atcctagaga tactgaagat gttccttgga taacactaaa tagcagcatt cagaaagtta
1261 atgagtggtt ttccagaagt gatgaactgt taggttctga tgactcacat gatggggagt
1321 ctgaatcaaa tgccaaagta gctgatgtat tggacgttct aaatgaggta gatgaatatt
1381 ctggttcttc agagaaaata gacttactgg ccagtgatcc tcatgaggct ttaatatgta
1441 aaagtgaaag agttcactcc aaatcagtag agagtaatat tgaagacaaa atatttggga
1501 aaacctatcg gaagaaggca agcctcccca acttaagcca tgtaactgaa aatctaatta
1561 taggagcatt tgttactgag ccacagataa tacaagagcg tcccctcaca aataaattaa
1621 agcgtaaaag gagacctaca tcaggccttc atcctgagga ttttatcaag aaagcagatt
1681 tggcagttca aaagactcct gaaatgataa atcagggaac taaccaaacg gagcagaatg
1741 gtcaagtgat gaatattact aatagtggtc atgagaataa aacaaaaggt gattctattc
1801 agaatgagaa aaatcctaac ccaatagaat cactcgaaaa agaatctgct ttcaaaacga
1861 aagctgaacc tataagcagc agtataagca atatggaact cgaattaaat atccacaatt
