The Molecular Aetiology of Inherited Breast Cancer in the South African Black Population

Wenlong Chen

0700425A

2015

Dissertation submitted to the Faculty of Health Science, University of the Witwatersrand, Johannesburg, in fulfilment of the requirements for the degree of Master of Science in Medicine.

DECLARATION

I, Wenlong Chen, declare that this dissertation is my own work. It is being submitted for the degree of Masters of Science in Medicine in the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at this or any other University.

_____

Wenlong Chen

_____ day of _____, 2015

DEDICATION

I dedicate this work to my parents, who have supported me whole-heartedly through this journey. Thank you.

I also dedicate this work to the participants of this study: without you this study would not have been possible. I hope the outcomes of this study will one day benefit you, as well as future generations burdened by this disease.

PRESENTATIONS ARISING FROM THIS STUDY

Poster presentations:

Chen, W., Kerr, R., Wainstein, T., & Krause, A., (2012). The Molecular Aetiology of Inherited Breast Cancer in the South African Black Population. Faculty of Health Science Research Day, University of the Witwatersrand, Johannesburg, South Africa (19th September 2012)

Chen, W., Kerr, R., Wainstein, T., & Krause, A., (2012). The Molecular Aetiology of Inherited Breast Cancer in the South African Black Population. 15th Southern African Society of Human Genetics Congress, Johannesburg, South Africa (6-9 October 2013)

ABSTRACT

Hereditary breast cancer (HBC), caused by mutations in tumour suppressor genes, has been well studied in Caucasian and Ashkenazi Jewish populations. Little is known about the genetic aetiology or the clinical epidemiology of HBC in African and South African black populations. Founder mutations within specific genes have been described for numerous populations worldwide. In South African black women, breast cancer often presents with early age of onset, rapid progression of disease and adverse histological phenotypes. It is possible that mutations within already described HBC-associated genes account for this phenotype in the South African black population. Genetic mutations within the *BRCA1* and *BRCA2* genes contribute to the development of breast cancer in high-risk cancer patients internationally. The proportion of breast cancers caused by mutations in the *BRCA1* and *BRCA2* genes in the local black population is currently not well understood.

This study screened 33 South African black women presenting with early onset, high-risk breast cancer, for genetic mutations in the *BRCA1* and *BRCA2* genes. These individuals were selected on predefined high-risk criteria, including early age of disease onset, family history and cancer histology. Mutation screening was done using bi-directional Sanger sequencing of the exons of both genes, and MLPA analysis.

A total of 59 distinct single nucleotide variants were found in the *BRCA1* and *BRCA2* genes. Three of these variants are pathogenic or likely pathogenic (*BRCA1:* c.431dupA; *BRCA2:* c.582G>A & c.7712A>G). Five novel variants with unknown clinical significance were also found (*BRCA1:* c.3751T>G, c.306A>C, c.5332+78C>T, c.212+66A>G; *BRCA2:* c.681+10T>G). There is little evidence to suggest that these novel variants are likely to have a functional impact and to be pathogenic. The remaining 51 variants were previously reported. No large gene/exon deletions or duplication mutations were found on MLPA analysis.

The allele frequency data of the 51 previously reported variants were compared to the allele frequencies of two separate control groups: an ethnically matched control group (established and investigated in-house by a previous Masters student in the Division) and the 1000 genomes project data. This comparison was done to screen for tag SNPs for future genetic association studies in black South Africans and to evaluate potential background genetic differences within the *BRCA* genes of different African populations. One allele within the *BRCA1* gene occurred at a frequency which was statistically significant as compared to the normal South African black population and one allele within the *BRCA2* gene was statistically significant. The two control groups were also compared against each other. The allele frequencies were significantly different between the two control

groups, suggesting that other African population groups (1000 genomes project data) do not serve well as control groups for sub-Saharan African populations.

Genetic mutations within the *BRCA1* and *BRCA2* genes account for approximately 10% (3/33) of the breast cancer cases within this high-risk cohort. Thus, the proportion of *BRCA* mutations contributing to the development of inherited breast cancer in sub-Saharan blacks appears to be in keeping with data reported for other ethnic groups. No evidence of any founder mutation/s was evident, although the sample size tested here was too small to exclude their existence.

ACKNOWLEDGMENTS

I would like to thank the supervisors for the tremendous contribution they have made towards this study. Thank you for mentoring and guiding me through this amazing journey, without you this study would not have been possible:

Dr Robyn Kerr and Professor Amanda Krause

I would like to thank Ms Tasha Wainstein for persuading me and leading me down this amazing journey. Tasha's work laid the foundation for this study, without her my study would not have been possible.

I would like to thank the following individuals for aiding me along my journey, I am truly thankful:

Ms Fahmida Essop

Mrs Elana Vorster

Mr Thijessen Naidoo

Ms Rajeshree Mahabeer

Ms Shelley Macaulay

Ms Chantel van Wyk

I would like to thank the staff of the clinical and counselling section, Division of Human Genetics, for recruiting the patients and collecting samples for this study.

I would like to thank the staff of the Molecular Diagnostic Laboratory for extracting the samples.

I would like to thank the Applied Polymorphism Laboratory and the Human Genomic Diversity and Disease Research Unit for laboratory access at times of need.

I would like to thank the fellow staff and students of the Molecular Research Unit. Thank you for all the joys and laughter's we shared over the years.

Lastly I would like to thank the University of the Witwatersrand and the National Health laboratory Service for housing this study and for the financial support towards this study.

The financial assistance of the National Research Foundation (NRF) towards this research is hereby acknowledged. Opinions expressed and conclusions arrived at, are those of the author and are not necessarily to be attributed to the NRF.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABBREVIATIONS

1KG	1000 Genomes Project
ATM	Ataxia Telangiectasia Mutated
BARD1	*BRCA1* Associated RING Domain 1
BIC	Breast Information Core
BRCA1	Breast Cancer Gene 1
BRCA2	Breast Cancer Gene 2
BRIP1	*BRCA1* interacting protein C-terminal helicase 1
CDCV	Common Disease Common Variant
CDRV	Common Disease Rare Variant
CDS	Coding DNA Sequence
CHEK2	Checkpoint Kinase 2
COGS	Collaborative Oncological Gene-Environment Study
DQ	Dosage Quotient
ER	Oestrogen Receptor
GWAS	Genome Wide Association Study
HBC	Hereditary Breast Cancer
HBOC	Hereditary Breast and Ovarian Cancer
HER2	Human Epidermal Growth Factor Receptor 2
HGMD	Human Gene Mutation Database
HWE	Hardy-Weinberg Equilibrium
INDELS	Insertions and Deletions
IUB	International Union of Biochemistry
LOVD	Leiden Open Variation Database
MAF	Minor Allele Frequency
MLPA	Multiplex Ligation Probe Amplification
MRI	Magnetic Resonance Imaging
NBA1	New Component of the BRCA1 A Complex
NBS1	Nijmegen Breakage Syndrome gene 1
NGRL	National Genetics Reference Laboratories
NHLS	National Health Laboratory Services
NMD	Nonsense Mediated Decay
OMIM	Online Mendelian Inheritance in Man
PALB2	Partner and Localizer of BRCA2
PR	Progesterone Receptor
PTEN	Phosphatase and Tensin Homologue
RAD50	DNA Repair Protein RAD50
RAD51	DNA Repair Protein RAD51
SNP	Single Nucleotide Polymorphism
SNV	Single Nucleotide Variant
STK11	Serine/threonine Kinase 11
TNBC	Triple Negative Breast Cancer
US	Universal Sequence
UTR	Untranslated Region
WITS	University of the Witwatersrand

1. INTRODUCTION

Cancer is a disease of uncontrolled, rapid, abnormal proliferation and growth of cells within the body, as a result of the accumulation of genetic damage/mutations within a single, original progenitor cell (NCI, 2012). Clonal reproduction of this cell leads to a tumour. Tumour cells can rapidly invade surrounding tissues as well as migrate to distant locations within the human body, disrupting normal cellular functions, ultimately leading to death of that individual. Currently, cancer is a leading cause of death, responsible for 13% (7.6 million) of all deaths worldwide in 2008 (WHO, 2012).

Breast cancer is the most commonly occurring cancer amongst females worldwide and is the leading cause of death due to cancer in females (Ferlay et al., 2010). Breast cancer alone accounted for 22.9% of the overall cancer incidence and 13.7% of the overall mortality rate in females worldwide (Ferlay et al., 2010). In South Africa, breast cancer is the second most prevalent cancer amongst black females after cervical cancer (NCR, 2012).

Cancer is categorized into two groups: sporadic cancer or hereditary cancer. The majority (90 – 95%) of cancers are sporadic arising in an adult cell due to damaging environmental factors and at-risk lifestyles such as alcohol intake, cigarette smoking, radiation, infections and other pollutants present in the surrounding environment (Anand et al., 2008). Only small percentages (5 – 10%) of cancers arise due to a pre-existing genetic mutation inherited as a germline mutation, and present from the time of conception. These are known as inherited/hereditary cancers. Individuals with an inherited genetic mutation have a predisposition to, and a significantly higher life time risk of, developing cancer.

The differentiation of hereditary cancers from sporadic cancers is important for numerous reason, one such reason is the risk implication for hereditary cancer mutation carriers. Sporadic cancers are the results of the accumulation of genetic damages the cell/body have sustained over a lifetime in individuals without germline mutations (NCI Dictionary of Cancer Terms, 2015). The lifetime risk for individuals to develop sporadic cancers is equal to that of the average population cancer risk for that cancer (Howlader et al., 2014). Genetic factor have been known to contribute to the development of sporadic cancers, these factors are known as tumour susceptibility genes (TSGs). Individuals who carries multiple TSGs have an increased risk to the development of sporadic cancers (reviewed in Fijneman, 2005). Hereditary cancers differs to sporadic cancers in that there are germline mutations in tumour suppressor genes, or DNA repair genes, that are passed down from generation to generation. This greatly increases the risk for these mutation carriers to develop cancer.

1.1. Identifying patients to test for inherited breast cancer in the South African black population

Inherited breast cancer usually presents with early age of disease onset and an aggressive disease phenotype. This is especially true in the local black populations of South Africa (Wainstein, 2011). The incidence of breast cancer is on the increase. The age of onset of cancer in this population is as young as 25 years of age and the clinical picture is of a rapidly progressing cancer phenotype (Wainstein, 2011). The high incidence of inherited breast cancer, combined with the observed early age of onset, is suggestive of an underlying genetic predisposition and the presence of inherited mutations.

Family history of certain cancers and/or breast cancer is a strong indicator of inherited breast cancer. The study conducted by Wainstein (2011) described the testing of risk models that have been designed to screen for and identify hereditary breast cancer patients amongst general breast cancer patients. One vital component of the risk model is the availability of patient family history detailing previous/present cancers (Wainstein, 2011). Positive family history for breast cancers or other related cancers is a strong indicator for inherited breast cancer. However, in the South African black population, family history is not always available. The lack of specialized medical services to state patients, unclear cause of death, and the lack of education and understanding of cancer in general, often results in a lack of vital information being collected regarding the patient or family members. Obtaining a clear family history is not always possible. In the absence of a clear family history, other cancer characteristics may need to be used to indicate underlying genetic predispositions.

Another “diagnostic clue” pointing to the presence of an inherited mutation is the presence of triple negative breast cancer. Breast cells normally express, amongst other receptors, the receptors for the hormones oestrogen (ER) and progesterone (PR). Another receptor, human epidermal growth factor receptor 2 (HER2), also acts to receive signals that stimulate growth. Only about 10 - 15% of breast cancers test negative for these three receptors, and are termed triple negative breast cancers (TNBC). TNBC is associated with earlier age at onset, more aggressive tumour progression and resistance to standard therapies. It is important to note however, that the incidence of TNBC status varies by race and age and is approximately three times higher in African-Americans than in white American women (Carey et al., 2006). If the breast cancer patient is of African American origin, the likelihood that the cancer will be triple negative is 20-37% (Carey et al., 2006). In a sample of unselected TNBC patients, the likelihood of finding a germline *BRCA* mutation was nearly 20% (Gonzalez-Angulo et al., 2011). Sixteen percent of these patients were found to carry a *BRCA1* mutation and 4% of patients carried a *BRCA2* mutation. It is important to note that patients’ ethnicity was taken into consideration by Gonzalez-Angulo, et al. (2011). Therefore, incorporating

TNBC status into the inclusion criteria for molecular genetic testing may increase the probability of finding *BRCA* mutations in early onset - South African black patients who lack a clear/positive family history.

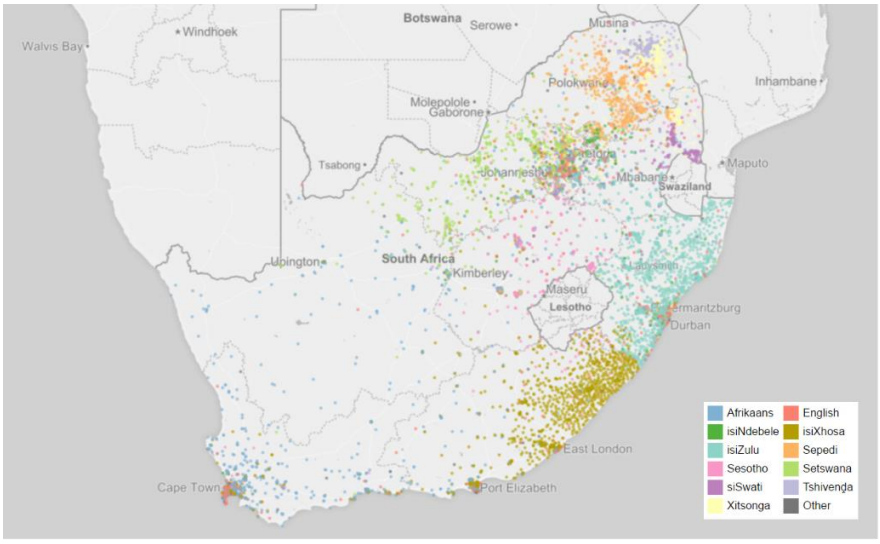


Figure 1.1: Map of the demographic of South Africa by first language. First language data obtained from the Census 2011. One dot represents 5000 people. Image courtesy of Adrian Frith (dotmap.adrianfrith.com).

It is important to note that the local black population is not a homogeneous population. The Census 2011 found that amongst the 51.77 million South Africans, 79.2% are African, 8.9% are white and coloured, 2.5% are Asian/Indian and 0.5% as others (Census 2011, 2012). 23.7% of the 51 million people reside in Gauteng province, from where this study recruited patients (Figure 1.1.). The African population within South Africa is largely comprised of four groups: the Nguni (Zulus, Xhosa, Ndebele and Swazi), Sotho-Tswana (Northern, Southern and Western Sotho), the Tsonga and the Venda (Census 2011, 2012). The population group of African patients are largely determined by the language spoken by the patients. There are nine officially recognized African languages in South Africa. The different languages and their corresponding percentage of home language speakers are as follows: IsiZulu (22.7%), IsiXhosa (16.0%), Sepedi (9.1%), Sesotho (7.6%), Setswana (8.0%), SiSwati (2.5%), Xitsonga (4.5%), Tshivenda (2.4%), and IsiNdebele (2.1%) (Census 2011, 2012).

1.2. Genetics of inherited breast cancer

Inherited breast cancer is an autosomal dominant trait with incomplete penetrance. Sporadic breast cancer is also a genetic disease, but here the predisposing mutation/s are obviously not inherited. To date, numerous genes have been described which contribute to the development of breast cancer or are the direct cause of the disease (Walsh & King, 2007). The genes that are responsible for the

development of breast cancer can largely be grouped into two groups: high penetrance genes and moderate/low penetrance genes.

1.3. High penetrance genes

While numerous genes are known to be involved in the aetiology of inherited breast cancer, mutations in the *BRCA1* or *BRCA2* genes are the most common cause of inherited breast cancer in populations studied to date (reviewed in van der Groep, et al., 2011). Mutations in these two genes confer a high risk of developing both breast and/or ovarian cancer in comparison to mutations in low penetrance genes. The average risk of developing breast cancer before the age of 70 years for *BRCA1* mutation carriers may be as high as 90% and the life time ovarian cancer risk ranges from 24% to 39% (Antoniou et al., 2003, Chen & Parmigiani, 2007). *BRCA2* mutation carriers have an estimated average lifetime risk of 45% for breast cancer development and 11% for ovarian cancer development before the age of 70 years (Antoniou et al., 2003, Chen & Parmigiani, 2007).

1.3.1. The *BRCA1* gene

The human *BRCA1* gene (OMIM 113705) was first identified through linkage analysis in families with early onset breast cancer by Hall, et al. (1990) and it was cloned four years later (Miki et al., 1994). Since then extensive research has gone into elucidating the function of the *BRCA1* gene and the implication of mutations within this gene.

1.3.1.1. Function of the BRCA1 protein

The *BRCA1* gene is a large gene located on chromosome 17q21 and contains 24 exons which encode an 1863 amino acid protein (Miki et al., 1994). The BRCA1 protein is essential for DNA damage repair, cell cycle checkpoint control and maintaining genomic stability (Venkitaraman, 2004; Boulton, 2006). The BRCA1 protein facilitates double-strand breakage repair in DNA through a process of homologous recombination. The BRCA1 protein interacts and forms DNA repair complexes with other proteins such as RAD51, BARD1, and NBA1 to repair DNA breakage and maintain genomic integrity (reviewed in Boulton, 2006). This protein super structure is known as the BRCA1-associated genome surveillance complex (BASC) (Wang et al., 2000). *BRCA1* functions as a tumour suppressor gene, and loss of one allele leads to haplo-insufficiency (Kinzler & Vogelstein, 1997): one normal copy is not enough and the loss of one copy of the *BRCA1* results in insufficient and inefficient DNA repair and a resulting accumulation of DNA damage. This usually leads to the knockout of other tumour-suppressor genes, activation of oncogenes and ultimately, the development of cancer (Kinzler & Vogelstein, 1997).

This observation is in line with Knudson's hypothesis or the multiple-hit hypothesis (Knudson, 1971). In this hypothesis, Alfred Knudson states that cancer occurs as a result of the accumulation of

genetic mutations within a cell's DNA. These genetic mutations knockouts the functions of important tumour suppressor genes, given there are two copies of each gene, both copies need to be knocked out thus disabling these genes. This will allow the proliferation of the damaged cells and leads to the development of cancer (Knudson, 1971). Individuals with a germline mutation in tumour suppressor genes such as *BRCA1* and *BRCA2* are at a significantly higher risk to develop cancer as they only have one functional copy of the tumour suppressor gene.

1.3.1.2 Mutation profile of BRCA1

Sequence changes within *BRCA1* are extremely heterogeneous. Most (~70%) sequence changes are single nucleotide changes in exonic sequences. These changes usually result in the formation of an altered BRCA1 protein. These proteins are commonly non-functional or have a greatly reduced functionality when compared to their wildtype counterparts (Welcsh & King, 2001).

Large gene rearrangements and whole/partial gene deletion/duplications have also been reported (Woodward et al., 2005). These result in the loss of a functional copy of the *BRCA1* gene.

Over 14911 known sequence variations have been reported for *BRCA1* and these variants are curated in an on-going manner in the Breast Cancer Information Core (BIC) database that was established in 2000 (Szabo et al., 2000). A large proportion of known sequence changes are pathogenic mutations are frameshifts (39%) or missense (30%) in nature, the rest of the variants consist of intronic region variants (14.5%), synonymous variants (5.3%), in frame deletions, in frame insertions, 5' UTR, and 3' UTR mutations. It is important to note that a large portion of the intronic variants and synonymous variants are of unknown significance.

1.3.2. The *BRCA2* gene

The human *BRCA2* gene (OMIM 600185) was identified by positional cloning using 15 high-risk breast cancer families unrelated to the *BRCA1* locus (Wooster et al, 1995). This gene is located on chromosome 13q12.3. It is approximately 70 kb in length, containing 27 exons and encodes a protein that is 3418 amino acids in length (Tavtigian et al., 1996).

1.3.2.1. Function of the BRCA2 protein

The BRCA1 and BRCA2 proteins participate in protein complexes responsible for double-strand breakage repair in DNA through homologous recombination (reviewed in Boulton, 2006). BRCA1 and BRCA2 proteins share a degree of sequence homology and characteristics in as much as both genes contain a very large central exon coding for an active binding domain which participate in the same biochemical pathway (Jensen et al., 1996).

BRCA2 is allelic to the Fanconi Anemia D1 gene. A bi-allelic knockout of *BRCA2* results in the Fanconi Anemia Type D1 (OMIM 605724) phenotype. Fanconi Anemia is characterised by the presence of frequent chromosome breakage and genomic instability. This observation suggests *BRCA2* functions as a double-strand breakage repair gene (Howlett et al., 2002).

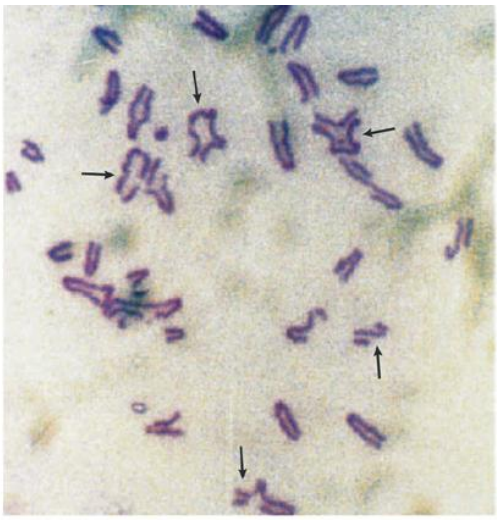


Figure 1.2: Chromosomal instability caused by homozygous BRCA2 deficiency. Giemsa-stained metaphase with chromosomal breakages indicated by arrows. Image courtesy of Venkitaraman, (2004).

1.3.2.2. *Mutation profile of BRCA2*

Sequence changes disrupting the *BRCA2* gene are largely attributed to single nucleotide changes that results in the production of altered or truncated BRCA2 proteins and a small percentage due to large gene rearrangements. To date, over 14823 known mutations have been reported for *BRCA2* and these mutations are curated in the BIC database. The most frequent sequence variant type in *BRCA2* differs from that of *BRCA1.* Missense mutations are most frequent (48%) in *BRCA2* and frameshift mutations are the second most common (24%). Other sequence variation types are less frequent in both genes, these mutation types include 3'UTR, 5'UTR, synonymous, in frame insertions, in frame deletions, splice site, intervening sequences, and nonsense mutations, as curated in the BIC database (Szabo et al., 2000a).

1.3.3. Functions of the BRCA proteins in DNA repair pathways

The BRCA1 and BRCA2 proteins function together and, together with other proteins, form complexes that are essential in DNA repair pathways. The full function of BRCA1 and BRCA2 is not yet fully understood.

The functions of BRCA1 and BRCA2 have been characterized in certain DNA repair pathways. BRCA1 is an essential component of BASC, which consists of BRCA1, MSH2, MSH6, MLH1, ATM, BLM proteins, and the RAD50-MRE11-NBS1 protein complex (Wang et al., 2000). This super complex is

responsible for transcription coupled repair. Transcription coupled repair is a form of nucleotide excision repair whereby the initial DNA damage recognition is different (Wang et al., 2000).

The BRCA1 and BRCA2 proteins are also essential for double stranded DNA breakage repair. Homology-directed repair: BRCA1 and BRCA2 are believed to act as the scaffold onto which other DNA repair proteins bind (Sharan & Bradley. 1998; Liu & West, 2002; Yoshida & Miki, 2004; Lu, et al., 2005). Homology-directed repair is a process where the homologous sister chromatid is used as template to repair the double stranded DNA breakage (Pardo et al., 2009). When a double stranded DNA break is detected, an exonuclease resects the breakage site, and the breakage site then unwinds. The sister chromatid unwinds and base-pairing occurs at the homologous sites. The broken DNA strand then extends using the sister chromatid as the template. The strands then disengage. Ligase then fills in the remaining gaps (reviewed in Shrivastav et al., 2008). When the sister chromatid is not available as the template for dsDNA repair, the dsDNA strands are joined together via non-homologous end joining. The breakage site is resected and joined together with limited base pairing, thereafter ligase fill the gaps. This results in the reformation of the double strand DNA helix (Frankenberg-Schwager et al., 2008).

1.3.4. Population specific mutations in *BRCA1* and *BRCA2*

The *BRCA1* and *BRCA2* genes have been extensively studied in numerous populations including American and European Caucasians, Ashkenazi Jews, and Asians (reviewed in Lacroix & Leclercq, 2005), but little work has been done in the African black populations, especially the different South African black populations.

A recent whole genome sequencing project that looked at Southern African black populations found that great sequence variation and sequence diversity exists between the Caucasian population and the ancestral African black population (Schuster et al., 2010). Further, the African Genome Variation Project provided detailed genotyping data describing the genetic diversity that exists in the sub-Saharan African black populations (Gurdasani, et al., 2015). These findings suggests that the genetic mutation profiles for particular diseases established in one population can not be used to extrapolate the genetic mutation profile data for another population, especially from the Caucasian to African black populations, as the genetic causality of diseases may be completely different in the different ethnical groups.

Numerous studies have shown that different, population specific, *BRCA* founder mutations exist in certain populations, for example, the Ashkenazi Jewish population, the South African white Afrikaner population and the Dutch population (reviewed in Neuhausen, 2000; Reeves et al., 2004). Further, the presence of *BRCA* founder mutations in certain black populations, such as the African-American

black population, Nigerian black population and west African black population (Gao et al., 1997; Mefford et al., 1999; Fackenthal et al., 2012; Zhang et al., 2012), suggest that founder mutations may also be present within the South African black populations.

Zhang, et al. (2012) described the high frequencies of 12 known, recurrent *BRCA1/2* mutations in the Nigerian, African American and Barbadian cohorts. These mutations are Y101X, 1742insG, 4241delTG, M1775R, 4359insC, C64Y, 1623delTTAAA, Q1090X, and 943ins10 in *BRCA1*, and 1538delAAGA, 2630del11, and 9045delGAAA in *BRCA2*. The presence of numerous recurrent *BRCA* mutations within the African cohorts complicates efforts to understand the genetic aetiology of breast cancer caused by *BRCA* mutations but it provides strong evidence that one or more founder *BRCA* mutations may be present within ancient African populations. This highlights the need to investigate Sub-Saharan African populations to determine if the recurrent mutations observed in Western Africa populations are also present in Southern Africa populations.

A recent study investigated a South African Xhosa black cohort (n=16) and a mixed ancestry coloured cohort (n=105) for *BRCA1* and *BRCA2* gene mutations, and the same common mutation was identified (4/16 individuals) in the Xhosa cohort and the coloured cohort (4/105 individuals) (van der Merwe et al., 2012). The p.Ile1924Argfs X38 mutation in *BRCA2* is believed to be a founder mutation in the Xhosa population and the population of mixed ancestry with Xhosa heritage. The presence of possible common founder mutation(s) in other Southern African black populations has yet to be investigated.

1.4. Low penetrance genes & disease susceptibility genes

BRCA1 and *BRCA2* are the most prevalent breast cancer susceptibility genes known to date, however the *BRCA* genes together only accounts for 10%-20% of causality within family breast cancers (Walsh & King, 2007). Other known breast cancer susceptibility genes (including *TP53, PTEN, STK11, PALB2, RAD50, RAD51, CHEK2, ATM, NBS1,* or *BRIP1*) each with variable penetrance, together account for around 35% of familial breast cancer (Walsh & King, 2007; Gracia-Aznarez et al., 2013). This still leaves over 50% of familial breast cancer cases with no known genetic causality.

Recently a large scale genotyping study identified 41 new disease susceptibility loci for breast cancer within populations of European ancestry (Michailidou et al., 2013). 29807 SNPs were selected, based on strong statistical risk correlations with breast cancer, from the meta-analysis of 9 previously published genome-wide genotyping studies. The combined cohort size was 10052 cases verses 12575 controls. The risk correlation to disease of the 29807 SNPs was investigated. The 29807 SNPs were genotyped in a separate cohort of 45290 cases verses 41880 controls. This study resulted in the description of 41 potential new breast cancer susceptibility loci in unselected sporadic breast

cancer cohort (Michailidou et al., 2013). This study highlights the polygenic nature of sporadic breast cancer susceptibility, and the complexity brought about when interpreting the risks these susceptibility loci carry. It is unclear if these susceptibility loci plays a role in hereditary breast cancer.

Such large scale genotyping studies are yet to be performed for the African populations. Due to the large genetic diversity that exists in populations of African ancestry as compared to populations of European ancestry, the valuable data produced by Michailidou et al. (2013) should not be extrapolated across to African populations. The mechanisms underlying breast cancer are similar to a large extent across different ethnic groups, however the underlying molecular contributions may be completely different. This reasoning is based on the different clinical presentation of breast cancer observed between African individuals and non-African individuals (Carey et al., 2006) and the difference in background molecular profile between African individuals and non-African individuals (Haffty et al., 2006). Thus, it is important to investigate the molecular causality of breast cancer in populations of African ancestry.

BRCA1 and *BRCA2* gene mutations plays little to no part in sporadic breast cancers. Sporadic breast cancers may share certain characteristics with breast cancers that arose from *BRCA* mutations, there are certain characteristics in the clinical and pathological data of the cancers that are distinctively different (reviewed in Turner et al., 2002: van der Groep et al., 2006).

1.5. Management of inherited breast cancer

Surgical interventions are often the first line treatment options available and recommended to breast cancer patients (NCCN, 2014a). Depending on the stage and grade of the cancer, surgical treatments may include lumpectomy (removal of a portion of the breast) and/or mastectomy (removal of whole breast/s). Chemotherapy may or may not be offered (NCCN, 2014a).

The identification of the causative genetic mutation for high risk breast cancer patients is vitally important. The presence of an identified genetic pathogenic mutation has a significant impact on the management and treatment of the disease. Management and surveillance programs for inherited breast cancer are much more stringent and require more frequent screening than that of sporadic breast cancer (NCCN, 2014a).

Breast cancer screening strategy recommended by the National Comprehensive Cancer Network (NCCN, 2014b) for inherited breast cancer includes:

For females:

- Women should be familiar with their breasts from the age of 18

- Annual or bi-annual clinical breast examinations from the age 25
- From 25-29 years of age, annual breast MRI
- From 30-75 years of age, annual mammogram and breast MRI
- Possible risk-reducing mastectomy
- Possible risk-reducing salpingo-oophorectomy
- Chemoprevention therapies, such as oral contraceptives and/or chemo-therapeutic drugs such as Tamoxifen for BRCA mutation carriers

For males:

- Breast self-examination from the age of 35 years
- Bi-annual clinical breast examinations from the age of 35 years
- Annual mammogram from the age of 40 years
- Prostate cancer screening from the age of 40 years, this is for *BRCA2* mutation carriers.

Genetic counselling is another vitally important component of the management of inherited breast cancer. It is important to educate patients and relatives regarding the risks and implications of inherited breast cancer. Through genetic counselling, at risk relatives of the patients can be identified (Wainstein, 2011). Treatments are more successful at early stages of the disease (NCCN, 2014a). If a positive *BRCA* mutation is identified in the patient then predictive testing can be offered to all at-risk family members. This will aid in the surveillance for the development of breast cancer and/or other related cancers. If a familial mutation is identified but it is not present in an at-risk family member, then the at-risk family member can be reassured. Genetic counsellors should be careful to explain to mutation negative females that they still have a population risk to the development of breast cancer.

1.6. Research motivation

Very little research investigating the molecular aetiology of inherited breast cancer amongst southern African black populations has been conducted (At the time of this study, molecular aetiology of inherited breast cancer have only been conducted in the Xhosa/mix ancestry population in the Western Cape, South Africa.). The research data available in the public domain are generated using populations primarily of European decent. It is vitally important to fill the knowledge gap that exists with regards to inherited breast cancers amongst women from the Southern African black population. The spectrum of mutations of inherited breast cancer differ greatly across different population groups thus leading us to speculate that the molecular aetiologies responsible for

inherited breast cancer amongst women of Southern African ancestry are possibly different to any other population.

By knowing and understanding the molecular aetiology that drives the disease, patients of Southern African ancestry may benefit from earlier disease detection, targeted disease treatments, effective disease management and ultimately, increased overall survival rate. This study will also investigate the possible presence of founder mutations within the Southern African black populations and to evaluate the prevalence of the previously described Xhosa founder mutations within high-risk breast cancer patients.

This study is a first step, looking into the molecular aetiology of inherited breast cancer in the South African black population, with the long term goal of identifying and understanding the role genetics plays in the development of breast cancer in this population group.

1.6.1. Aims

The main aim of this study is to gain a better understanding of the genetic aetiology underlying breast cancer in South African black women and specifically to determine whether inherited breast cancers are caused by mutations in the *BRCA1* and *BRCA2* genes. DNA sequencing analysis and Multiplex Ligation-dependent Amplification (MLPA) analysis of these genes will allow us to establish whether any germline mutations are present and if these mutations are contributing to the increased incidence of breast cancer in this population, perhaps because of population-specific founder mutations.

1.6.2. Objectives

1. To recruit a minimum of 30 South African black female individuals with breast cancer, at high risk of having inherited disease, based on a predetermined high risk profile: <50 years old at age of onset of breast cancer (with or without family history) OR >50 years old at age of onset of cancer with positive family history of breast cancer and/or high risk histology.

2. To perform sequencing analysis using Sanger sequencing chemistry, to sequence all the exons and flanking regions of *BRCA1* and *BRCA2* to screen for and analyse small putative pathogenic mutation/s and sequence variants.

3. To perform *BRCA1* and *BRCA2* MLPA analysis to screen for possible large gene rearrangements or gene dosage alterations.

4. To perform genotype allele frequency comparison between the BRCA cohort and the control groups.

2. SUBJECTS AND METHODS

This chapter discusses the subjects and methods of this study. Selection criteria for the subjects as well as methodologies for both data collection and analysis used in this study are detailed below.

In summary, 33 high risk black breast cancer patients were recruited for this study. Sanger sequencing and Multiplex Ligation-dependent Probe Amplification (MLPA) analysis were performed on genomic DNA (extracted from peripheral blood samples) to screen for pathogenic mutations in the *BRCA1* and *BRCA2* genes.

2.1. Ethics clearance

This study made use of retrospectively collected blood samples collected from black breast cancer patients recruited at the Breast and Plastic Clinic, Chris Hani Baragwanath Hospital, Johannesburg, South Africa. DNA analysis was performed on these samples. Ethics approval for this study was obtained from the Human Research Ethics Committee (Medical), Faculty of Health Sciences, the University of the Witwatersrand, reference number: M110922 (Appendix A).

An ethnically matched Sub-Saharan African control cohort was genotyped in a separate study (May et al., 2013). These genotype data were used for allelic frequency comparison and statistical analysis. The genotyping study was performed under the ethical approval certificate reference number: M10745 (Appendix B).

2.2. Subjects and controls

2.2.1. Subjects

Patients for this study were selected from a larger cohort of high risk (with positive family history and/or early-onset) breast cancer patients recruited (between 2009 - 2011) at the Breast and Plastic Clinic, Chris Hani Baragwanath Hospital, Johannesburg, South Africa. Patients attending the clinic were routinely counselled by the genetic counsellors, registrars, and medical geneticists of the Clinical Section of the Division of Human Genetics, National Health Laboratory Service (NHLS), and the University of the Witwatersrand (WITS). Patients were routinely offered the option of giving a blood sample and of having their DNA banked, with informed consent, giving the NHLS permission for the purpose of possible future diagnostic testing and/or the use of the DNA sample in research projects.

Any positive findings found in this study will be confirmed by the Molecular Diagnostic Laboratory, NHLS under a diagnostic setting and the results will be fed back to the patients. Further predictive screening will also be made available to the at-risk family relatives.

A total of 33 female South African black patients, falling into the high risk category, were selected for this study. Selection criteria were as follows:

1. Patients with early-onset breast cancer (≤50 years of age at time of disease diagnosis) or patients with positive family history (multiple cases of breast or other associated cancers occurring within first and/or second degree relatives).
2. Patients with bilateral breast cancer.
3. Patients with triple negative breast cancer diagnosed before the age of 50 years.

Breast cancer patients fulfilling any one of the three criteria above were considered for this study.

The patients' age at diagnosis, family history, breast cancer receptor status, breast cancer laterality and risk categorization are presented in Table 2.1. This cohort will be referred to as the "BRCA cohort" in the dissertation.

Full family histories of disease and disease related conditions were not available for all patients. For certain patients, only partial family history was available. Full family history is defined as when there is history of the disease and/or related disease in the family and the patients were able to provide detailed medical history relating to the disease from both sides of the family (maternal and paternal line). Patients' family histories were classified as "partial" when patients were unable to provide a full family history, from both sides of the family and from all individuals within the family. Patients either did not know the medical history of certain individuals within the family or no information was available. With regards to TNBC classification, patients were classified as "yes" for TNBC when histology was negative for all three receptors. The receptors are the oestrogen receptor (ER), progesterone receptor (PR) and the human epidermal growth factor receptor 2 (HER2/neu). If the patients' histology was positive for any of the three receptors then they were classified as "no" for TNBC status.