1921 caaaagcacc taaaaagaat aggctgagga ggaagtcttc taccaggcat attcatgcgc
1981 ttgaactagt agtcagtaga aatctaagcc cacctaattg tactgaattg caaattgata
2041 gttgttctag cagtgaagag ataaagaaaa aaaagtacaa ccaaatgcca gtcaggcaca
2101 gcagaaacct acaactcatg gaaggtaaag aacctgcaac tggagccaag aagagtaaca
2161 agccaaatga acagacaagt aaaagacatg acagcgatac tttcccagag ctgaagttaa
2221 caaatgcacc tggttctttt actaagtgtt caaataccag tgaacttaaa gaatttgtca
2281 atcctagcct tccaagagaa gaaaaagaag agaaactaga aacagttaaa gtgtctaata
2341 atgctgaaga ccccaaagat ctcatgttaa gtggagaaag ggttttgcaa actgaaagat
2401 ctgtagagag tagcagtatt tcattggtac ctggtactga ttatggcact caggaaagta
2461 tctcgttact ggaagttagc actctaggga aggcaaaaac agaaccaaat aaatgtgtga
2521 gtcagtgtgc agcatttgaa aaccccaagg gactaattca tggttgttcc aaagataata
2581 gaaatgacac agaaggcttt aagtatccat tgggacatga agttaaccac agtcgggaaa