The number of patients recruited for this study was set at 33 patients due to limited funding resources for this project.

Table 2.1. Summary of the patient information of the BRCA cohort.

Sample ID	Age at Dx	Family History	TNBC	Laterality	Risk Category*
BRCA001	48	no	no	unilateral	moderate
BRCA002	38	yes	no	unilateral	high
BRCA003	39	yes	yes	unilateral	high
BRCA004	38	partial	yes	unilateral	high
BRCA005	48	partial	no	unilateral	moderate
BRCA006	40	no	yes	unilateral	moderate
BRCA007	30	yes	yes	unilateral	high
BRCA008	44	partial	no	unilateral	moderate
BRCA009	42	partial	no	unilateral	moderate
BRCA010	37	partial	no	unilateral	high
BRCA011	41	no	no	unilateral	high
BRCA012	32	no	no	unilateral	high
BRCA013	43	yes	no	unilateral	high
BRCA014	42	yes	yes	unilateral	high
BRCA015	33	no	no	unilateral	high
BRCA016	45	no	no	unilateral	moderate
BRCA017	40	partial	unknown	unilateral	moderate
BRCA018	49	partial	no	unilateral	moderate
BRCA019	27	yes	no	unilateral	high
BRCA020	50	yes	no	unilateral	high
BRCA021	36	no	unknown	unilateral	high
BRCA022	43	no	unknown	unilateral	moderate
BRCA023	45	no	no	unilateral	moderate
BRCA024	31	partial	no	unilateral	high
BRCA025	32	yes	no	unilateral	high
BRCA026	49	no	no	bilateral	high
BRCA027	39	yes	no	unilateral	high
BRCA028	30	yes	Yes	unilateral	high
BRCA029	31	no	unknown	unilateral	high
BRCA030	45	yes	no	unilateral	high
BRCA031	29	yes	no	unilateral	high
BRCA032	28	no	no	bilateral	high
BRCA033	46	yes	no	unilateral	moderate

*-Risk calculation and categorisation was performed by Wainstein (2011) in a related study. Matrices described by Wainstein (2011) scored the patients to be either Moderate risk or High risk based on available clinical information of the patients. Please refer to Wainstein (2011) for a detailed explanation of the matrices.

2.2.2. Controls

Two data sets were used as controls for comparison of allele frequencies for this study:

2.2.2.1. The Control group

Firstly, genotyping data from an unselected, random, ethnically-matched control group of 94 individuals was available and was used in this study to obtain the background population allele frequencies. This control group comprised a subset of 94 black southern African individuals participating in a large study within the University of the Witwatersrand, the Birth-to-Twenty study. These individuals were recruited from the metropolitan area of Johannesburg-Soweto. These individuals were recruited in 1990, at the start of the Birth-to-Twenty study, to participate in a longitudinal study that followed and studied various aspect of their health over the next twenty years and the cohort and study was thus named Birth-to-Twenty (Richter et al., 2007). This Control group was not gender and age matched to the BRCA cohort.

This subset of 94 individuals was genotyped on a genome wide level, in a separate study undertaken by another MSc student in the department, using the Illumina Human Omni 5 BeadChip (May et al., 2013). A total of 4.3 million SNPs were genotyped in this cohort and the SNP data specific to the *BRCA1* and *BRCA2* gene were mined and used for comparative statistical analysis against the BRCA cohort's allele frequency data. The Illumina Human Omni 5 Beadchip is designed to contain the relatively frequent or common SNPs previously detected. The absence of a SNP on this array does not imply that the SNP is not in the sample cohort, but rather a limitation of design of the Omni 5 BeadChip. This black southern African control group is ethnically matched to the BRCA cohort and they are also from the same geographical location, however no phenotypical data of this group was made available to our study. Thus this control group would serve only as a population allele frequency reference for genotype risk analysis and comparison.

This group will be referred to as the "Control group" in the dissertation.

2.2.2.2. The 1KG group

Secondly, the 1000 Genomes (1KG) Project phase I data were released during the course of our study (1000 Genomes Project Consortium, 2012). This enabled us to mine and utilize the SNP data from the 1KG project for statistical comparisons and genotype risk analysis calculations. The African Super population data subset from the 1KG project was mined and utilized as a second reference / normal population for statistical analysis.

The African Super population consisted of individuals of the following ethnicities and origins: Yoruba in Ibadan, Nigeria; Luhya in Webuye, Kenya; Gambian in Western Division, The Gambia; Mende in Sierra Leone; Esan in Nigeria; African Ancestry in Southwest United States of America; and African

Caribbean in Barbados (1000 Genome Project Consortium, 2012). This super population consists of 246 individuals.

This group will be referred to as the "1KG group" in the dissertation.

However, it is important to note that no southern African populations were sampled and studied in the 1000 Genome Project. The 1KG cohort is therefore not a genetically well-matched reference population for southern African populations. This study also investigated the feasibility and viability of using the 1KG cohort as a population proxy for southern African population studies looking at the genetic aetiology of breast cancer.

2.3. Methods

Sequencing analysis was undertaken to screen for point mutations and small insertions and deletions (INDELs) at the DNA nucleotide level within the *BRCA1* and *BRCA2* genes, using Sanger sequencing chemistry and to screen for large gene/exonic deletions and/or duplications using MLPA analysis.

2.3.1. Sample preparation

Genomic DNA was extracted from peripheral blood samples collected from the breast cancer patients. DNA extraction was performed by the staff of the Molecular Diagnostic Laboratory, NHLS, and WITS. The standard salting out method was used (Miller et al., 1988). The DNA samples were stored at 4°C until use.

The quality and integrity of the DNA samples were ascertained by running an aliquot of the DNA samples on a 0.8% agarose gel (Bioline, UK) (Appendix C) and DNA concentration was determined by Nanodrop quantification using the ND1000 (Thermo Scientific, USA). All DNA samples were diluted to 50ng/µl for the PCR applications and Sanger sequencing.

2.3.2. PCR and Sanger sequencing

This study made use of oligo primer sequences already described in the literature. The molecular techniques, PCR and Sanger sequencing, used in this study are also well described (Sanger et al., 1977; Bartlett & Stirling 2003).

2.3.2.1. Primer design

The sequences of oligonucleotide primers designed for the amplification of the coding regions of the *BRCA1* and *BRCA2* gene were obtained from the National Genetics Reference Laboratory (NGRL) (Wessex, UK) (Mattocks et al., 2010). The oligonucleotide primers were designed to flank the coding regions (coding exons), including the intron/exon boundaries of the *BRCA1* and *BRCA2* genes. This will be referred to as the target region. All primers incorporated a 5' universal tag sequence to allow

for high throughput Sanger sequencing preparation. The universal tag sequence setup allows for the sequencing of all tagged amplicons with one primer (Figure 2.1.).

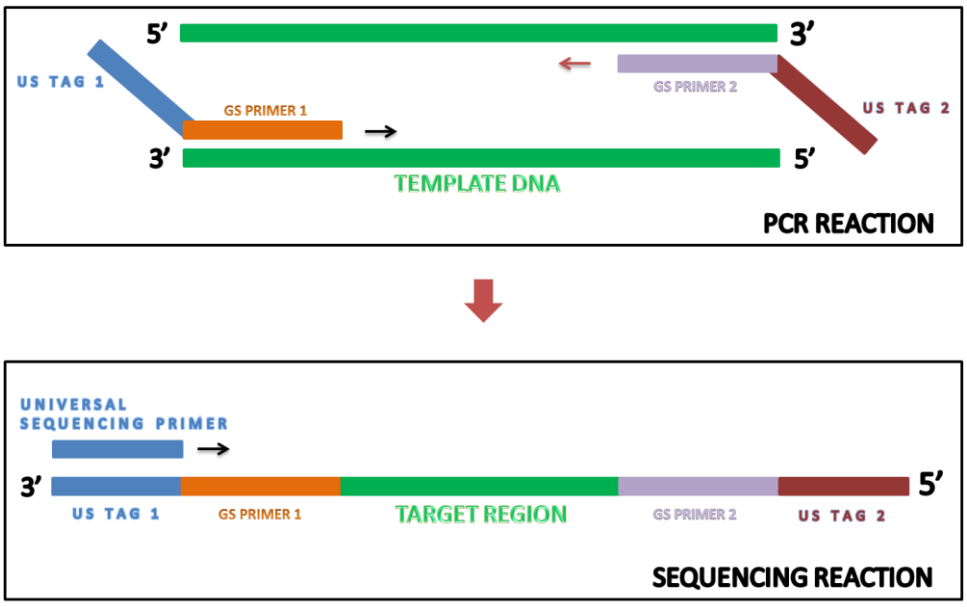


Figure 2.1. Universal tag sequencing primer design. The 5' end of all gene-specific (GS) primers were tagged with universal primer sequences for high-throughput Sanger sequencing applications. Two different universal tag sequences were used to tag the GS primers, to allow for forward and reverse sequencing reactions.

2.3.2.1.1. *BRCA1* primers

The *BRCA1* gene target regions for PCR amplification comprised of 22 coding exons. Certain exons were too large to be amplified and sequenced in a single amplicon. These exons were amplified and sequenced in multiple fragments. The target regions were amplified using 33 primer pairs, generating ~12kb of readable sequence per patient. Amplicon maps for the *BRCA1* gene can be found in Appendix D (*BRCA* primer list). The amplicons are shown in relation to the target regions and coding exons of the *BRCA1* gene. *BRCA1* primers were designed based on the reference sequence version HGVS: NM_007294.2. The primers were checked for primer-binding polymorphisms using dbSNP build 129 (Mattocks et al., 2010).

2.3.2.1.2. *BRCA2* primers

The *BRCA2* gene target regions for PCR amplification comprised of 25 coding exons. As for *BRCA1,* large exons were amplified and sequenced in multiple fragments. The *BRCA2* target regions were amplified using 46 primer pairs, generating ~19kb of readable sequence per patient. Amplicon maps for the *BRCA2* gene can be found in Appendix D (*BRCA* primer list). The amplicons are shown in

relation to the target regions and coding exons of the *BRCA2* gene. *BRCA2* primers were designed based on the reference sequence version HGVS: NM_00059.3. The primers were checked for primer-binding polymorphisms using dbSNP build 129 (Mattocks et al., 2010).

Full primer design information and sequences for the *BRCA1, BRCA2* and the Universal Sequencing (US) primers can be found in Mattocks et al. (2010), supplementary data B. This document is attached in Appendix D.

Numerous sets of *BRCA* specific oligonucleotide primers have been published in the literature. The set designed by Mattocks et al. (2010) was chosen due to the benefit of the universal sequencing tag sequence present at the 3' end of each primer, which streamlines the downstream sequencing reaction.

2.3.2.2. PCR amplification conditions

The *BRCA1* target regions were amplified using the MyTAQ HS RED polymerase system (Bioline, UK). Each PCR mix comprised the following ingredients: 50 ng of gDNA; 0.1 pmol of forward primer; 0.1 pmol of reverse primer; 1x MyTAQ HS RED reaction buffer, 1 unit of MyTAQ HS RED polymerase and ddH_2O, made up to the final reaction volume of 25 µl (Appendix E – Section 7.5.1.).

The *BRCA2* target regions were amplified using the KAPA Taq ReadyMix with loading dye (Kapa biosystems, SA). Each PCR mix comprised the following ingredients: 50 ng of gDNA; 0.1 pmol of forward primer; 0.1 pmol of reverse primer; 1 x Kapa Taq ReadyMix with loading dye and ddH_2O, made up to the final reaction volume of 25 µl (Appendix E – 7.5.2.).

The change of the polymerase system as the project progressed was done purely for financial reasons. The KAPA Taq ReadyMix with loading dye was significantly cheaper than the MyTaq HS RED polymerase system. Both polymerase systems performed equally well under the same cycling conditions (data not shown).

One thermal cycling condition was used to amplify all the target regions. Thermal cycling was performed in the T100 thermal cycler (Biorad, USA). The thermal cycling conditions consisted of an initial denaturation step of 2 minutes at 95°C, followed by 40 cycles of: 25 seconds at 95°C; 30 seconds at 61°C; 60 seconds at 72°C; and a final elongation of 5 minutes at 72°C. The PCR products of the target regions were held at 15°C until the next step.

2.3.2.3. *Post PCR gel electrophoresis*

The PCR products were checked for amplification and sized on a 1.5% EtBr-stained agarose gel (Bioline, UK) (Appendix C). Five µl of PCR product for each reaction was sized against a 1KB^{+} ladder (Invitrogen, USA). The agarose gels were resolved at 3 volt/cm for 30 minutes.

The agarose gel was visualized under UV using the G: BOX Gel documentation system (Syngene, UK).

2.3.2.4. *Post PCR clean-up*

PCR products were isolated from the reaction mix prior to Sanger sequencing setup. The clean-up/isolation step was performed using the Montage PCR96 Cleanup kit (Merck, UK) with the amended protocol below:

Eighty µl of ddH_2O was added to the 20 µl of PCR product, the mixture was then transferred into an empty well on the Montage PCR 96 plate. The plate was placed onto a vacuum manifold and vacuumed for 8 minutes at 381 mm Hg or 7.36 psi. The PCR products are retained by the filter membrane within the wells while all other reaction components and wastes are removed through the vacuum suction.

Once the vacuuming was completed, the Montage PCR96 plate was removed from the vacuum manifold. The bottom of the Montage PCR96 plate was then blotted dry with a paper towel.

Eighty µl of ddH_2O was added to the wells containing the PCR products to wash the PCR products. The Montage PCR96 plate was then vacuumed again, for 5 minutes, at 381 mm Hg or 7.36 psi.

The Montage PCR96 plate was then dried again after washing. The PCR products were resuspended in 20 µl of ddH_2O. The 20 µl ddH_2O was pipetted up and down twenty (20) times to allow adequate mixing and resuspension of the PCR products.

Resuspended PCR products were transferred to a 96 well V-bottom plate (Axygen, USA) for storage at -20°C and used in Sanger sequencing setup.

2.3.2.5. *Sanger sequencing setup*

Sanger sequencing was performed using the BigDye Terminator v3.1 CycleSeq kit (Life Technologies, USA). Each cycle sequencing reaction consisted of the following ingredients: 4 µl of cleaned PCR product; 0.1 pmol of US tag primer; 1 µl of BigDye Terminator v3.1 Reaction mix; 1.5 µl of BigDye Terminator v3.1 Reaction buffer; and ddH_2O, making up to a total reaction volume of 12 µl (Appendix E – Section 7.5.3.).

Thermal cycling of the cycle sequencing was performed in the T100 thermal cycler (Biorad, USA). The cycling conditions consisted of 30 cycles of 30 seconds at 95°C, 15 seconds at 60°C, 2 minutes at 72°C. The cycle sequencing products were held at 24°C until the next step.

2.3.2.6. *Post cycle sequencing clean-up*

Post cycle sequencing clean-up was performed using the Montage PCR96 plate (Merck, UK). The following amended protocol was used:

The cycle sequencing products were loaded onto the Montage PCR96 plate with 20 µl of Montage Injection Solution (Merck, UK). The Montage PCR96 plate was then placed under vacuum for 2 minutes at 381 mm Hg or 7.36 psi. The bottom of the Montage PCR96 plate was dried after vacuuming, using paper towels.

The cycle sequencing products were then washed twice. Each wash consisted of adding 20 µl of Injection Solution into each well and then vacuuming for 1 minute 30 seconds at 7.36 psi. The bottom of the Montage PCR96 plate was dried using paper towels after each wash step.

The cycle sequencing products were resuspended in 20 µl of Injection Solution (Merck, UK). The cycle sequencing products were adequately mixed by pipetting up and down for 20 times. The cycle sequencing products were then transferred to a semi-skirted 96 well plate (Axygen, USA) for capillary gel electrophoresis.

2.3.2.7. *Capillary gel electrophoresis and base calling*

Capillary gel electrophoresis of the cycle sequencing products was performed on the ABI 3130xl genetic analyser (Life Technologies, USA). The cycle sequencing products were resolved using the ABI POP7 polymer (Life Technologies, USA) on the ABI 3130xl genetic analyser. The ABI POP7 polymer allowed for the resolution of fragments with one nucleotide difference in size. The laser excited the fluorescently labelled ddNTPs at the end of the fragments; the fluorescent signal was then collected by the Data Collection v3.0 software (Life Technologies, USA).

Base calling was done using the KB Basecaller v1.2 algorithm in the Sequencing Analysis Software v5.2 (Life Technologies, USA) software package. Raw sequencing data were presented as chromatograms for downstream sequence analysis.

2.3.3. MLPA analysis

MLPA (multiplex ligation-dependent probe amplification) analysis was performed to screen for gene rearrangements and/or gene dosage alterations, including whole gene and/or whole exon deletions and/or duplications, in both the *BRCA1* and *BRCA2* genes.

BRCA1 MLPA analysis was performed using the SALSA MLPA P002-C2 (Lot. 0811) probemix (MRC-Holland, the Netherlands). The P002-C2 probemix consists of 35 MLPA probes. Of these, 26 are *BRCA1* specific probes, and 9 genomic reference probes. In addition to the 35 MLPA probes, there are 9 control probes, which consist of 4 DNA quality control probes, 3 denaturation control probes, 1 X chromosome probe, and 1 Y chromosome probe, to check the MLPA reaction quality.

BRCA2 MLPA analysis was performed using the SALSA MLPA P090-A4 (Lot. A4-0712) probemix (MRC-Holland, the Netherlands). The P090-A4 probemix consists of 43 MLPA probes. Of these, 31 are *BRCA2* specific probes, 2 *BRCA2* gene flanking probes and 10 genomic reference probes. In addition to the 43 MLPA probes, there are 9 control probes, which consist of 4 DNA quality control probes, 3 denaturation control probes, 1 X chromosome probe, and 1 Y chromosome probe, to check the MLPA reaction quality.

All probes were fluorescently labelled with FAM dye for fragment detection and sizing using capillary gel electrophoresis on the ABI 3130xl genetic analyser.

MLPA setup was performed according to the manufacturer's protocol. The MLPA protocol in summary is as follows: DNA samples were first standardised to a concentration of 40ng/µl. 200ng of DNA was then denatured at 98°C for 5 minutes. 1.5 µl of the SALSA MLPA probe mix and 1.5 µl of MLPA buffer were then added to the denatured DNA. Probes were allowed to hybridize to the denatured DNA at 60°C for 18 hours. 1.0 µl of Ligase-65, 3.0 µl of Ligase-65 buffer-A and 3.0 µl of Ligase-65 buffer-B were then added to the reaction mix to ligate the hybridized probes. Ligation took place at 54°C for 15 minutes, followed by 5 minutes at 98°C to inactivate the Ligase-65 enzyme. PCR amplification of the hybridized probes then took place to amplify the ligated probes. 2.0 µl of the SALSA PCR primer mix, 7.5 µl of ddH_2O, 0.5 µl of the SALSA polymerase and 40 µl of the MLPA ligated product made up the reaction mix. The PCR conditions were: 35 cycles of 30 seconds at 95°C; 30 seconds at 60°C; 1 minute at 72°C, followed by 20 minutes at 72°C. The reaction mix was held at 4°C post PCR amplification. The amplified probes were then resolved and sized using capillary gel electrophoresis. The probes were sized against the ABI LIZ-500 marker. The signal intensities of the amplified probes were quantified using the native ABI Data Collection software v.3 (Life, USA) to determine whether large gene/exonic deletions and/or duplications are present. Detailed protocol can be found in Appendix E – Section 7.5.4.

The P002 and P090 probe sequences can be found in Appendix F.

The use of MLPA analysis to detect the presence of gene/exon deletions and/or duplications for *BRCA1* and *BRCA2* has been well described in the literature (Hogervorst et al., 2003; Engert et al., 2008; Kuusisto et al., 2011; Ruiz de Garibay et al., 2012).

2.3.4. Data analysis

The data analysis carried out in this study followed a step-wise workflow. The steps are outlined below in Figure 2.2. and Figure 2.3.

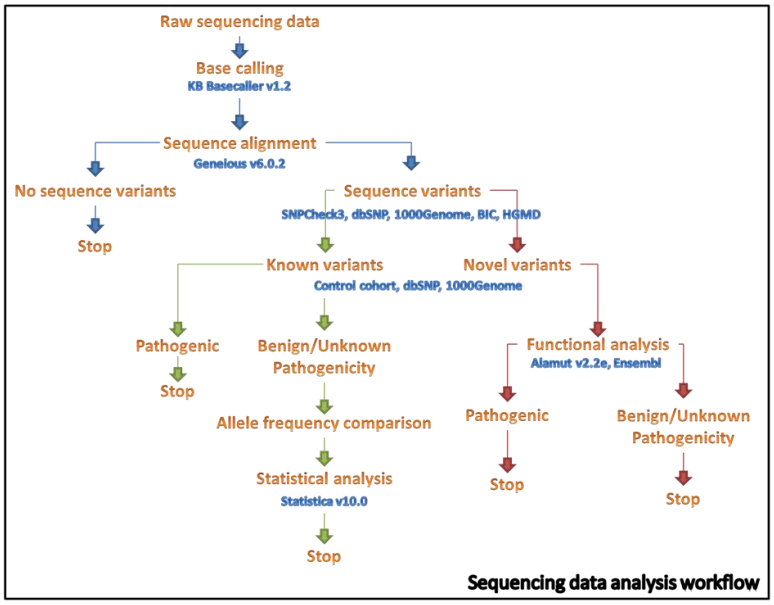


Figure 2.2. Sequencing data analysis workflow. Orange labels indicate milestones/markers critical to decision making. Blue labels indicate bioinformatic tools, online databases and computational programs used for analysis. Different colour arrows indicate workflow pathway.

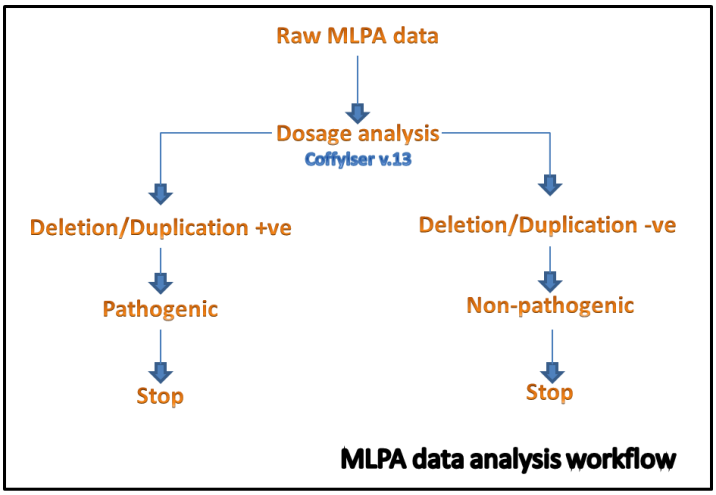


Figure 2.3. MLPA data analysis workflow. Step-wise data analysis used in the analysis of MLPA data..

2.3.4.1. *Sequence alignment*

Sequence alignments were performed using Geneious v6.0.2 (Biomatters, New Zealand). The Reference sequence GRCh37.p5 Primary Assembly NC_000017.10, residues 41196312 to 41277500, was used as the reference sequence for *BRCA1* sequence alignments. The mRNA transcript HGVS: NM_007294.3 was used as the primary mRNA transcript for variant naming and analysis.

The Reference sequence GRCh37.p5 Primary Assembly NC_000013.10, residues 32889617 to 32973809, was used as the reference sequence for *BRCA2* sequence alignments. The mRNA transcript HGVS: NM_000059.3 was used as the primary mRNA transcript for variant naming and analysis.

Pairwise/Multiple align was performed using the Geneious alignment algorithm. The IUB sequence scoring cost matrix, set out by the International Union of Biochemistry, was used to allow standardization of sequence analysis and to ensure reproducibility of the results in future studies. Alignment parameters used were as follows: Cost matrix: 70% similarity (IUB) (5.0/-4.5); Gap open penalty: 12; Gap extension penalty: 3; Alignment type: Global alignment (Needleman-Wunsch).

Nucleotide heterozygosity was annotated automatically using the Find Heterozygosity plugin tool in Geneious v6.0.2 then confirmed with manual analysis. Any overlapping nucleotide signals (nucleotide signals overlapping on the same location) with signal peak heights within 60% or higher ratio/proportion to the other on the chromatogram were flagged as a heterozygous nucleotide.

2.3.4.2. *Sequencing variants mapping and naming*

Sequence variants (defined as any sequences that differs from the reference sequence) detected post sequence alignment were mapped to the reference genome sequence built 19 (GRCh37 – hg19) using SNPCheck3 (NGRL, UK). SNPCheck3 was designed for the purpose of screening for primer-binding polymorphisms. By using the variant nucleotide as part of the primer sequence, SNPCheck3 was able to BLAST the sequence variant against dbSNP, Ensembl and the 1000 Genomes database (1KG) to determine if the variant had been previously reported. For ease of analysis, the sequence variants found during sequence alignments were designated as the last base of the forward primer.

2.3.4.2.1. Novel sequence variants

A sequence variant was classified as novel when no dbSNP or Ensembl or 1KG data entry was found by SNPCheck3. The novel variants were then named using the Mutalyzer v2.0 online software (Wildeman et al., 2008). Mutalyzer v2.0 was able to generate cDNA (c.XXX) nomenclature and gDNA (g.XXX) nomenclature for the variant, based on the genomic location of the variant determined by SNPCheck3 BLAST.

The functional *in-Silico* analysis of the novel variants was conducted using the Ensembl Variant effect predicator (VEP) online software (McLaren et al., 2010) and the Alamut version 2.2e software package (Interactive Biosoftware, France). Alamut then categorizes novel variants based on the combined evidence presented by the pathogenicity predication algorithms into one of the 5 class scheme: Class 1 – Benign/Certainly not pathogenic; Class 2 – Unlikely pathogenic; Class 3 – Unknown pathogenicity; Class 4 – Likely pathogenic; Class 5 – Certainly pathogenic.

The pathogenicity prediction algorithms used for the analysis are discussed later in this chapter.

2.3.4.2.2. Previously reported variants (Known variants)

A sequence variant was classified as previously reported when data entry for the variant was found by SNPCheck3. SNPCheck3 data set included: cDNA nomenclature; gDNA nomenclature; chromosomal location; rsID; reference allele; variant allele; Breast cancer Information Core (BIC) mutation database data entry; dbSNP database data entry, Ensembl data entry and 1KG database data entry.

Functional *in-Silico* analysis was performed for the previously reported variants using VEP and Alamut v2.2e. *In-Silico* analysis data were used in conjunction with public domain data to classify the functions of the sequence variants as the pathogenicity of previously reported variants were not always clear and available from public domain data.

The variants were subsequently classified as per the class scheme mentioned above, based on the combination of data available in the public domain data and in-silico analysis data.

For "Variant of unknown pathogenicity", the allele frequency of the variants was compared with general population allele frequency data obtained from the 1KG cohort (1000 Genome Project Consortium, 2012), as well as the Control cohort allele frequency data obtained using the Human Omni 5 BeadChip (Illumina, USA) (May et al., 2013).

A "Variant of unknown pathogenicity" was then classified as "Likely to be pathogenic" or "Unlikely to be pathogenic" where possible. This was done only when sufficient combined evidences from population allelic frequency data and *in-Silico* analysis data were available. Population allelic frequency threshold were set to less than 0.001.

2.3.4.3. Bioinformatic analysis

Bioinformatic analysis of the sequence data was conducted using a combination of software packages and web based tools. Several prediction algorithms were used in conjunction to strengthen the functional predictions of the sequence variants. The Alamut v2.2e and Ensembl VEP were

primarily used for functional analysis of the sequence variants. The Alamut v2.2e software package combined the predictions from the three missense prediction algorithms and five splice site prediction algorithms to provide evidence for classifying the sequence variants. Mutation databases such as BIC, LOVD and HGMD were also incorporated and accessed through Alamut v2.2e for ease of data management.

2.3.4.3.1. Mutation databases

Three human mutation databases were used during this study. Sequence variants were searched against mutation databases to determine if the variants have been previously reported and to determine the pathogenicity of the previously reported variants.

a) HGMD® Professional – **H**uman **G**ene **M**utation **D**atabase (http://www.hgmd.org/) Professional is a comprehensive professionally curated mutation database containing genetic and genomic data for human inherited disease research (Biobase, Germany). This database operates on an annual subscription fee for access basis. This database was used to check for previously reported genetic mutations within the *BRCA1* and *BRCA2* genes.
b) LOVD – **L**eiden **O**pen-source **V**ariation **D**atabase (http://www.lovd.nl/) is an open source, free to use, web based mutation database. This database curates human genetic variations based on HGVS recommendation. Genotype and phenotype data and their association with disease are stored and referenced in this database (Fokkema et al., 2011).
c) BIC – **B**reast cancer **I**nformation **C**ore (http://lgdfm3.ncifcrf.gov/bic/BIC.html) is an open access online breast cancer mutation database curated and hosted by the National Human Genome Research Institute (NHGRI), NIH, USA. This database is free to access and it hosts comprehensive sequence variants entries for the *BRCA1* and *BRCA2* genes (Szabo et al., 2000).

2.3.4.3.2. *In-Silico* analysis algorithms

In total, four missense mutation pathogenicity prediction algorithms were used to predict the function and potential pathogenicity of coding sequence variants. Five splice site prediction algorithms were used to predict the pathogenicity of coding sequence variants and intronic variants.

2.3.4.3.2.1. Missense prediction algorithms

a) SIFT – SIFT predicts pathogenicity of amino acid substitution by **s**orting **i**ntolerant **f**rom **t**olerant amino acid substitutions (Ng & Henikoff, 2001). SIFT searches and aligns the protein sequence to species closely related to humans. The reasoning is that amino acid codons important to the function of the protein would be highly conserved across species. Alteration to these highly conserved amino acids would be intolerable. Physical amino acid

properties are taken into consideration too. SIFT calculates normalized probabilities of the positional amino acid substitution. A score below 0.05 is predicted deleterious, and a score more than or equal to 0.05 is tolerated.

b) PolyPhen-2 – **Poly**morphism **Phen**otyping v**2** predicts the impact of the amino acid substitution on the structure of the protein by analysing the structural and functional change induced by the substitution, phylogeny of the substitution and the sequence impact of the substitution (Adzhubei et al., 2010). Amino acid substitutions are scored into one of four categories: benign; possibly damaging (less confident prediction); probably damaging (more confident prediction); unknown (prediction not possible).

c) Align GVGD – Align GVGD predicts the pathogenicity of missense variants by using a combination of amino acid biophysical characteristics and multiple protein sequence alignments (Tavtigian et al., 2006). Amino acid substitutions pathogenicity is reflected within a spectrum from deleterious to neutral.

d) MutationTaster – MutationTaster predicts the functions of variants and alterations by using the Bayes classifier in conjunction with HapMap SNP data (Schwarz et al., 2010). Depending on the features of the allelic alteration and the availability of SNP data from HapMap, MutationTaster predicts the functional impact of the alteration. A p-value closer to 1.0 indicates a strong prediction of potential pathogenicity. If the alteration has been described in HapMap as a SNP then it is automatically classified as polymorphic by MutationTaster.

2.3.4.3.2.2. Splice site prediction algorithms

Five splice site prediction algorithms were incorporated by Alamut v2.2e. Splice site alteration predictions were conducted using the combination of all five algorithms to determine if a sequence variant potentially altered the donor splice site; acceptor splice site; branch site; or created a cryptic splice site. These five algorithms are: SpliceSiteFinder-like; MatEntScan (Yeo & Burge, 2004); NNSPLICE (Reese, Eeckman, Kulp, & Haussler, 1997); GeneSplicer (Pertea, Lin, & Salzberg, 2001); and Human Splicing Finder (Desmet et al., 2009).

SpliceSiteFinder-like and Human Splicing Finder utilize algorithms based on position weight matrices computed from known human constitutive exonic and intronic boundaries and position-dependent logic to predict splice sites. These algorithms evaluate and predict the donor site, acceptor and branch site.

MatEntScan algorithm is based on Maximum Entropy principle. It statistically computes the thermodynamics of the splice site alternation with a mathematical model.

NNSPLICE algorithm makes splice site predictions by Neural networks.

GeneSplicer algorithm is based a combination of mathematical models including the Markov model.

2.3.4.4. *Statistical analysis*

Statistical analysis was performed to determine if the genotypes identified in the BRCA cohort conferred any risk associations with breast cancer by calculating the genotype relative risk associated with the identified known alleles using the two control groups as population references. A second statistical component was added to this study to determine if the allele frequencies of the observed SNPs differed significantly between the different African populations from West Africa and sub-Saharan Africa. All statistical tests were conducted using Statistica Pro version 10 (StatSoft, USA).

2.3.4.4.1. Statistical evaluation of difference in observed SNPs: case versus control

A Two tailed Chi – squared test, testing for the log of likelihood ratios, was performed for the BRCA cohort versus the Control group and for the BRCA cohort versus the 1KG group. Two p-value thresholds were chosen to evaluate significant differences between SNPs observed in the case and control group: p=<0.01 and p=<0.001, p>0.01 were classified as not significantly different, p=<0.01 but p=>0.001 were classified as potentially different, and p=<0.001 were classified as significantly different.

The allele frequencies of the Control group were also tested against the 1KG group to determine if any allele frequencies were significantly different between different African populations. The same p-value thresholds as above were used.

2.3.4.4.2. Genotype relative risk

Odds ratios were calculated for each polymorphic allele. Genotype relative risk was estimated for each individual SNP, under the assumption that the population was in Hardy-Weinberg equilibrium.

For each individual patient, a combined genotype relative risk was estimated, based on their observed genotypes, to determine whether any combined genotypes present in the breast cancer cohort conferred risk for breast cancer development. Statistical evaluation methodology was outlined in (Adank et al., 2011).

Genotype relative risks were estimated for three scenarios: 1) relative risk to population average for homozygotes of the minor allele, 2) relative risk to population average for heterozygotes, 3) relative risk to population average risk for homozygotes of the major allele. For the basis of comparison, the 3 genotypes of each allele were compared to a population average risk of 1. This is under the assumption that the risk allele is a rare allele and it is not present in the majority of the population,

thus the population would have the lowest risk. And by definition that risk value is 1 (DecodeHealth, 2012).

Genotype relative risk was estimated for both the BRCA cohort versus Control group and for BRCA cohort versus 1KG group. Comparison analysis was also performed to determine whether genotype relative risks differed significantly when different populations were used for baseline risk calculation.

2.3.4.5. Identification of co-segregated alleles

The Control group SNP data were further analysed to determine if allele co-segregation was present in the *BRCA1* & *BRCA2* genes in the sub-Saharan African black populations. The SNP data pertaining to the *BRCA1* & *BRCA2* genes for each of the 94 individuals screened in the Control group study were analysed.

2.3.4.6. MLPA data analysis

MLPA data analysis was performed using the Coffalyser.Net software package (v.130112.0327) designed by MRC-Holland (MRC-Holland, the Netherlands).

Coffalyser.Net software package analyses and interprets the raw fragment electrophoresis data collected by the ABI 3130xl genetic analyser. Probe signal peak heights are normalized and dosage quotients (DQ) are calculated and analysed in Coffalyser.Net to determine if any gene/exon dosage alterations exists in the region of interest.

Gene/exon dosage or DQ is visually represented in the form of histograms. The DQ of each probe is represented by one bar on the histogram. Normal gene/exon dosage (DQ) is represented by the peak value range of 0.85 to 1.15; exon deletions are represented by DQ with peak value range of 0.35 to 0.65; and exon duplications are represented by DQ with peak range of 1.35 to 1.65. Peak value ranges of 0.65 to 0.85 and 1.15 to 1.35 are listed as equivocal. Probes with DQs in this range do not yield clear definitive results as this phenomenon may occur due to the chemistry limitations of MLPA.

All results from this study are presented in the next chapter.

3. RESULTS

This chapter describes the findings of this study, divided into four sub sections:

Section – 3.1. Pathogenic mutations as detected on sequence analysis

Section – 3.2. Novel sequence variants

Section – 3.3. Known Single Nucleotide Variants (SNV)

Section – 3.4. MLPA analysis results

In this study, a total of 59 different sequence changes were found in the *BRCA1* and *BRCA2* genes in the study cohort of 33 black, early onset breast cancer patients. Three variants were very likely pathogenic, 5 were novel sequence variants, and 51 were previously reported single nucleotide variants (SNV). No large gene deletions and/or duplications were detected in the patient cohort using MLPA analysis. The sequence variants are summarized in Table 3.1. and Table 3.2. below for *BRCA1* and *BRCA2* respectively and presented in detail in the sections that follow. Detailed sequence variants per patient can be found in Appendix G.