2641 caagcataga aatggaagaa agtgaacttg atgctcagta tttgcagaat acattcaagg
2701 tttcaaagcg ccagtcattt gctccgtttt caaatccagg aaatgcagaa gaggaatgtg
2761 caacattctc tgcccactct gggtccttaa agaaacaaag tccaaaagtc acttttgaat
2821 gtgaacaaaa ggaagaaaat caaggaaaga atgagtctaa tatcaagcct gtacagacag
2881 ttaatatcac tgcaggcttt cctgtggttg gtcagaaaga taagccagtt gataatgcca
2941 aatgtagtat caaaggaggc tctaggtttt gtctatcatc tcagttcaga ggcaacgaaa
3001 ctggactcat tactccaaat aaacatggac ttttacaaaa cccatatcgt ataccaccac
3061 tttttcccat caagtcattt gttaaaacta aatgtaagaa aaatctgcta gaggaaaact
3121 ttgaggaaca ttcaatgtca cctgaaagag aaatgggaaa tgagaacatt ccaagtacag
3181 tgagcacaat tagccgtaat aacattagag aaaatgtttt taaagaagcc agctcaagca
3241 atattaatga agtaggttcc agtactaatg aagtgggctc cagtattaat gaaataggtt
3301 ccagtgatga aaacattcaa gcagaactag gtagaaacag agggccaaaa ttgaatgcta