Table 3.1. Summary of all sequence variants detected in the *BRCA1* gene. The single nucleotide variants are ordered by cDNA nucleotide number. The novel variants are highlighted in blue and the novel pathogenic variant are highlighted in red.

cDNA Nomenclature	rsID	Protein change	Mutation class	Pathogenicity	Observed frequency in BRCA cohort	Present in Control group*	Present in 1KG group*
c.212+66 A>G	Novel	-	Intronic Variant	Unlikely	1/66	No	No
c.306 A>C	Novel	-	Synonymous	Unlikely	1/66	No	No
c.431 dupA	Novel	p.Asn144LysfsX15	Frameshift	Pathogenic	1/66	No	No
c.1724 A>G	rs111539978	p.Glu575Gly	Missense	Unlikely	3/66	No	No
c.1971 A>G	rs28897679	-	Synonymous	Benign	1/66	Yes	Yes
c.1972 A>G	rs55932871	p.Met658Val	Missense	Unlikely	1/66	No	No
c.2082 C>T	rs1799949	-	Synonymous	Benign	1/66	Yes	Yes
c.2120 G>A	rs80357192	p.Gly707Asp	Missense	Unlikely	1/66	No	No
c.2311 T>C	rs16940	-	Synonymous	Benign	2/66	Yes	Yes
c.2612 C>T	rs799917	p.Pro871Leu	Missense	Benign	62/66	Yes	Yes
c.3113 A>G	rs16941	p.Glu1038Gly	Missense	Benign	3/66	Yes	Yes
c.3418 A>G	rs2227945	p.Ser1140Gly	Missense	Benign	4/66	Yes	Yes
c.3548 A>G	rs16942	p.Lys1183Arg	Missense	Benign	4/66	Yes	Yes
c.3751 T>G	Novel	p.Cys1251Gly	Missense	Unknown	1/66	No	No
c.4113 G>A	rs147448807	-	Synonymous	Benign	1/66	No	No
c.4308 T>C	rs1060915	-	Synonymous	Benign	2/66	Yes	Yes
c.4600 G>A	rs55815649	p.Val1534Met	Missense	Unlikely	1/66	No	Yes
c.4837 A>G	rs1799966	p.Ser1613Cys	Missense	Benign	4/66	Yes	Yes
c.4987-20 A>G	rs80358035	-	Intronic Variant	Benign	1/66	Yes	Yes
c.5332+78 C>T	Novel	-	Intronic Variant	Unlikely	1/66	No	No
c.5406+8 T>C	rs55946644	-	Intronic Variant	Benign	1/66	Yes	Yes
c.5468-10 C>A	rs8176316	-	Intronic Variant	Benign	1/66	Yes	Yes

* Allele frequencies of the *BRCA1* SNVs found in the Control group are listed in Table 3.3. and the allele frequencies for the *BRCA1* SNVs found in the 1KG group are listed in Table 3.5.

Table 3.2. Summary of all sequence variants detected in the *BRCA2* gene. The single nucleotide variants are ordered by cDNA nucleotide number. The novel variants are highlighted in blue and the pathogenic variants are highlighted in red.

cDNA Nomenclature	rsID	Protein change	Mutation Class	Pathogenicity	Observed frequency in BRCA Cohort	Present in Control group	Present in 1KG group
c.-11C>T	rs76874770	-	5' UTR	Benign	3/66	Yes	Yes
c.-26G>A	rs1799943	-	5' UTR	Benign	3/66	Yes	Yes
c.517-19C>T	rs11571623	-	Intronic	Benign	1/66	Yes	Yes
c.517-4C>G	rs81002804	-	Intronic	Benign	3/66	No	Yes
c.582G>A	rs80358810	p.Trp194X	Nonsense	Pathogenic	1/66	No	No
c.681+10T>G	novel	-	Intronic	Unlikely	1/66	No	No
c.1114A>C	rs144848	p.Asn372His	Missense	Unlikely	4/66	Yes	Yes
c.3264T>C	rs36060526	-	Synonymous	Benign	2/66	No	Yes
c.3396A>G	rs1801406	-	Synonymous	Benign	5/66	Yes	Yes
c.3807T>C	rs543304	-	Synonymous	Benign	15/66	Yes	Yes
c.4563A>G	rs206075	-	Synonymous	Benign	64/66	Yes	Yes
c.4681C>A	rs2219594	p.His1561Asn	Missense	Benign	2/66	Yes	Yes
c.5198C>T	rs55639415	p.Ser1733Phe	Missense	Unlikely	2/66	Yes	Yes
c.5414A>G	rs80358765	p.Asn1805Ser	Missense	Benign	5/66	No	Yes
c.5418A>G	rs34351119	-	Synonymous	Benign	2/66	Yes	Yes
c.6412G>T	rs11571659	p.Val2138Phe	Missense	Benign	8/66	No	Yes
c.6513G>C	rs206076	-	Synonymous	Benign	62/66	No	Yes
c.6842-73T>A	rs11571673	-	Intronic	Benign	8/66	No	Yes
c.6938-120T>C	rs206080	-	Intronic	Benign	62/66	No	Yes
c.7017G>C	rs45574331	p.Lys2339Asn	Missense	Unlikely	1/66	Yes	Yes
c.7242A>G	rs1799955	-	Synonymous	Benign	4/66	Yes	Yes
c.7397C>T	rs169547	p.Val2466Ala	Missense	Benign	65/66	Yes	Yes
c.7435+6G>A	rs81002852	-	Intronic	Benign	2/66	No	Yes
c.7626G>A	rs61754138	-	Synonymous	Benign	1/66	Yes	Yes
c.7712A>G	rs55689095	p.Glu2571Gly	Missense	Pathogenic	1/66	No	Yes
c.7805+6C>G	rs81002819	-	Intronic	Benign	3/66	No	Yes
c.7806-14T>G	rs9534262	-	Intronic	Benign	32/66	Yes	Yes
c.8092G>A	rs80359052	p.Ala2698Thr	Missense	Benign	3/66	No	Yes
c.8487+8G>A	rs81002838	-	Intronic	Benign	1/66	No	Yes
c.8487+19A>G	rs11571743	-	Intronic	Benign	1/66	Yes	Yes
c.8487+47C>T	rs11571744	-	Intronic	Benign	5/66	Yes	Yes
c.8632+131_8632+132insC	rs201392123	-	Intronic	Unlikely	1/66	No	No
c.8830A>T	rs4987047	p.Ile2944Phe	Missense	Unlikely	3/66	Yes	Yes
c.9235G>A	rs55933907	p.Val3079Ile	Missense	Benign	1/66	No	Yes
c.9648+54G>A	rs11571823	-	Intronic	Benign	7/66	No	Yes
c.9730G>A	rs11571831	p.Val3244Ile	Missense	Benign	2/66	Yes	Yes
c.10234A>G	rs1801426	p.Ile3412Val	Missense	Benign	6/66	Yes	Yes

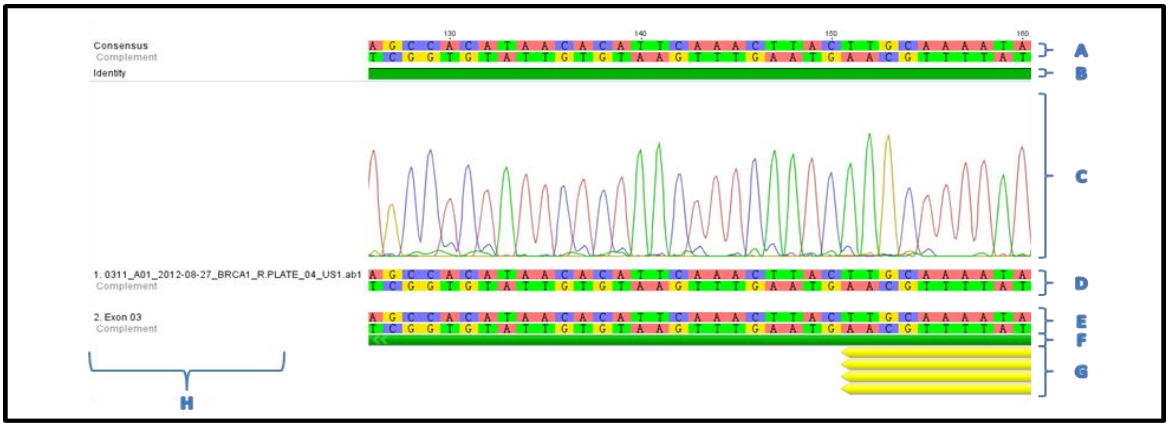
* Allele frequencies of the *BRCA2* SNVs found in the Control group are listed in Table 3.4. and the allele frequencies for the *BRCA2* SNVs found in the 1KG group are listed in Table 3.6.

3.1. Pathogenic or likely pathogenic mutations

A total of 3 pathogenic mutations were found in 3 out of the cohort of 33 patients (9.1% pick-up rate). One patient tested positive for a *BRCA 1* mutation, and two patients tested positive for *BRCA 2* mutations. All three mutations were unique to the patients and only occurred once.

The sequence alignments were performed using the Geneious software package (Biomatters, New Zealand). Figure 3.1. contains a detailed explanation of the graphical sequence alignment output. The sequence alignment tracks and annotations are explained in this figure. The graphical sequence alignment outputs are explained ahead of the results-figures to avoid overcrowding of the result-figures.

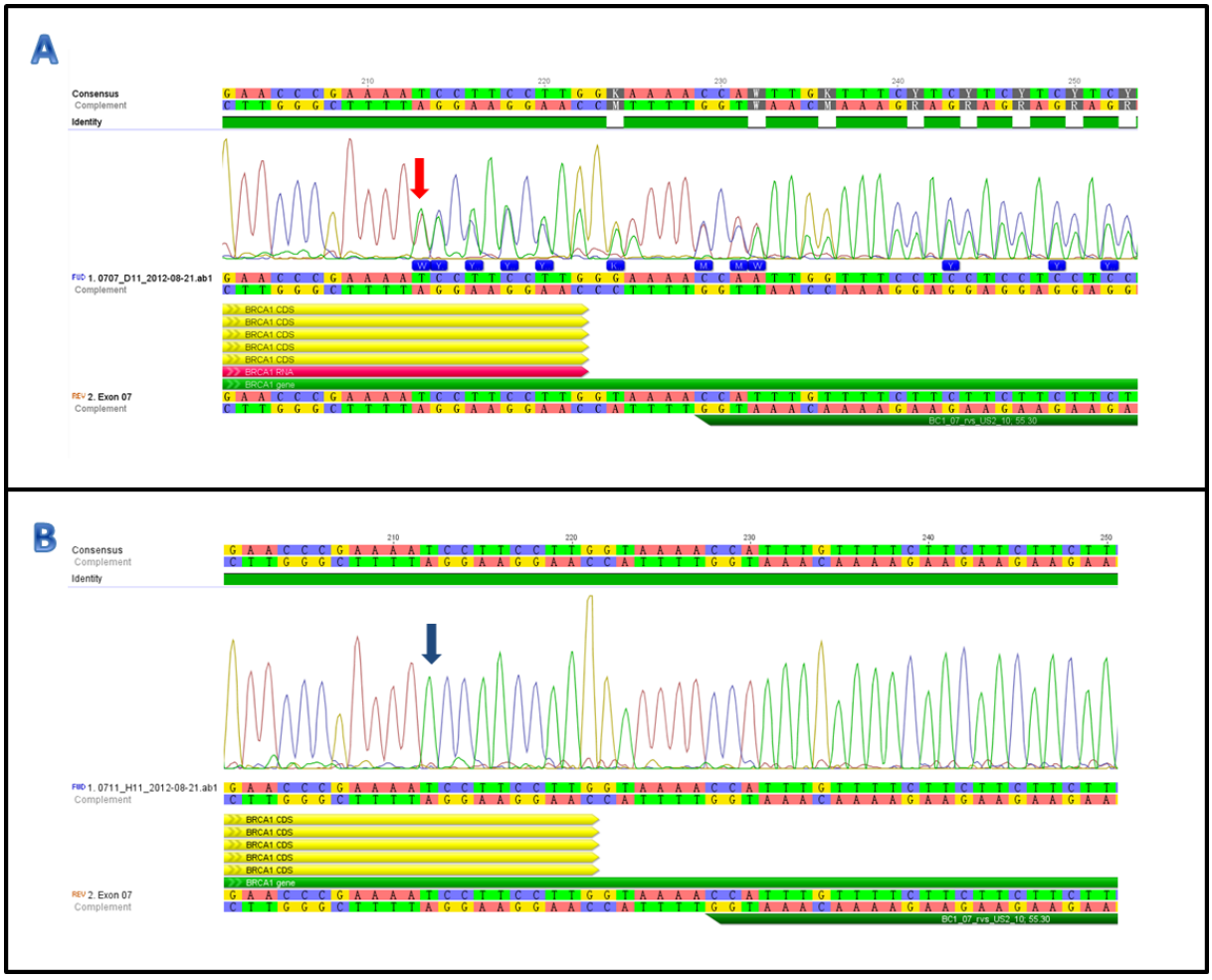
Figure 3.1. Geneious v.6.0.2. graphical sequence alignment output format. The Geneious software package exports the sequence alignment tracks and annotations graphically for ease of viewing. Each track is labelled (A-G) and is explained as follows: A – Consensus sequence track – this track shows sequence consensus between the sequence of interest and the reference sequence against which the sequence of interest was aligned. Ambiguous sequences are represented by the symbols "M,R,W,S,Y & K" depending on the nucleotide mismatch; B – Identity track – this track shows the percentage of identical sequence present at each nucleotide position. This is used in conjunction with track A to annotate ambiguous sequence. Since only two sequences (sequence of interest and reference sequence) were used in our alignment, this track will either show 100% identity or 0% identity. The green bar represents 100% identity, gaps represent mismatch sequence at a given position; C – Chromatogram – this track shows the chromatogram input of the sequence of interest. Physical peak heights and peak quality can be viewed here; D – Sequence of interest (patient sequence) – the nucleotide sequence and its reverse complement is displayed here. The nucleotide sequences were extracted from the chromatogram; E – Reference sequence – the reference sequence nucleotide and its reverse complement sequence is displayed here; F – Gene track – this annotation track displays the full gene region. The track annotations are features of Geneious that allow the user to lay annotation tracks over the reference sequence. This was used to identity different regions of the gene. This includes both intronic and exonic regions; G – CDS (Coding DNA sequence) track – this annotation track displays the physical location of the CDS regions of the gene of interest. Each yellow bar represents one known transcript of the gene. Regions not covered by the CDS tracks represent the intronic regions. The direction of the arrows indicates the orientation of the gene, with the 5' end flowing towards the 3' end; H – Name of sequence files – the file names of the sequences used are displayed here.



3.1.1. *BRCA1*: c.431dupA (p.Asn144Lysfs*15)

This mutation is novel. Patient BRCA007 tested positive for a c.431dupA mutation in the *BRCA 1* gene. The duplication of 1 bp is predicted to cause a frame shift mutation in exon 7 of the coding sequence of *BRCA 1*. This frame shift mutation creates a STOP codon 15 amino acids downstream of the mutation. This mutation results in the production of truncated mRNA transcripts. This truncated transcript may be subjected to nonsense mediated decay (NMD). Alternatively this mRNA transcript may produce truncated protein products which may be dysfunctional and may interfere with the BRCA protein complex. Figure 3.2. shows the c.431dupA mutation against the reference sequence for this region.

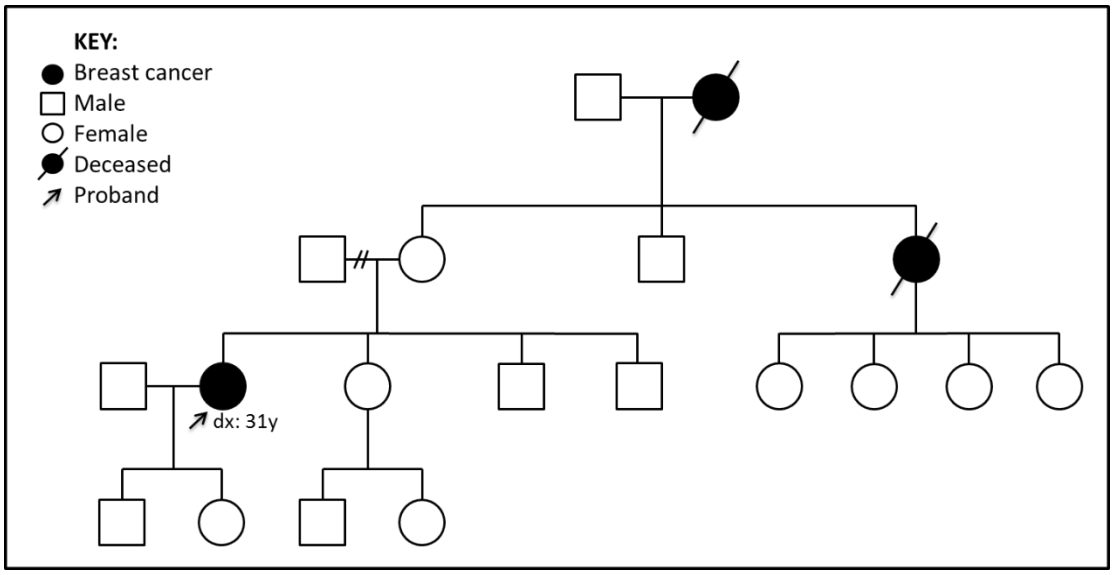
Figure 3.2. The c.431dupA mutation in *BRCA1*. A – This mutation was detected in patient BRCA007. The red arrow indicates the point of duplication of the A nucleotide. This mutation results in a frameshift, creating a stop codon 15 codons downstream. B – The wildtype sequence for the region is shown. The blue arrow indicates the wildtype A nucleotide position.



This is a novel mutation which has not been reported to date in any of the BRCA mutation databases listed in section 2.2.4.4.1.

Patient BRCA007 was diagnosed with Grade II invasive ductal carcinoma of the left breast at the age of 31 years. The patient's tumour tested negative for all three of the receptors: oestrogen receptors, progesterone receptors and human epidermal growth factor receptor 2 (HER2). The patient's maternal grandmother and maternal aunt were diagnosed with breast cancer around the age of 40. Figure 3.3. shows the family pedigree for patient BRCA007.

Figure 3.3. Family pedigree of patient BRCA007. The arrow indicates the proband (BRCA007).