3361 tgcttagatt aggggttttg caacctgagg tctataaaca aagtcttcct ggaagtaatt
3421 gtaagcatcc tgaaataaaa aagcaagaat atgaagaagt agttcagact gttaatacag
3481 atttctctcc atatctgatt tcagataact tagaacagcc tatgggaagt agtcatgcat
3541 ctcaggtttg ttctgagaca cctgatgacc tgttagatga tggtgaaata aaggaagata
3601 ctagttttgc tgaaaatgac attaaggaaa gttctgctgt ttttagcaaa agcgtccaga
3661 aaggagagct tagcaggagt cctagccctt tcacccatac acatttggct cagggttacc
3721 gaagaggggc caagaaatta gagtcctcag aagagaactt atctagtgag gatgaagagc
3781 ttccctgctt ccaacacttg ttatttggta aagtaaacaa tataccttct cagtctacta
3841 ggcatagcac cgttgctacc gagtgtctgt ctaagaacac agaggagaat ttattatcat
3901 tgaagaatag cttaaatgac tgcagtaacc aggtaatatt ggcaaaggca tctcaggaac
3961 atcaccttag tgaggaaaca aaatgttctg ctagcttgtt ttcttcacag tgcagtgaat
4021 tggaagactt gactgcaaat acaaacaccc aggatccttt cttgattggt tcttccaaac
4081 aaatgaggca tcagtctgaa agccagggag ttggtctgag tgacaaggaa ttggtttcag
4141 atgatgaaga aagaggaacg ggcttggaag aaaataatca agaagagcaa agcatggatt
```

Sequencing results for all the clones (B, F, G and K) and for MCF-10A WT:

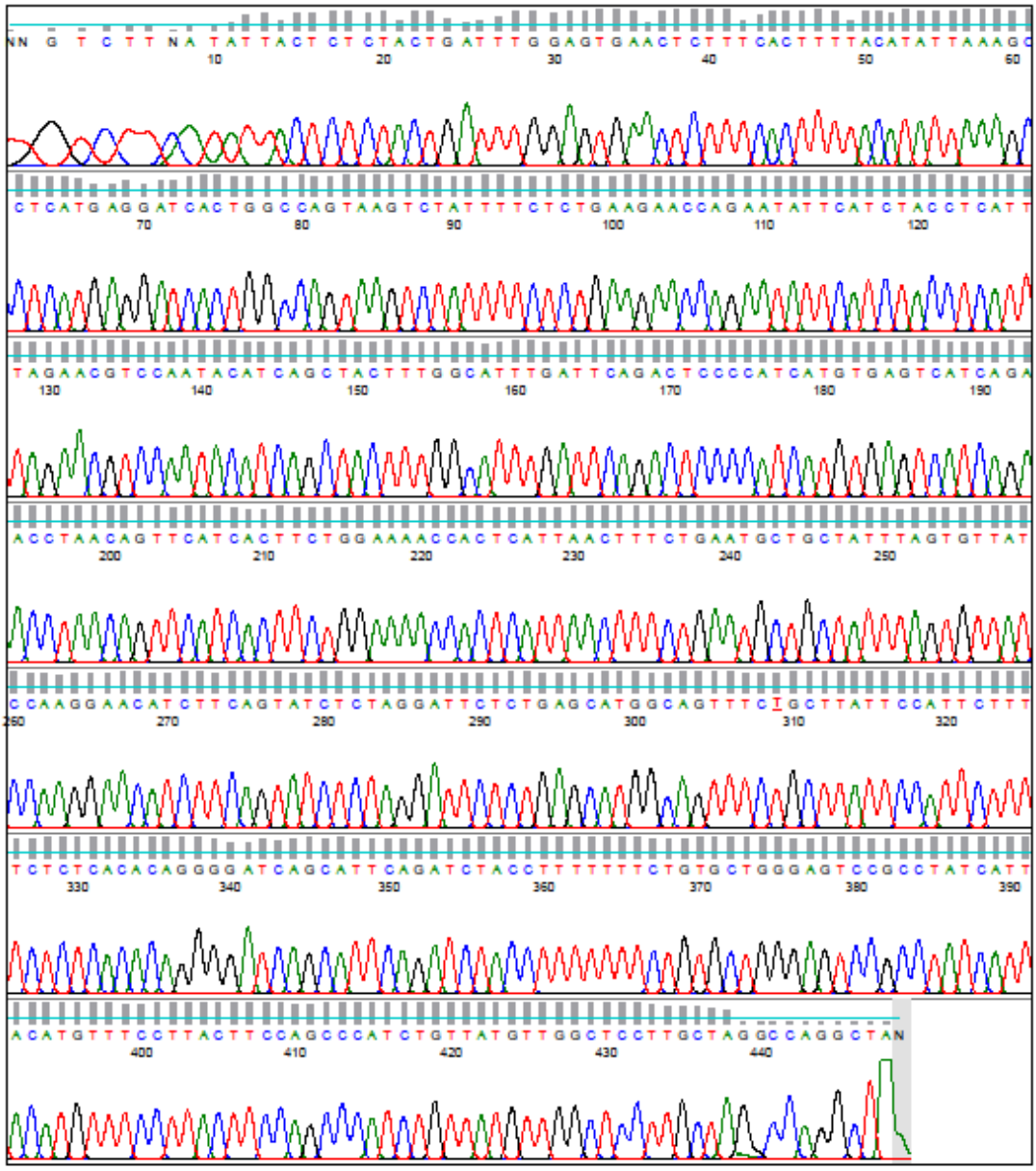


File: BRCA1_B+PR.ab1

Geospiza
www.geospiza.com

Sample Name: 1731635_D9_BRCA1_B+PR
Mobility: KB_3730_POP7_BDTv3.mob
Spacing: 15.4191
Comment: Premium reaction

Signal Strengths: A = 3832, C = 3742, G = 2302, T = 4574
Lane/Cap#: 73
Matrix: n/a
Direction: Native



Printed: nov 25, 2022 4:55PM

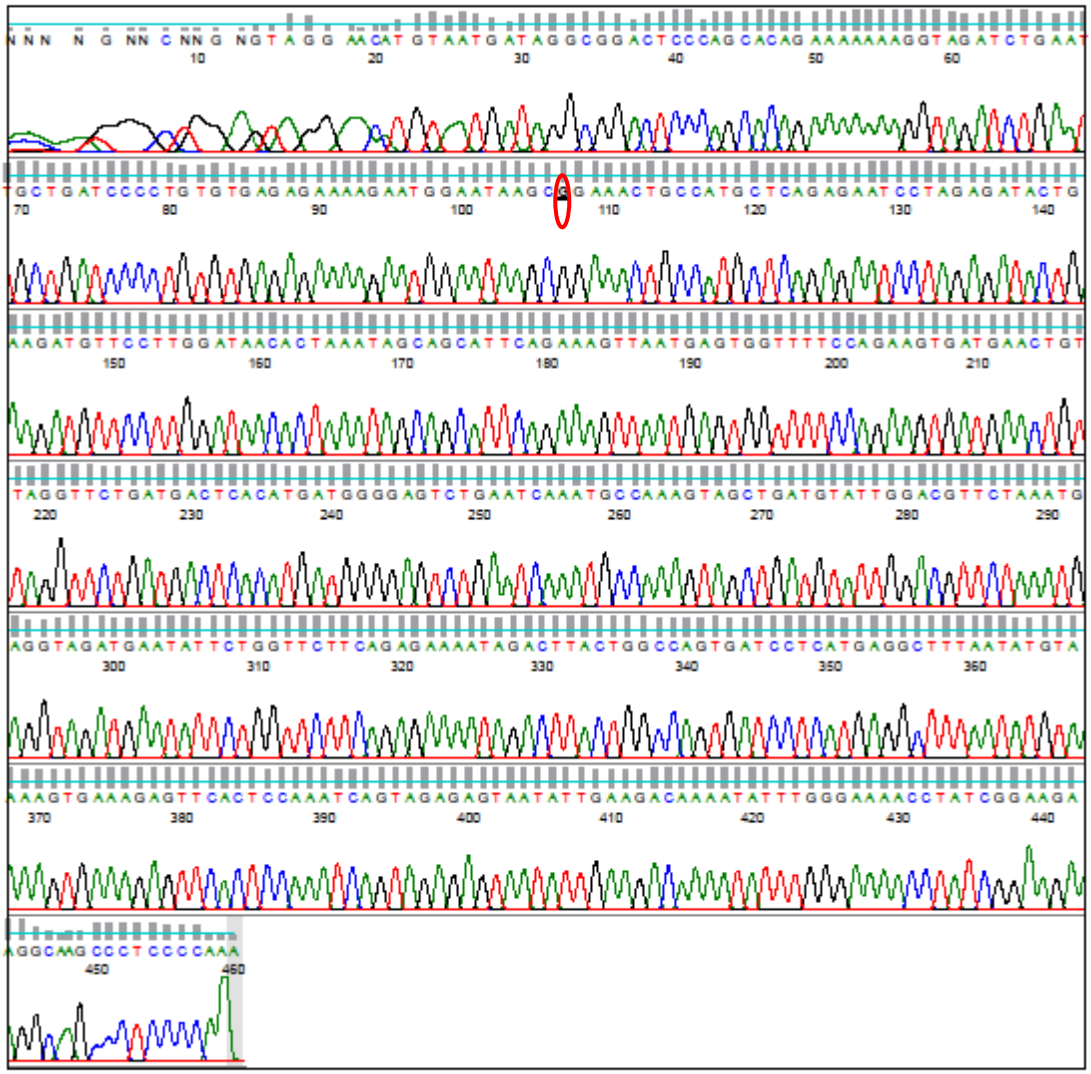
FinchTV v.1.4.0

Page 1 of 1

File: BRCA1_F+PF.ab1

Geospiza
www.geospiza.com

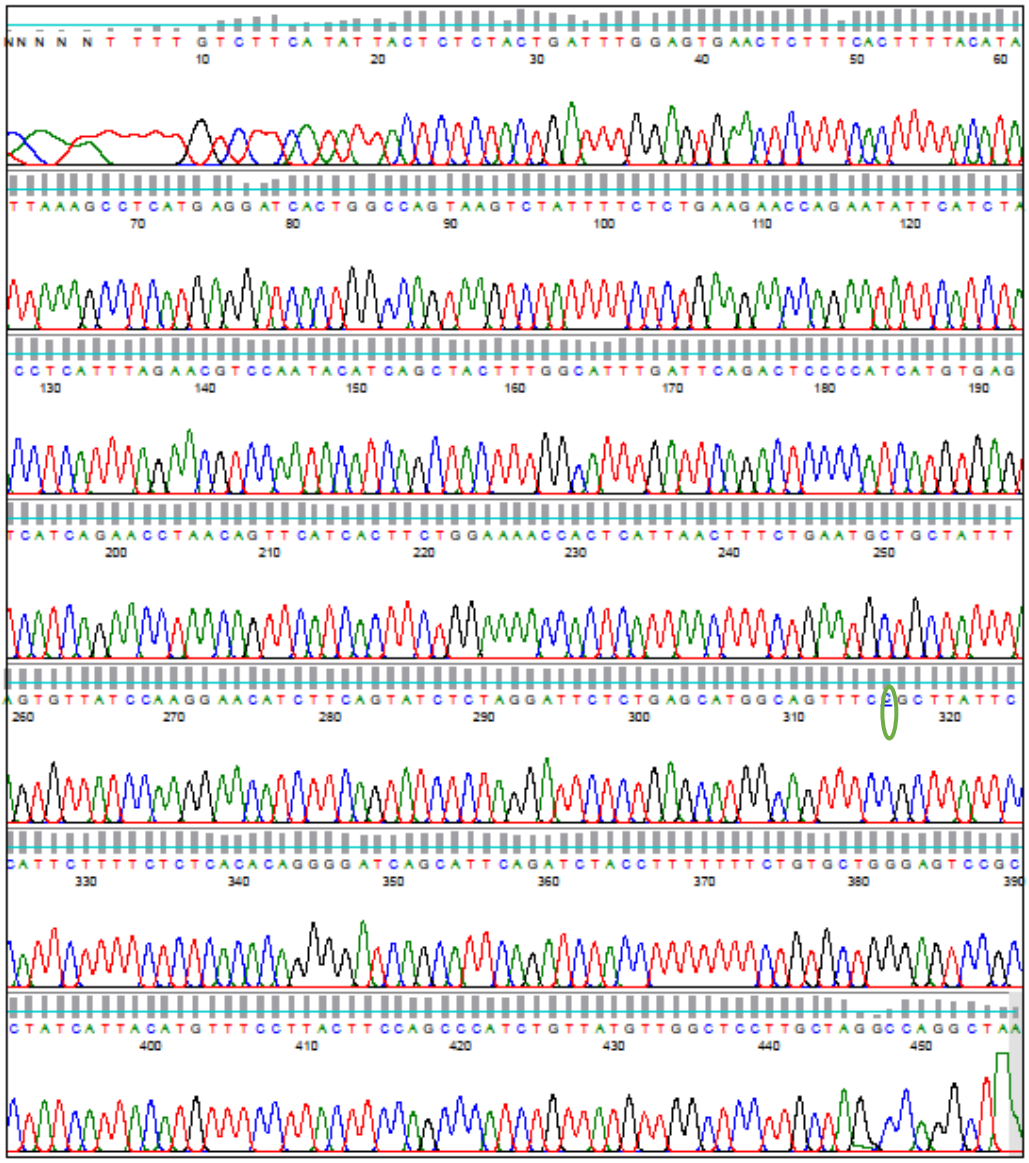
Sample Name:	1731631_H8_BRCA1_F+PF	Signal Strengths:	A = 4630, C = 2660, G = 3311, T = 4088