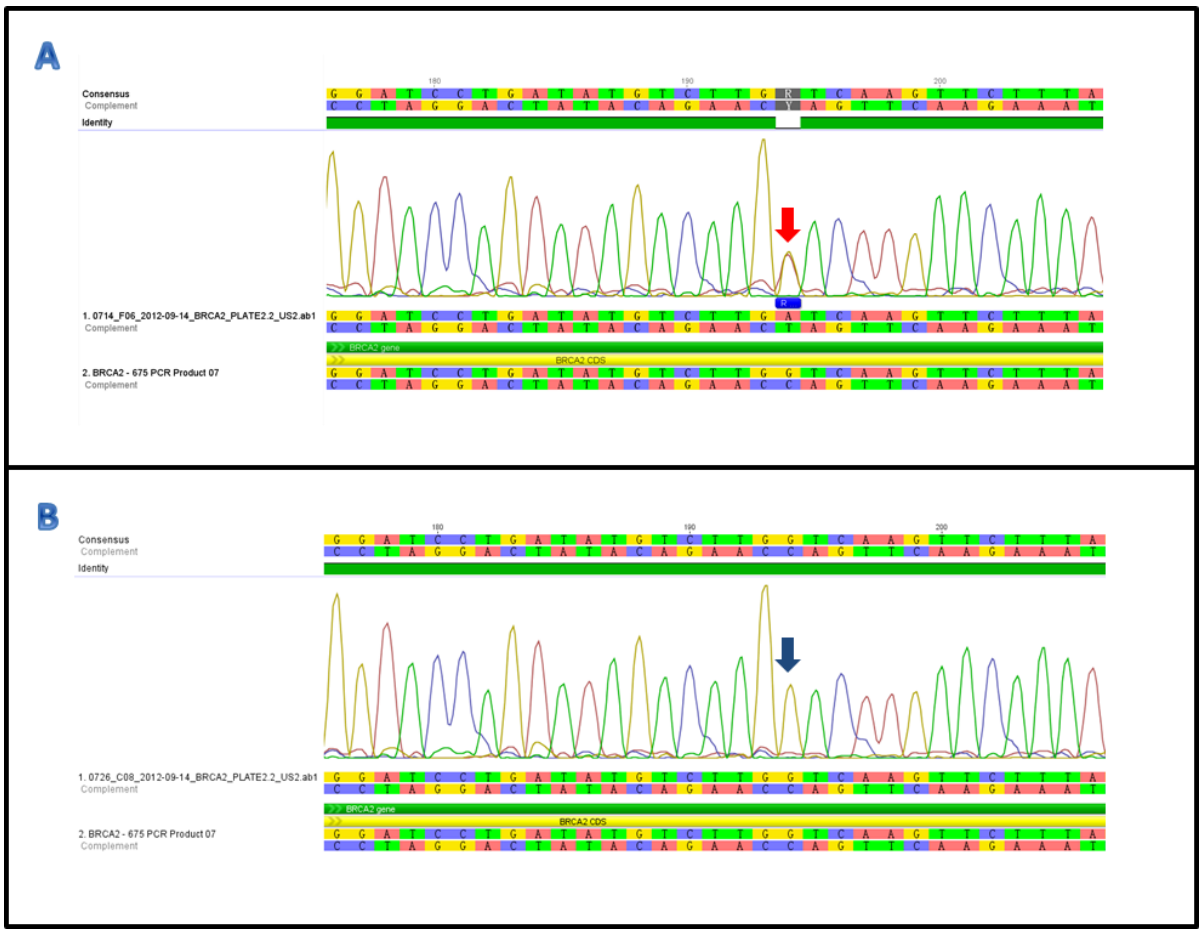


This result will be verified under diagnostic settings in the Molecular Diagnostic Laboratory, NHLS. Patient and at-risk relatives will be informed once this result have been verified. Predictive testing will be offered to at-risk relatives.

3.1.2. *BRCA2*: c.582G>A (p.Trp194*)

This mutation have been previously reported to be pathogenic. Patient BRCA014 tested positive for the c.582G>A mutation in the *BRCA 2* gene. This mutation is a nonsense mutation in exon 7 of the *BRCA 2* gene. The mRNA reading frame is interrupted by a premature STOP codon. This mutation generates truncated mRNA transcripts that may be targeted by NMD. Figure 3.4. shows this mutation aligned against the reference sequence.

Figure 3.4. The c.582G>A mutation in *BRCA2* in patient BRCA014. A – The red arrow indicates the position of the single nucleotide change. This mutation creates a premature stop codon. B – The wildtype sequence for the region is shown. The blue arrow indicates the location of the wildtype nucleotide.

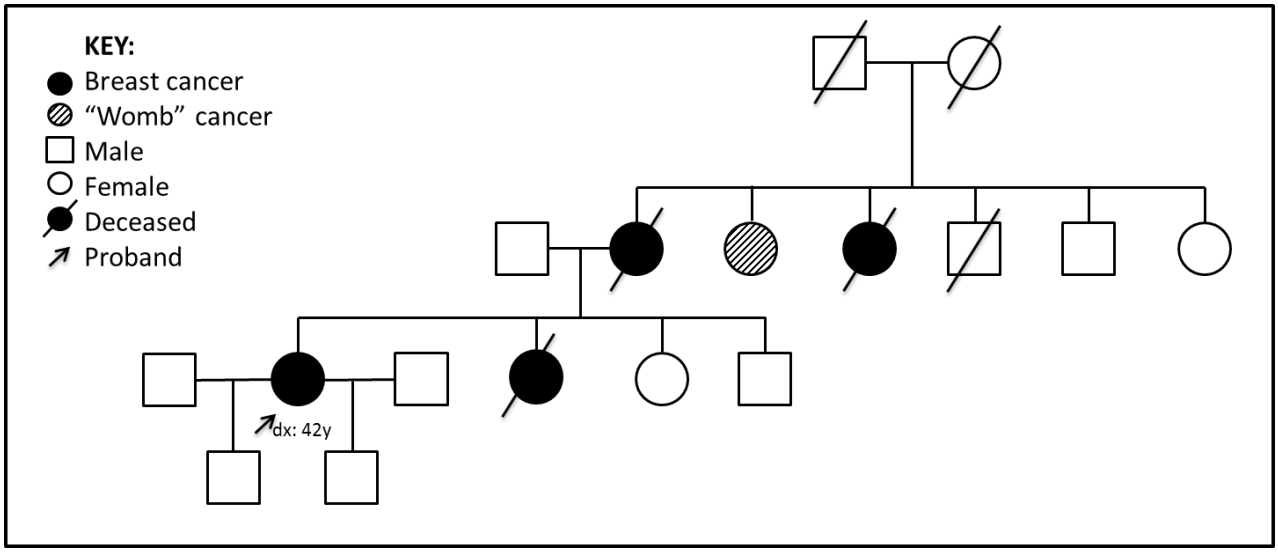


This mutation has been previously reported in dbSNP (rs80358810) but this variant has not been confirmed in larger studies such as HapMap or 1KG. This variant was not detected by the Omni 5 BeadChip (Illumina, USA).

Patient BRCA014 was diagnosed with invasive ductal carcinoma of the right breast, stage 2, triple negative, at the age of 42 years. Three other family members were affected with cancer in her family: her sister was diagnosed with unilateral breast cancer at 48 years of age and passed away at 51

years of age; her mother was diagnosed with unilateral breast cancer and passed away in her late 50's; her maternal aunt was diagnosed with breast cancer (laterality unknown) and passed away in her 50's. Another maternal aunt had "womb" cancer. Histological diagnosis of this cancer was not confirmed. Figure 3.5. presents the family pedigree for patient BRCA014.

Figure 3.5. Family pedigree for patient BRCA014. The arrow indicates the proband (BRCA014).

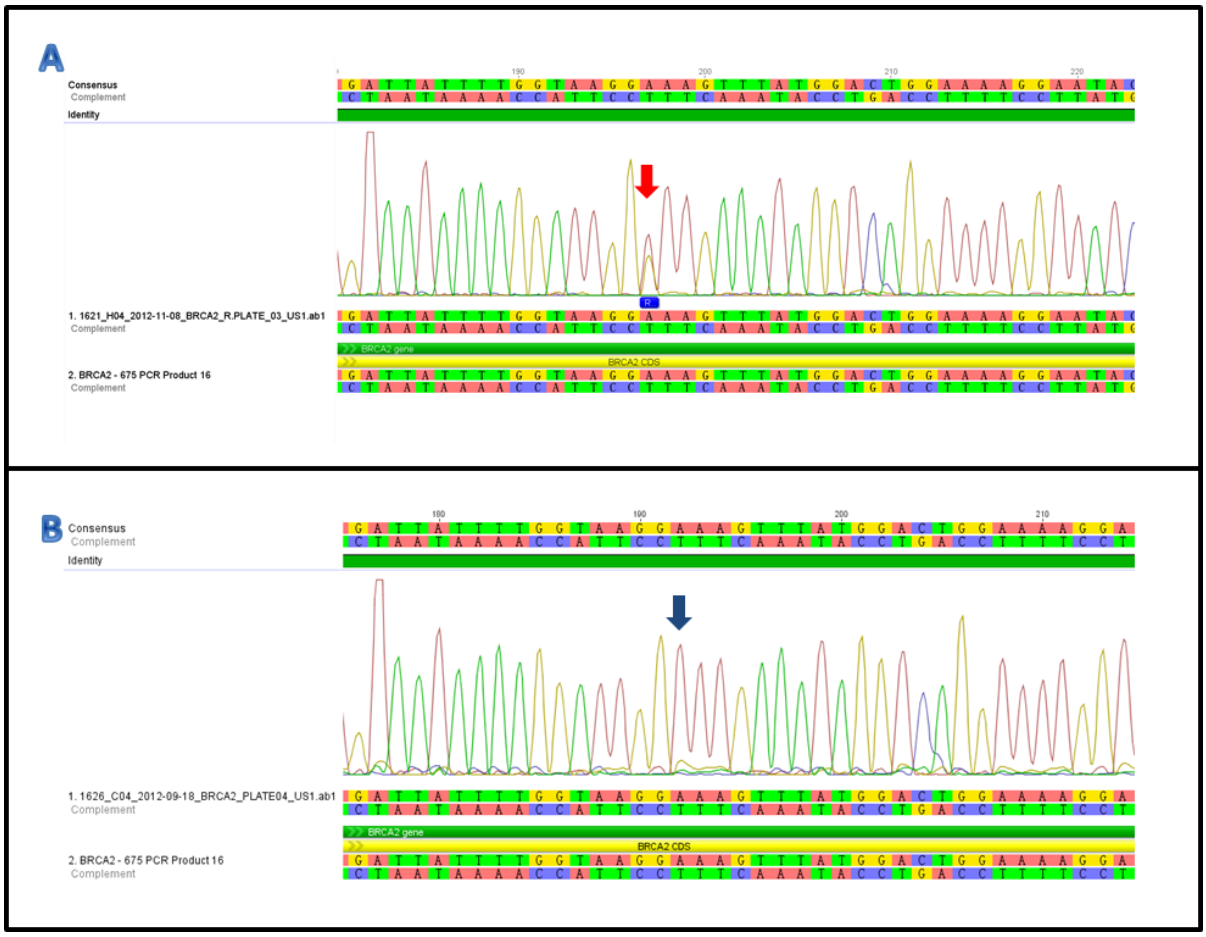


This result will be verified under diagnostic settings in the Molecular Diagnostic Laboratory, NHLS. Patient and at-risk relatives will be informed once this result have been verified. Predictive testing will be offered to at-risk relatives.

3.1.3. *BRCA2*: c.7712A>G (p.Glu2571Gly)

This substitution have been previously reported. Patient BRCA021 tested positive for the c.7712A>G substitution in exon 16 of the *BRCA2* gene. This is a missense mutation. The 2571th amino acid glutamic acid, is substituted with glycine. A substitution that results in a different missense change at this codon has been previously reported in a patient in a prostate cancer study (Kote-Jarai et al., 2011a). *In Silico* evidence provided by Kote-Jarai et al. (2011) suggested the change of the codon 2571 is potentially damaging. In their study, the substitution c.7711G>A was found, which caused the amino acid codon change of glutamic acid to lysine. Figure 3.6. shows this substitution against the reference sequence.

Figure 3.6. The c.7712A>G substitution in *BRCA2*. A – This substitution is a missense mutation resulting in the change of the amino acid codon from glutamic acid to glycine. Variation at this amino acid codon has been previously implicated in prostate cancer. The red arrow indicates the point of mutation. B – The wildtype sequence for the region is shown. The blue arrow indicates the location of the wildtype nucleotide.



This mutation is predicted as deleterious by 4/4 prediction algorithms used. SIFT: deleterious (score: 0.00); PolyPhen_2: possibly damaging (score: 0.874); Align GVGD*: deleterious (score: C65); MutationTaster: disease causing (p-value: 0.839). This nucleotide and amino acid are both highly

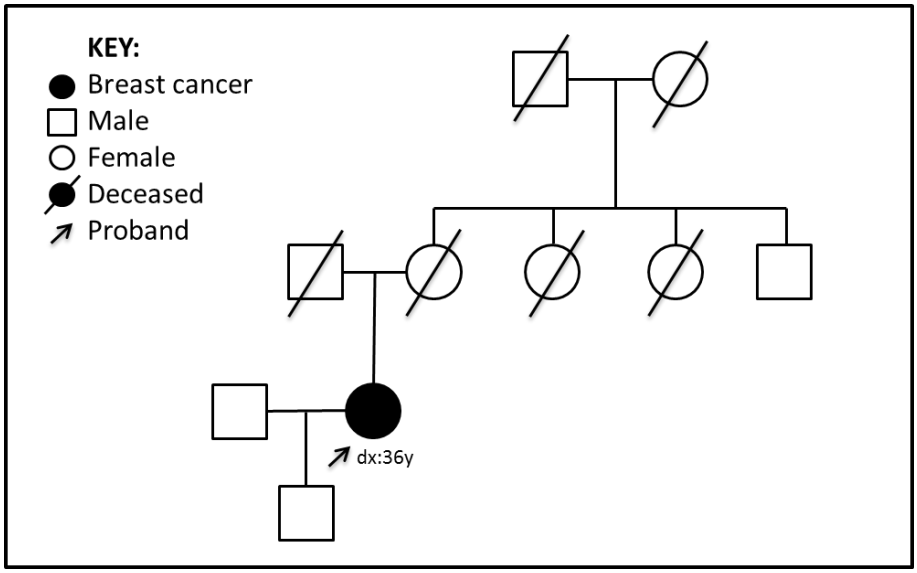
conserved. This mutation is located in the helical domain of the BRCA 2 protein. No predicted effects on splicing of the exon.

*It is noted that in the recent updated version of Alamut (April 2015), Align GVGD scored this substitution to be benign (score: C0).

This substitution has been reported in dbSNP, rs55689095, with MAF (minor allele frequency) of 0.002 for the G allele. This variant was not included in the Omni 5 BeadChip (Illumina, USA). This variant was detected in the 1KG project, with an MAF of 0.002 for the G allele in the African super-population (referred to as the 1KG group in this study).

This patient was diagnosed with breast cancer of the left breast at the age of 36 years. Receptor status for patient BRCA021 is unknown. No significant family history of cancers was reported. Figure 3.7. presents the family pedigree for BRCA021.

Figure 3.7. Family pedigree for patient BRCA021. The arrow indicates the proband (BRCA021).



This result will be verified under diagnostic settings in the Molecular Diagnostic Laboratory, NHLS. Patient and at-risk relatives will be informed once this result have been verified. Predictive testing will be offered to at-risk relatives.

3.2. Novel sequence variants

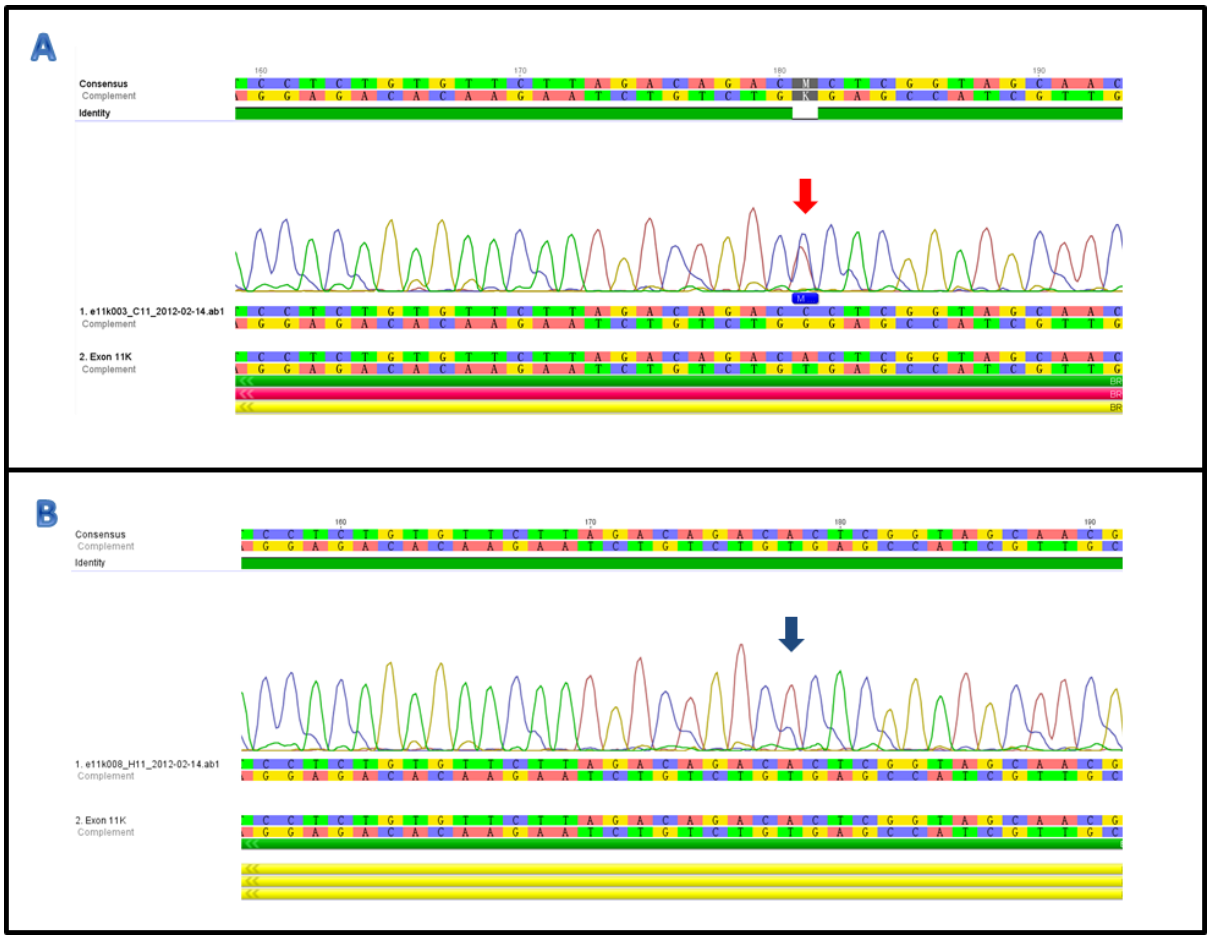
A total of five novel sequence variants were found in this study. Four novel sequence variants were found in *BRCA 1* and one novel sequence variant was found in *BRCA 2*. All novel sequence variants were unique to the individual patients and only occurred once. The novel sequence variants have not been previously reported in any studies.