Mobility:	KB_3730_POP7_BDTv3.mob	Lane/Cap#:	50
Spacing:	15.2991	Matrix:	n/a
Comment:	Premium reaction	Direction:	Native



File: BRCA1_F+PR.ab1

Geospiza
www.geospiza.com

Sample Name:	1731636_E9_BRCA1_F+PR	Signal Strengths:	A = 4086, C = 4008, G = 2423, T = 4995
Mobility:	KB_3730_POP7_BDTv3.mob	Lane/Cap#:	71
Spacing:	15.4027	Matrix:	n/a
Comment:	Premium reaction	Direction:	Native

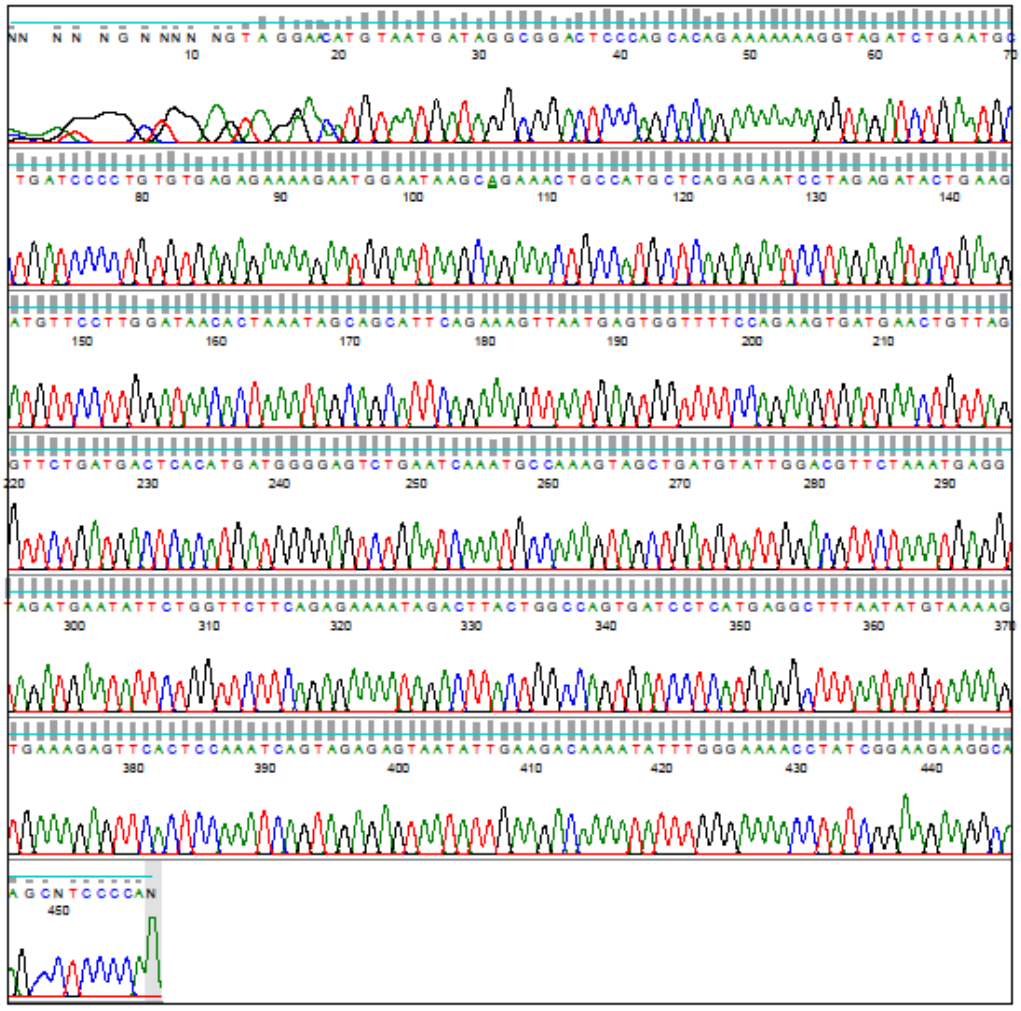


Printed: nov 25, 2022 3:51PM | FinchTV v.1.4.0 | Page 1 of 1

File: BRCA1_G+PF.ab1

Geospiza
www.geospiza.com

Sample Name:	1731632_A9_BRCA1_G+PF	Signal Strengths:	A = 4159, C = 2060, G = 3165, T = 3274
Mobility:	KB_3730_POP7_BDTv3.mob	Lane/Cap#:	79
Spacing:	15.3828	Matrix:	n/a
Comment:	Premium reaction	Direction:	Native

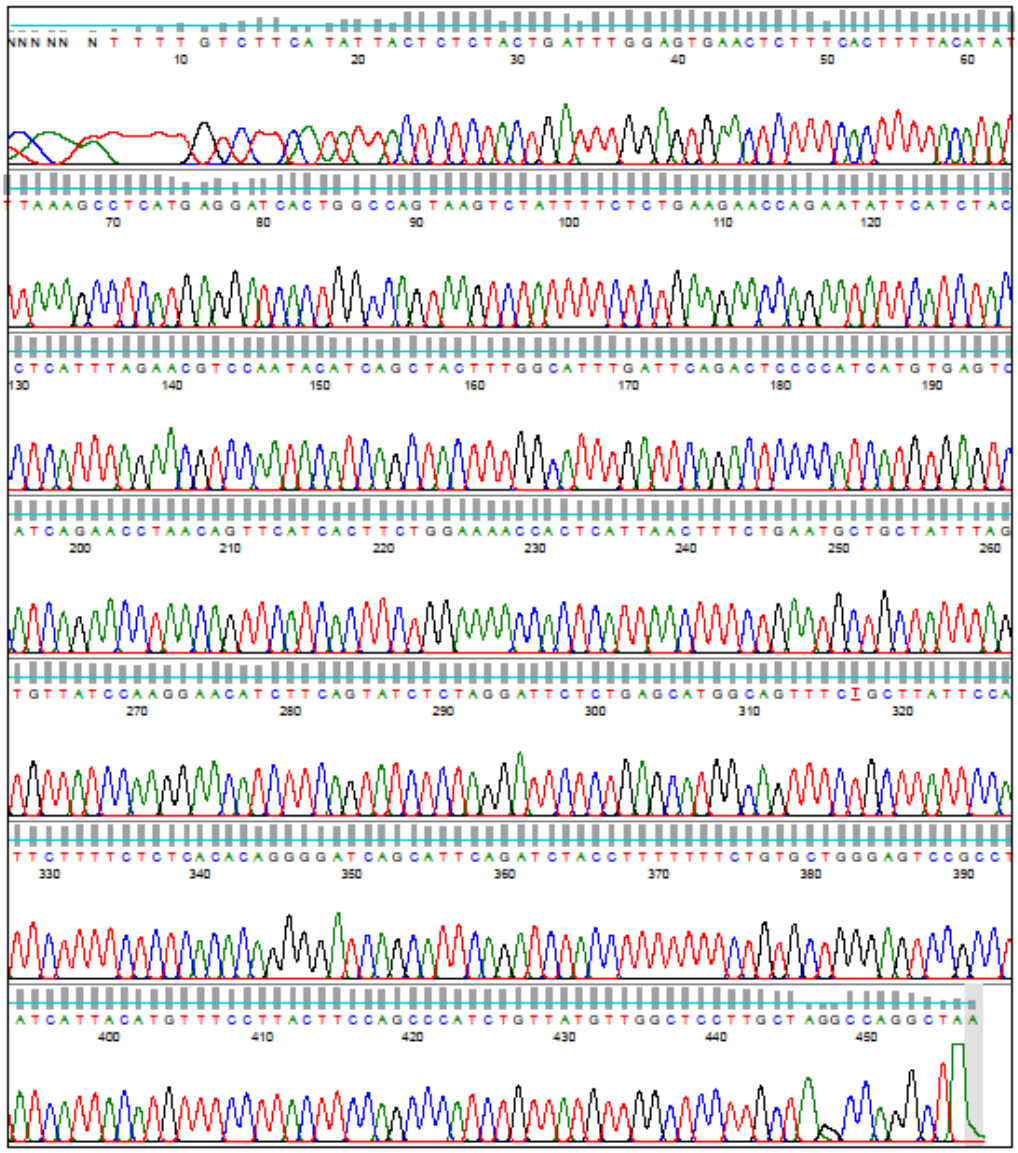


Printed: nov 25, 2022 4:58PM FinchTV v.1.4.0 Page 1 of 1