These novel sequence variants were not screened for in the Control cohort due to financial constraints.

3.2.1. *BRCA1:* c.3751T>G (p.Cys1251Gly)

This variant is located in exon 11 of the *BRCA 1* gene. c.3751T>G results in a missense mutation. The amino acid cysteine at position 1251 is substituted with glycine. This variant was found in patient BRCA003. Figure 3.8. shows this variant against the reference sequence.

Figure 3.8. The c.3751T>G mutation in *BRCA1.* A – The red arrow indicates the point of mutation. B – The wildtype sequence for the region is shown. The blue arrow indicates the location of the wildtype nucleotide.



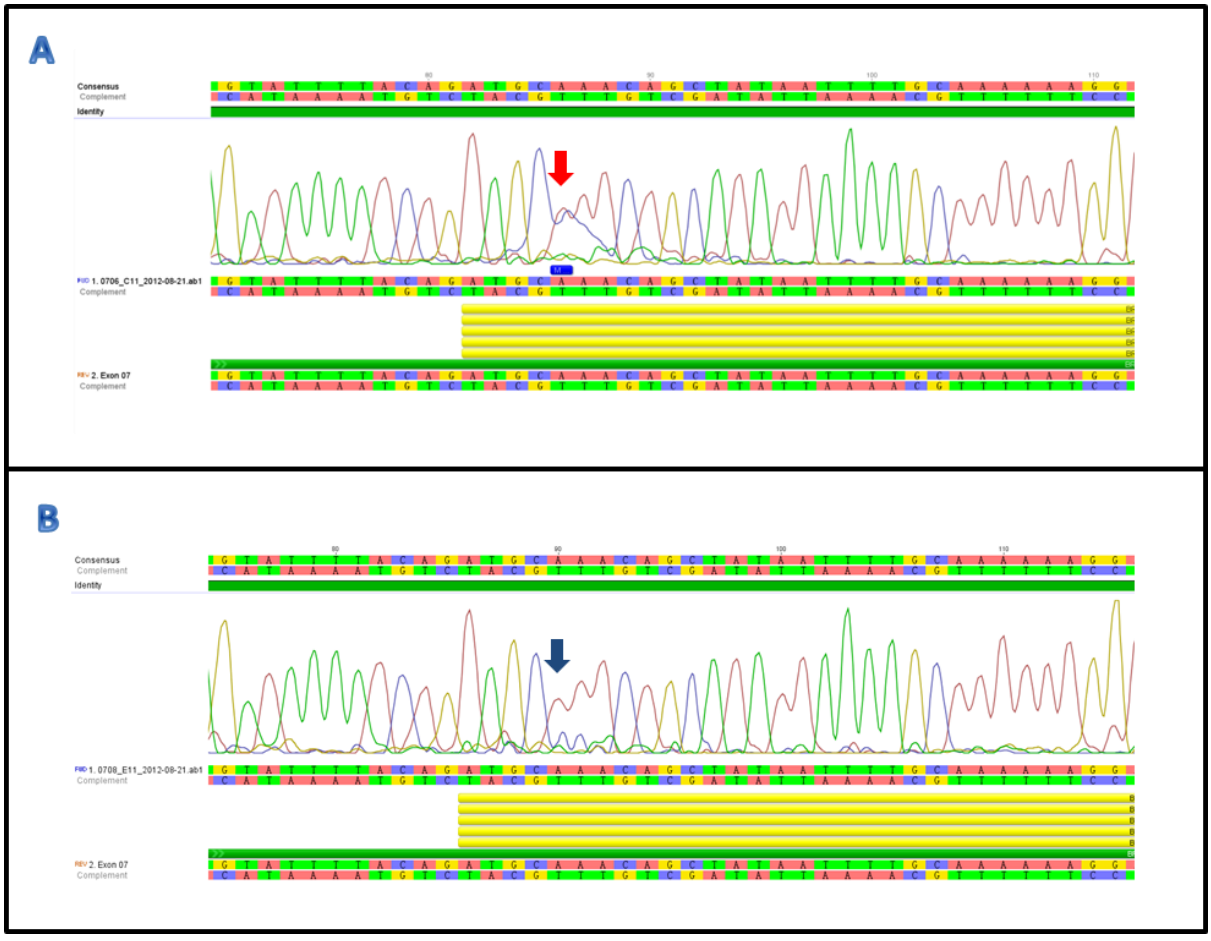
This variant is predicted to be benign by 3/4 prediction algorithms and predicted deleterious by 1/4 predication algorithm: SIFT: tolerated (score: 0.56); PolyPhen_2: benign (score: 0.427); AlignGVGD:

benign (score: C0); MutationTaster: disease causing (p-value: 0.773). This nucleotide is not conserved and the amino acid is weakly conserved. No splicing effects were predicted.

3.2.2. c.306A>C (p.=)

This variant is located in exon 7 of the *BRCA 1* gene. c.306A>C is a synonymous substitution, with no alteration of the amino acid sequence. Figure 3.9. shows this variant aligned against the reference sequence. No splicing effects were predicted.

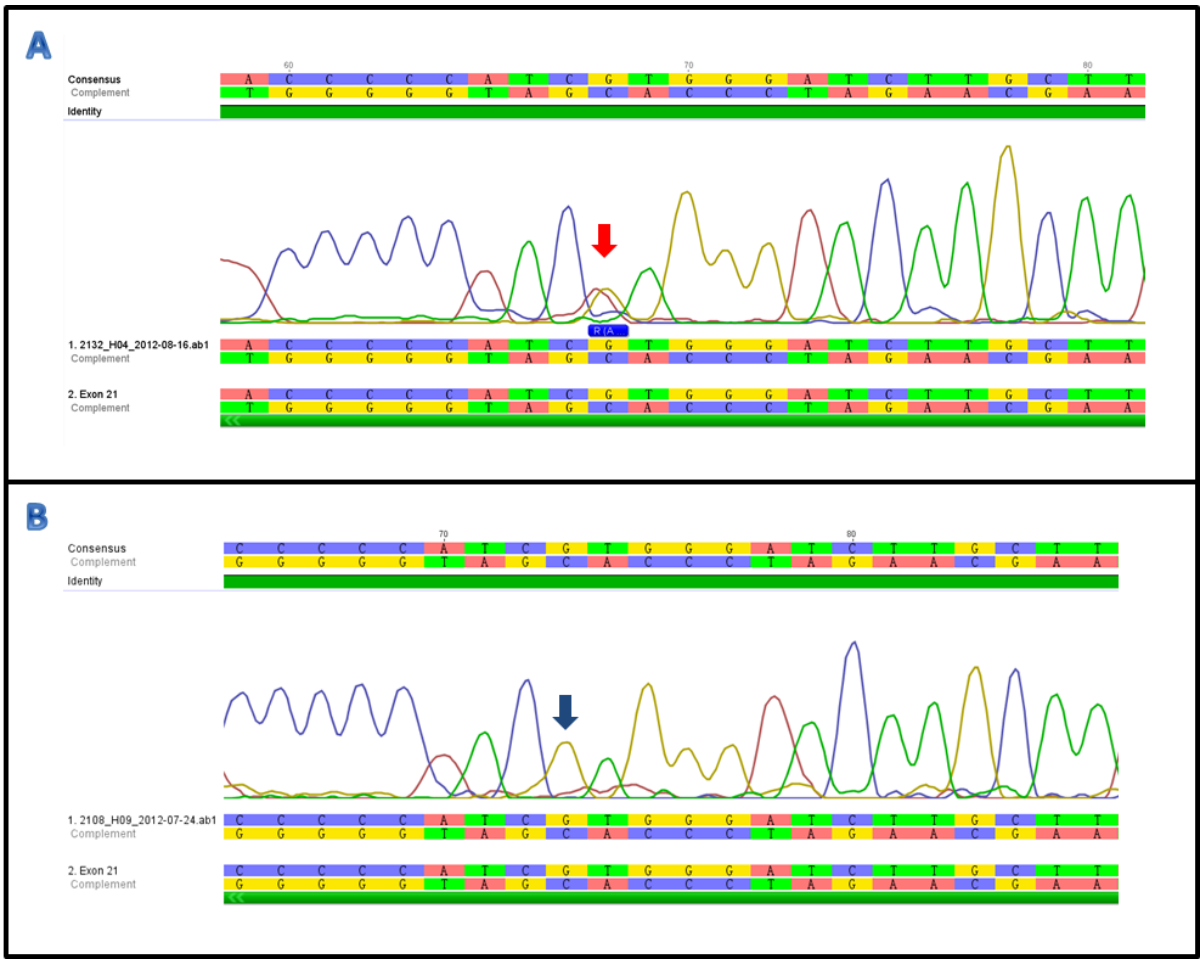
Figure 3.9. The c306A>C mutation in *BRCA1.* A – The red arrow indicates the point of mutation. B – The wildtype sequence for the region is shown. The blue arrow indicates the location of the wildtype nucleotide. The electropherogram represents the reverse strand.



3.2.3. *BRCA1:* c.5332+78C>T

This variant is located in intron 21 of the *BRCA 1* gene. No obvious alterations to the splice site junction were predicted by *in Silico* analysis using the Alamut software package. Figure 3.10. shows this variant aligned against the reference sequence. No splicing effect predicted.

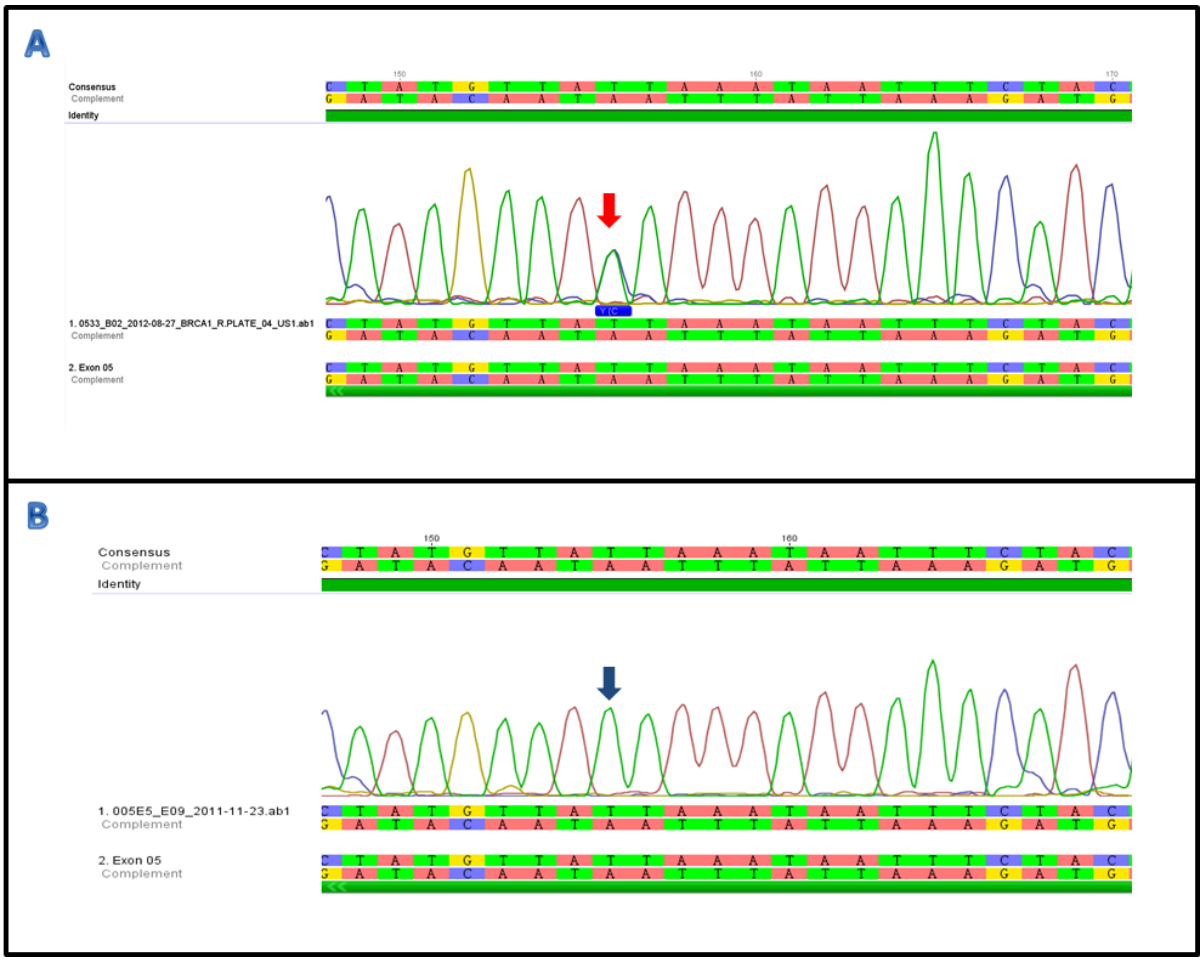
Figure 3.10. The c.5332+78C>T mutation in *BRCA1.* A – The red arrow indicates the point of mutation. B – The wildtype sequence for the region is shown. The blue arrow indicates the location of the wildtype nucleotide.



3.2.4. *BRCA1:* c.212+66A>G

This variant is located in intron 5 of the *BRCA 1* gene. No obvious alterations to the splice site junction were predicted by the *in Silico* analysis using the Alamut software package, no novel splice site were created. Figure 3.11. shows this variant aligned against the reference sequence.

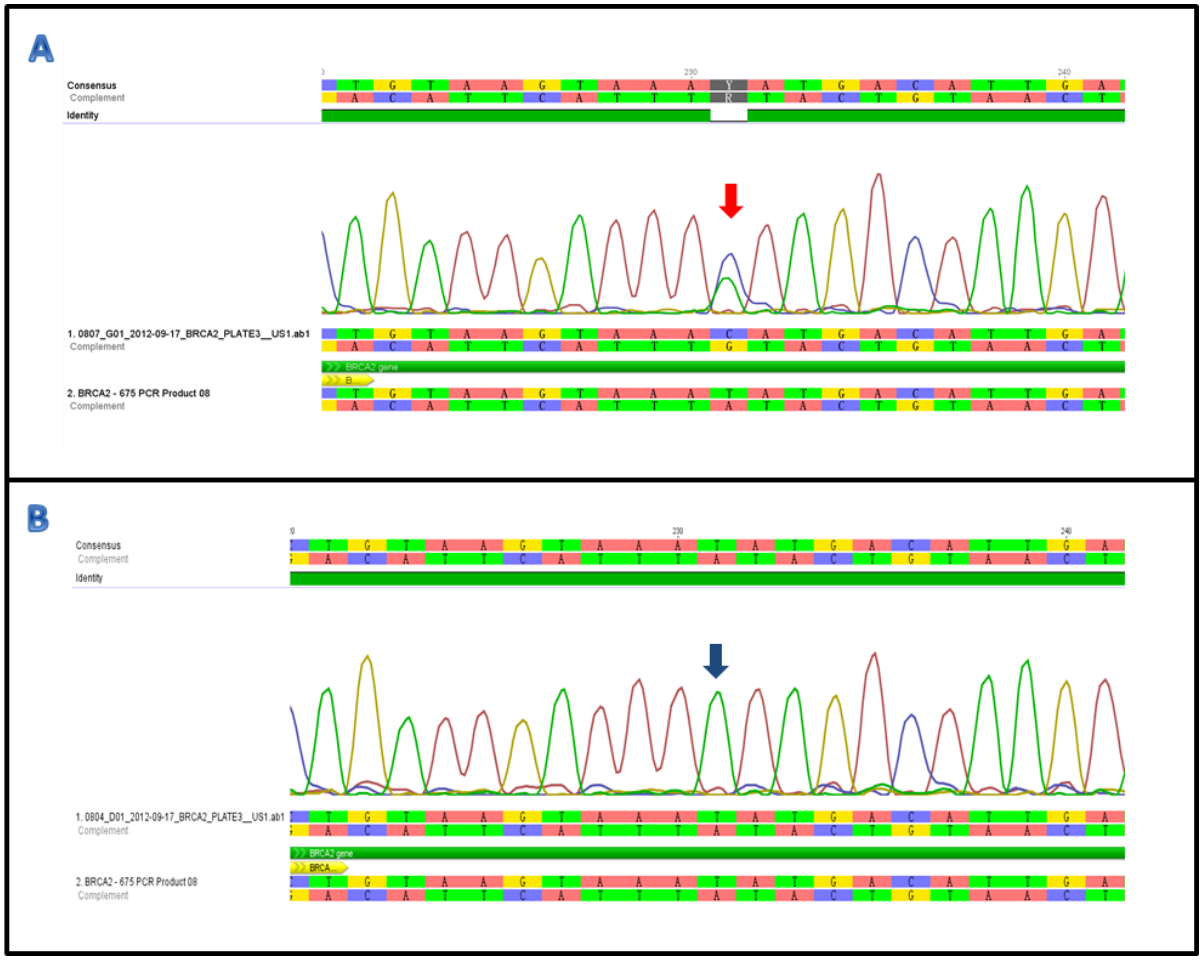
Figure 3.11. The c.212+66A>G mutation in *BRCA1.* A – The red arrow indicates the point of mutation. B – The wildtype sequence for the region is shown. The blue arrow indicates the location of the wildtype nucleotide.



3.2.5. *BRCA2:* c.681+10T>G

This variant is located in intron 8 of the *BRCA 2* gene. No obvious alterations to the splice site junction were predicted by the *in Silico* analysis using the Alamut software package, no novel splice sites were created. Figure 3.12. shows this variant aligned against the reference sequence.

Figure 3.12. The c.681+10T>G mutation in *BRCA2.* A – The red arrow indicates the point of mutation. B – The wildtype sequence for the region is shown. The blue arrow indicates the location of the wildtype nucleotide. The gap in sequence identity shows where the nucleotide does not match the reference sequence, as the signal for the variant is stronger than the wildtype allele.