File: BRCA1_G+PR.ab1

Geospiza
www.geospiza.com

Sample Name:	1731637_F9_BRCA1_G+PR	Signal Strengths:	A = 2145, C = 2033, G = 1405, T = 2416
Mobility:	KB_3730_POP7_BDTv3.mob	Lane/Cap#:	69
Spacing:	15.4482	Matrix:	n/a
Comment:	Premium reaction	Direction:	Native

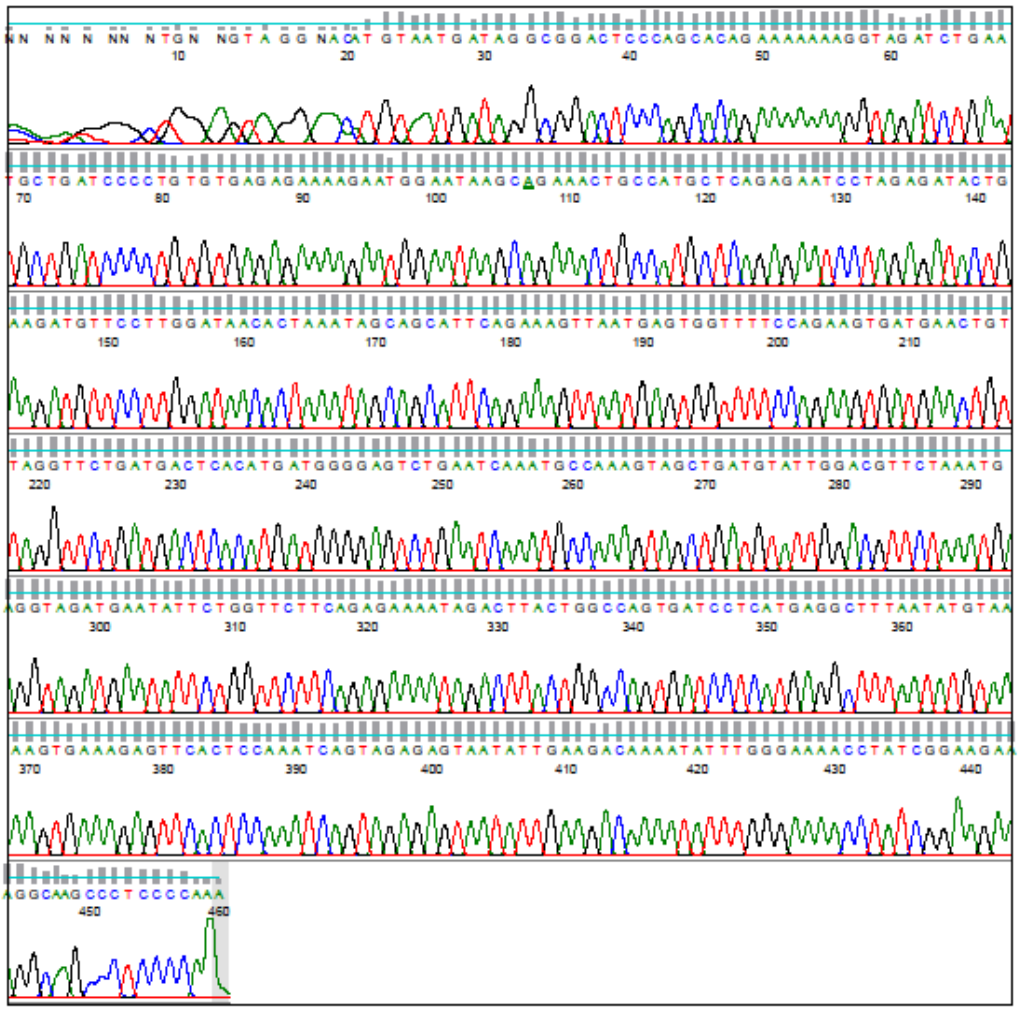


Printed: nov 25, 2022 5:00PM FinchTV v.1.4.0 Page 1 of 1

File: BRCA1_K+PF.ab1

Geospiza
www.geospiza.com

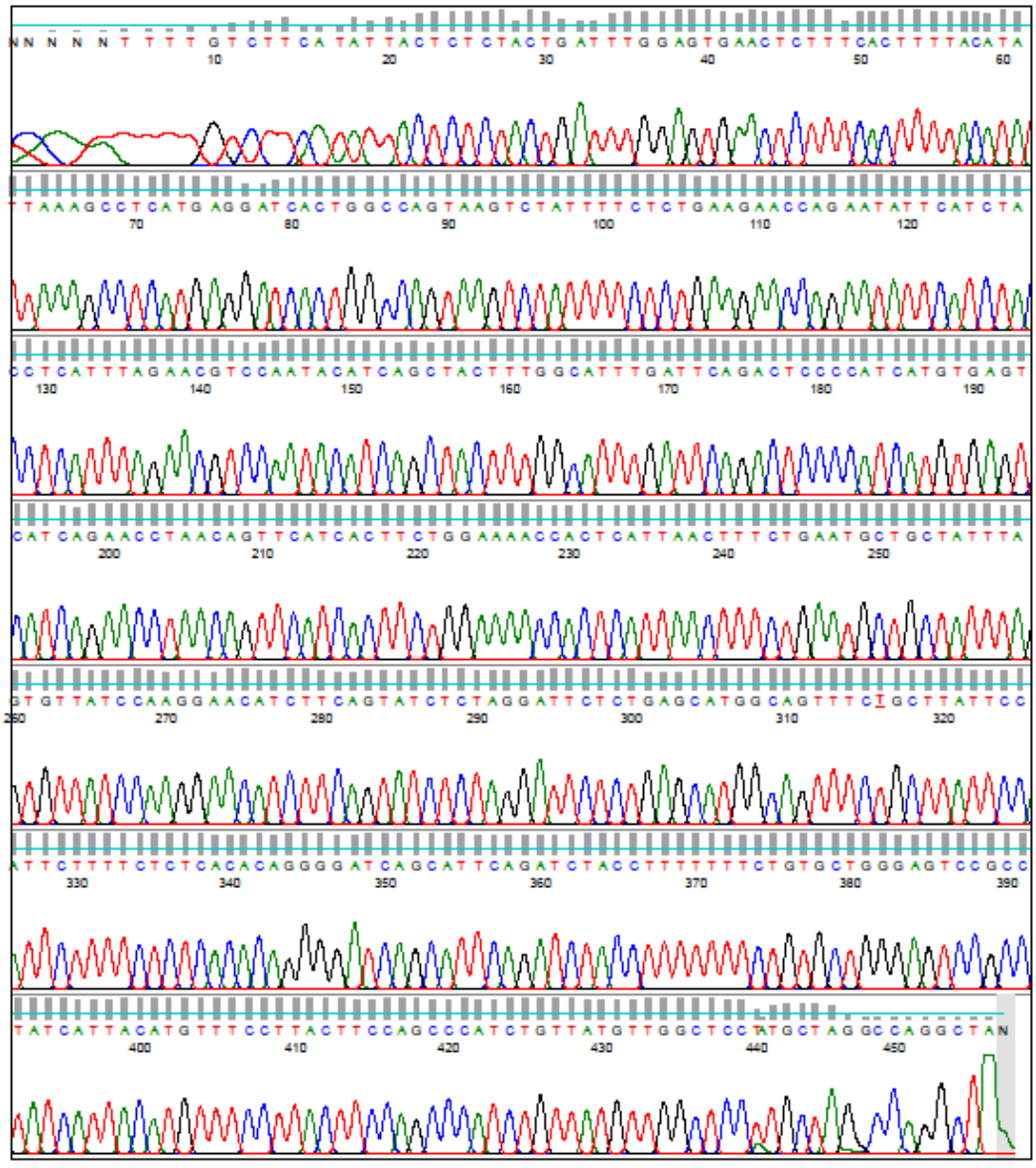
Sample Name:	1731633_B9_BRCA1_K+PF	Signal Strengths:	A = 4731, C = 2683, G = 3474, T = 3960
Mobility:	KB_3730_POP7_BDTv3.mob	Lane/Cap#:	77
Spacing:	15.4049	Matrix:	n/a
Comment:	Premium reaction	Direction:	Native



File: BRCA1_K+PR.ab1

Geospiza
www.geospiza.com

Sample Name:	1731638_G9_BRCA1_K+PR	Signal Strengths:	A = 3525, C = 3465, G = 2229, T = 4101
Mobility:	KB_3730_POP7_BDTv3.mob	Lane/Cap#:	67
Spacing:	15.4861	Matrix:	n/a
Comment:	Premium reaction	Direction:	Native

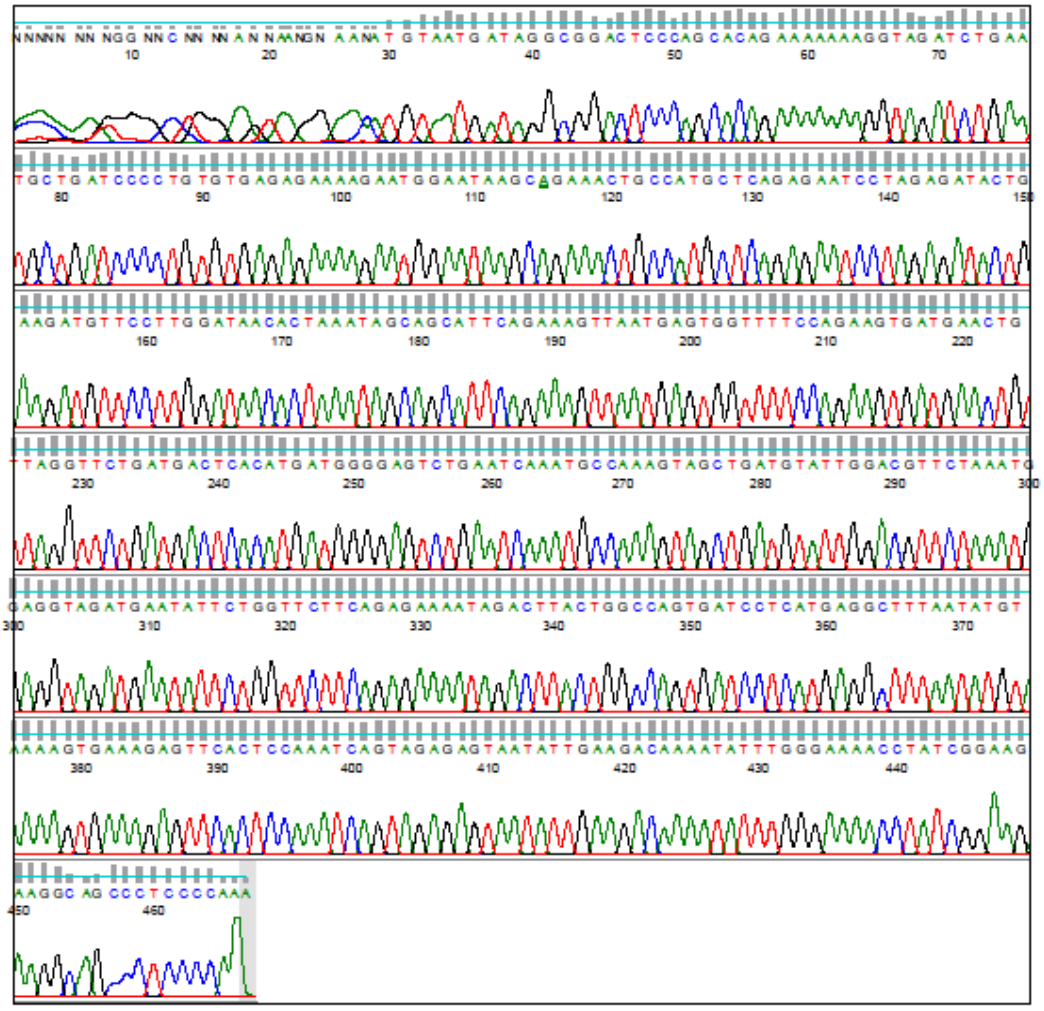


Printed: nov 25, 2022 5:03PM FinchTV v.1.4.0 Page 1 of 1

File: MCF10A-WT+PF.ab1

Geospiza
www.geospiza.com

Sample Name:	1731629_F8_MCF10A-WT+PF	Signal Strengths:	A = 5221, C = 2736, G = 3514, T = 4679
Mobility:	KB_3730_POP7_BDTv3.mob	Lane/Cap#:	54
Spacing:	15.1739	Matrix:	n/a
Comment:	Premium reaction	Direction:	Native

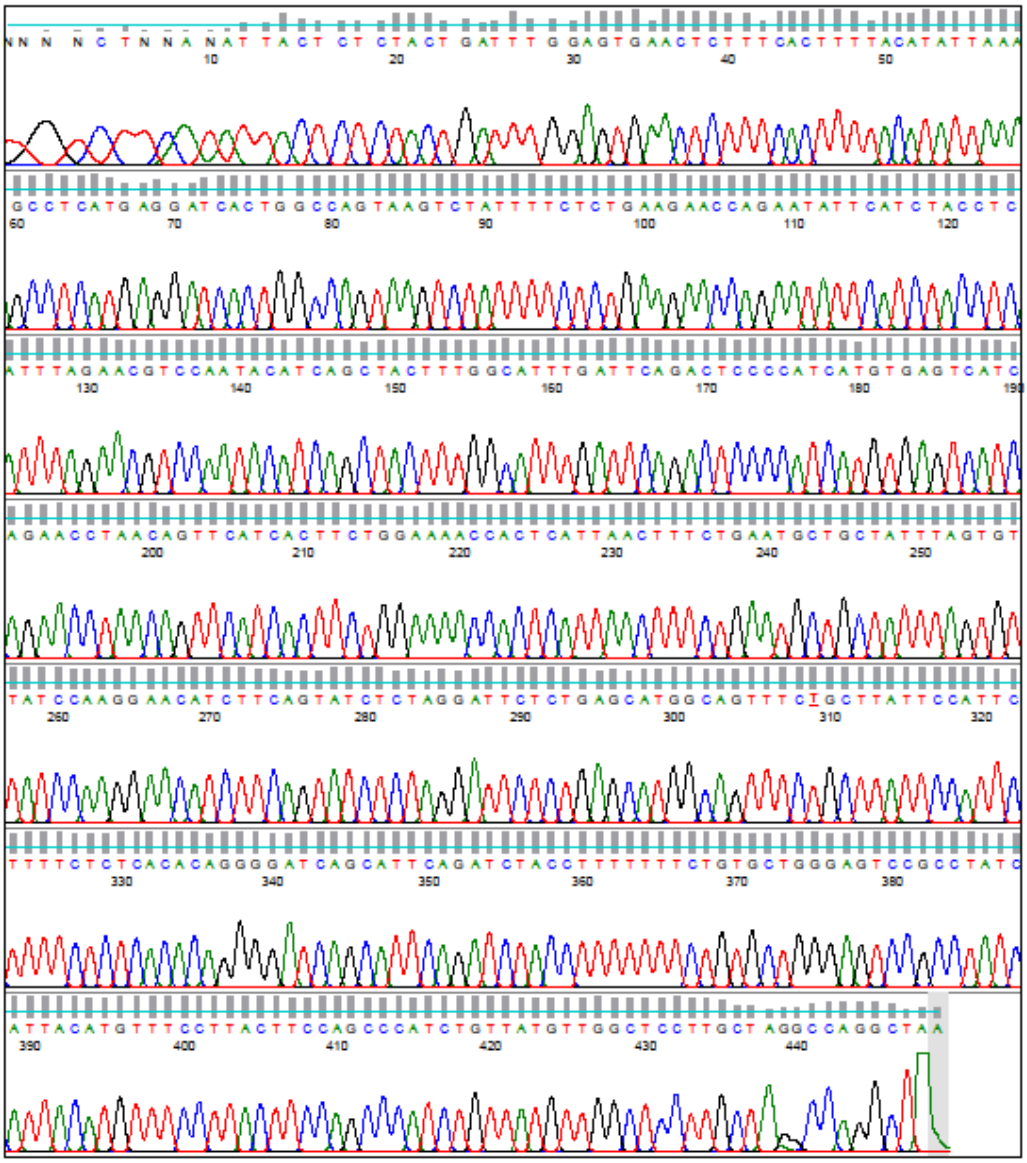


File: MCF10A-WT+PR.ab1

Geospiza
www.geospiza.com

Sample Name: 1731634_C9_MCF10A-WT+PR
Mobility: KB_3730_POP7_BDTv3.mob
Spacing: 15.5023
Comment: Premium reaction

Signal Strengths: A = 2576, C = 2508, G = 1634, T = 3124
Lane/Cap#: 75
Matrix: n/a
Direction: Native



Printed: nov 25, 2022 4:13PM

FinchTV v.1.4.0

Page 1 of 1