3.3. Known SNVs

A total of 51 distinct, previously reported, SNVs were found in this study.

Of the 17 SNVs found in *BRCA1*, nine were missense variants, five were synonymous and three were intronic (Table 3.1.).

Of the 34 SNVs found in *BRCA2*, twelve were missense variants, eight were synonymous and fourteen were intronic (Table 3.2.).

The allelic frequencies of the SNVs were compared with the Control group data generated using the Omni 5 chip (May et al., 2013) and the 1KG group data (1000 Genomes Project Consortium, 2012) and statistically evaluated to determine if the SNVs conferred any risks or protective effects associated with breast cancer.

All alleles were checked for Hardy-Weinberg equilibrium (HWE). All genotypes were in HWE for the BRCA cohort, the Control group and the 1KG group.

In the tables below, the alleles that either confer an altered risk profile or if the alleles were statistically different in allele frequency are highlighted in green. The pathogenic allele that is present in either of the two control data sets is highlighted in yellow. This is done to differentiate the pathogenic alleles to non-pathogenic alleles. The statistically significant alleles and risk conferring alleles are discussed in detail in the Discussion chapter.

3.3.1. *BRCA1* data versus Control group data

Out of the 17 SNVs found in the *BRCA1* gene, only 12 SNVs were present on the Omni 5 chip (Table 3.1.). Chi-squared test, V-squared test and Yates corrected Chi-squared test were performed to evaluate if differences in allelic frequencies existed between the *BRCA* cohort and the Omni 5 control group (Table 3.4.). The allele frequencies were tested on the basis of the null hypothesis that the minor allele observed in the BRCA cohort does not confer a risk for the development of breast cancer. Alleles with p values below 0.01 were deemed significant and thus may confer an increase in risk for the development of breast cancer and led to the rejection of the null hypothesis.

V-squared (corrected Chi-square or Cramér's V) and Yates corrected Chi-square test are variations of the standard Chi-square test that takes into consideration the small allele count and corrects for the small number of observed allelic counts as certain SNVs were only observed once amongst the entire group. The Yates correction may adjust too far and lead to bias in the statistical outputs (Yates, 1934). The V-squared Chi-square test was used in combination with Yates corrected Chi-square test

to reduce the impact of over statistical corrections. The total allelic count for the BRCA cohort was 66 for any given SNV and the total allelic count for the Control group is 188 alleles for any given SNV.

Odds ratios with 95% confidence intervals and genotype relative risks were calculated. Risks conferred by genotypes, either homozygous for the minor allele, heterozygous, or homozygous for the major allele are tabulated (Table 3.3.).

For *BRCA1* one allele frequency was significantly different in the BRCA cohort and the Control group at $p<0.001$ level. Homozygosity for the major allele of rs28897679 (T nucleotide) confers a 1.36 fold relative risk increase in the development for breast cancer (Table 3.3.). Homozygosity for the minor allele and heterozygosity conferred no risk.

No other SNVs showed significant difference in allele frequencies between the BRCA cohort and the Control group.

3.3.2. *BRCA2* data versus Control group data

Out of the 37 SNVs found in the *BRCA2* gene, only 20 SNVs were present on the Omni 5 chip (Table 3.4.). Alleles were statistically evaluated using the same parameters are the *BRCA1* alleles, and tested under the same null hypothesis.

The rs169547 allele frequency was significantly different in the BRCA cohort and the Control group, at $p<0.001$ levels. Homozygosity for the minor allele of rs169547 (T nucleotide) conferred approximately a 500 fold increase in the relative risk of developing breast cancer when compared to the average population risk. Although it is important to note that this may be due to the incorrect reference sequence error, whereby the minor allele was called as the reference allele. It is important to note there this observation may likely be a genotyping error from the Omni 5 data.

Homozygosity for the minor allele of rs2219594 (A nucleotide) conferred a relative risk of 2, but the allele frequency did not differ significantly between the BRCA cohort and the Control group. The relative risk is defined as the risk, in addition to the general population risk, for the development of disease.

Due to the small sample size of the BRCA cohort and the Control cohort, there is no power in this statistically evaluation. The possibility of over analysis may also exist.

Table 3.3. Statistical comparison of BRCA1 SNV allelic frequencies between the BRCA cohort and the Control group.

Allele Reference		BRCA Data	Omni5 Data	Chi-square test (p<0.01)					Odds Ratio (95% C.I.)			Relative Risk (95% C.I.)			Relative Risk to Population Average Risk		
rsID	Minor Allele	MAF	MAF	Chi-square (df=1)	V-squared (df=1)	Yates (df=1)	p<0.01	p<0.001	Odds Ratio	Lower Limt OR (95% C.I.)	Upper Limit OR (95% C.I.)	Relative Risk	Lower Limt RR (95% C.I.)	Upper Limit RR (95% C.I.)	RR(AA)*	RR(Aa)**	RR(aa)***
rs16941	G	0.045455	0.069150	0.4955	0.4963	0.6986			0.6410	0.1768	2.3363	0.4522	0.1596	1.2816	0.2508	0.3913	0.6104
rs16942	G	0.060606	0.117000	0.1933	0.1942	0.2870			0.4869	0.1613	1.4761	0.4258	0.1686	1.0750	0.1606	0.3298	0.6773
rs1799966	G	0.060606	0.117000	0.1933	0.1942	0.2870			0.4869	0.1613	1.4761	0.4258	0.1686	1.0750	0.1606	0.3298	0.6773
rs2227945	G	0.060606	0.074470	0.7057	0.7063	0.9213			0.8018	0.2543	2.5401	0.4849	0.1991	1.1812	0.3572	0.4455	0.5556
rs799917	C	0.060606	0.074470	0.7057	0.7063	0.9213			0.8018	0.2543	2.5401	0.4849	0.1991	1.1812	0.3572	0.4455	0.5556
rs1060915	C	0.030303	0.069150	0.2494	0.2504	0.3963			0.4207	0.0924	1.9276	0.4011	0.1085	1.4824	0.1249	0.2969	0.7059
rs16940	C	0.030303	0.069150	0.2494	0.2504	0.3963			0.4207	0.0924	1.9276	0.4011	0.1085	1.4824	0.1249	0.2969	0.7059
rs1799949	T	0.015152	0.117000	0.0131	0.0133	0.0256			0.1161	0.0153	0.8865	0.3249	0.0473	2.2336	0.0122	0.1053	0.9069
rs28897679	G	0.015152	0.531900	0.0000	0.0000	0.0000	Significant	Significant	0.0135	0.0018	0.1004	0.4347	0.0613	3.0831	0.0003	0.0185	1.3686
rs8176316	A	0.015152	0.026880	0.5913	0.5920	0.9465			0.5570	0.0639	4.9002	0.4309	0.0711	2.6103	0.1993	0.3578	0.6425
rs80358035	G	0.015152	0.042550	0.3002	0.3012	0.5163			0.3462	0.0425	2.8460	0.3764	0.0586	2.4171	0.0891	0.2575	0.7437
rs55946644	C	0.015152	0.021980	0.7353	0.7359	0.8625			0.6846	0.0751	6.2946	0.4675	0.0800	2.7319	0.2782	0.4064	0.5937

*Relative risk to population average risk for homozygotes of the minor allele

**Relative risk to population average risk for heterozygotes

***Relative risk to population average risk for homozygote of the major allele

Table 3.4. Statistical comparison of BRCA2 SNV allelic frequencies between the BRCA cohort and the Control group.

Allele Reference		BRCA Data	Omni5 Data	Chi-square test (p<0.01)					Odds Ratio (95% C.I.)			Relative Risk (95% C.I.)			Relative Risk to Population Average Risk		
rsID	Minor Allele	MAF	MAF	Chi-square (df=1)	V-squared (df=1)	Yates (df=1)	p<0.01	p<0.001	Odds Ratio	Lower Limt OR (95% C.I.)	Upper Limit OR (95% C.I.)	Relative Risk	Lower Limt RR (95% C.I.)	Upper Limit RR (95% C.I.)	RR(AA)*	RR(Aa)**	RR(aa)***
rs11571831	A	0.030303	0.085110	0.1355	0.1363	0.2248			0.3359	0.0751	1.5113	0.3823	0.1018	1.4358	0.0849	0.2527	0.7522
rs144848	C	0.060606	0.058510	0.9505	0.9506	0.8093			1.0381	0.3189	3.3961	0.5261	0.2213	1.2509	0.5287	0.5093	0.4906
rs169547	T	0.984848	0.037230	0.0000	0.0000	0.0000	Significant	Significant	1680.9038	202.8900	14046.6667	0.9083	0.1284	6.4229	504.7078	0.3003	0.0002
rs1801426	G	0.090909	0.117000	0.5601	0.5608	0.7231			0.7547	0.2919	1.9586	0.4798	0.2285	1.0073	0.3257	0.4316	0.5718
rs2219594	A	0.030303	0.010640	0.2696	0.2706	0.5966			2.9058	0.4010	21.2250	0.7560	0.2774	2.0602	2.1613	0.7438	0.2560
rs45574331	C	0.015152	0.085110	0.0504	0.0509	0.0949			0.1654	0.0215	1.2830	0.3331	0.0492	2.2555	0.0236	0.1428	0.8633
rs4987047	T	0.045455	0.048390	0.9234	0.9236	0.8101			0.9364	0.2457	3.5885	0.5125	0.1880	1.3968	0.4529	0.4837	0.5165
rs55639415	T	0.030303	0.015960	0.4704	0.4713	0.8362			1.9268	0.3148	11.8799	0.6570	0.2200	1.9622	1.2681	0.6582	0.3416
rs1799955	G	0.060606	0.085110	0.5249	0.5257	0.7112			0.6935	0.2232	2.1644	0.4650	0.1886	1.1461	0.2846	0.4104	0.5918
rs1801406	G	0.075758	0.085110	0.8125	0.8128	0.9820			0.8811	0.3096	2.5181	0.4999	0.2258	1.1069	0.4131	0.4688	0.5321
rs206075	A	0.030303	0.053190	0.4508	0.4517	0.6768			0.5563	0.1187	2.6239	0.4311	0.1196	1.5545	0.1991	0.3579	0.6434
rs34351119	G	0.030303	0.09043	0.1102	0.1109	0.1850			0.3143	0.0706	1.4075	0.3776	0.1001	1.4242	0.0756	0.2405	0.7651
rs543304	C	0.227273	0.2128	0.8056	0.8059	0.9422			1.0880	0.5549	2.1393	0.5290	0.3233	0.8654	0.5647	0.5190	0.4770
rs61754138	A	0.015152	0.005319	0.4369	0.4378	0.9746			2.8770	0.1774	47.1970	0.7579	0.1866	3.0787	2.1348	0.7420	0.2579
rs11571623	T	0.015152	0.0266	0.5984	0.5991	0.9556			0.5630	0.0646	4.9528	0.4287	0.0708	2.5974	0.2028	0.3603	0.6400
rs11571743	G	0.015152	0.06989	0.0953	0.0960	0.1754			0.2047	0.0263	1.6100	0.3445	0.0515	2.3036	0.0349	0.1706	0.8333
rs11571744	T	0.075758	0.08065	0.8996	0.8998	0.8896			0.9344	0.3258	2.6910	0.5129	0.2330	1.1291	0.4515	0.4832	0.5172
rs1799943	A	0.045455	0.03191	0.6087	0.6094	0.9006			1.4447	0.3509	5.9830	0.5905	0.2288	1.5240	0.8533	0.5907	0.4089
rs76874770	T	0.045455	0.03723	0.7677	0.7681	0.9423			1.2314	0.3090	4.9354	0.5582	0.2115	1.4731	0.6794	0.5517	0.4480
rs9534262	T	0.484848	0.4628	0.7571	0.7576	0.8682			1.2333	0.7035	2.1672	0.5216	0.3444	0.7901	0.6487	0.5260	0.4265

*Relative risk to population average risk for homozygotes of the minor allele

**Relative risk to population average risk for heterozygotes

***Relative risk to population average risk for homozygotes of the major allele

3.3.3. *BRCA1* data versus 1KG group data

The SNVs found in the BRCA cohort were also tested against the data from the 1000 Genomes Project (1000 Genomes Project Consortium, 2012). The BRCA cohort data was tested against the African super population data set from the 1KG project. The African samples within the 1KG project comprised samples from the Yoruba in Ibadan, Nigera; Luhya in Webuye, Kenya; Gambian in Western Divisions in The Gambia; Mende in Sierra Leone; Esan in Nigera; Americans of African Ancestry in SW USA; and African Carribeans in Barbados. This group is comprised of 246 individuals. This data set forms the 1KG group data set. Refer to section 2.2.2.

Amongst the 17 SNVs found in the *BRCA1* gene of the BRCA cohort, 13 were present in the 1KG group data (Table 3.5.). Statistical analysis was performed using the same parameters as mentioned above (Section 3.3.1.). Two allele frequencies differed significantly at the $p<0.01$ level (rs16942 and rs1799966), rs1799949 also differed significantly at the $p<0.001$ level.

Two different risk conferring alleles were also found: homozygosity for the minor allele (A nucleotide) of rs55815649 confers a 6.6 fold increased risk for the development of breast cancer; homozygosity for the minor allele (G nucleotide) of rs80358035 confers a 1.22 fold increase in risk for the development of breast cancer. However the observed frequency of these two alleles was not statistically different between the two cohorts (Table 3.5.), thus suggesting the increased risk may not be a true risk association.

3.3.4. *BRCA2* data versus 1KG group data

Amongst the 37 SNVs found in the *BRCA2* gene of the BRCA cohort, 34 were present in the 1KG group data (Table 3.6.). Nine allele frequencies differed significantly at the $p<0.01$ level of which six also frequency differed significantly at the $p<0.001$ level (rs11571659; rs169547; rs80358765; rs206076; rs11571673; and rs11571823).

Sixteen risk conferring alleles were present. Amongst them are the nine statistically significant alleles mentioned above, six statistically non-significant allele frequency differences and the pathogenic allele rs55689095 (Table 3.6.). The conferred risk ranged from 1.49 fold increase to 92.2 fold increased risk for the development of breast cancer.

3.3.5. Control group data versus 1KG group data (*BRCA1*)

In addition to the statistical comparisons done for the BRCA cohort against the Control group (Section 3.3.1. & Section 3.3.2.) and the BRCA cohort against the 1KG group data (Section 3.3.3. & Section 3.3.4.), the comparable allele frequencies from the Control group were statistically

evaluated against the allele frequencies from the 1KG group to evaluate whether the observed allele frequencies were different across different African populations.

Twelve alleles within the *BRCA1* gene were compared and statistically evaluated. Amongst the twelve alleles, six alleles were statistically different at the $p<0.01$ level of which two alleles were also statistically different at the $p<0.001$ level (Table 3.7.).

In addition, three alleles appear to confer risk for the development of breast cancer, however the statistical significance needs to be evaluated in a larger sample size. It is still interesting to note that of the three alleles, rs28897679 (C nucleotide) confers an increase of 58.8 fold risk and this allele is significantly different between the Control group and the 1KG group.

Table 3.5. Statistical comparison of *BRCA1* SNV allelic frequencies between the BRCA cohort and the 1KG group.

Allele Reference		BRCA Data	1KG Data	Chi-square test					Odds Ratio (95% C.I.)			Relative Risk (95% C.I.)			Relative Risk to Population Average Risk		
rsID	Minor Allele	MAF	MAF	Chi-square (df=1)	V-squared (df=1)	Yates (df=1)	p<0.01	p<0.001	Odds Ratio	Lower Limt OR (95% C.I.)	Upper Limit OR (95% C.I.)	Relative Risk	Lower Limt RR (95% C.I.)	Upper Limit RR (95% C.I.)	RR(AA)*	RR(Aa)**	RR(aa)***
rs16941	G	0.045455	0.128	0.0511	0.0513	0.0805			0.3243	0.0988	1.0690	0.1735	0.0561	0.5368	0.0803	0.2476	0.7636
rs16942	G	0.060606	0.215	0.0030	0.0030	0.0050	Significant		0.2349	0.0836	0.6633	0.1748	0.0650	0.4700	0.0463	0.1972	0.8396
rs1799966	G	0.060606	0.220	0.0025	0.0025	0.0042	Significant		0.2294	0.0816	0.6475	0.1747	0.0649	0.4701	0.0445	0.1938	0.8448
rs2227945	G	0.060606	0.049	0.6795	0.6797	0.9100			1.2581	0.4225	3.7629	0.2598	0.1018	0.6632	0.7005	0.5568	0.4426
rs55815649	A	0.015152	0.002	0.0940	0.0943	0.5634			7.5538	0.4668	123.6339	0.6169	0.1514	2.5133	6.6706	0.8831	0.1169
rs799917	C	0.060606	0.136	0.0836	0.0839	0.1252			0.4092	0.1442	1.1668	0.1836	0.0689	0.4893	0.1202	0.2936	0.7175
rs1060915	C	0.030303	0.128	0.0201	0.0202	0.0340			0.2128	0.0508	0.8963	0.1606	0.0403	0.6405	0.0378	0.1778	0.8353
rs16940	C	0.030303	0.128	0.0201	0.0202	0.0340			0.2128	0.0508	0.8963	0.1606	0.0403	0.6405	0.0378	0.1778	0.8353
rs1799949	T	0.015152	0.215	0.0001	0.0001	0.0002	Significant	Significant	0.0560	0.0077	0.4118	0.1535	0.0215	1.0936	0.0031	0.0555	0.9903
rs28897679	G	0.015152	0.018	0.8567	0.8568	0.7540			0.8256	0.1029	6.6798	0.2186	0.0336	1.4232	0.3734	0.4523	0.5478
rs8176316	A	0.015152	0.016	0.9465	0.9465	0.6540			0.9308	0.1146	7.6275	0.2295	0.0357	1.4772	0.4487	0.4821	0.5179
rs80358035	G	0.015152	0.008	0.5698	0.5701	0.8988			1.8769	0.2066	17.2053	0.3175	0.0542	1.8603	1.2244	0.6524	0.3476
rs55946644	C	0.015152	0.020	0.7765	0.7767	0.8512			0.7415	0.0934	5.9378	0.2097	0.0319	1.3782	0.3158	0.4258	0.5743

*Relative risk to population average risk for homozygotes of the minor allele

**Relative risk to population average risk for heterozygotes

***Relative risk to population average risk for homozygotes of the major allele

Table 3.6. Statistical comparison of *BRCA2* SNV allelic frequencies between the BRCA cohort and the 1KG group.

Allele Reference		BRCA Data	1KG Data	Chi-square test					Odds Ratio (95% C.I.)			Relative Risk (95% C.I.)			Relative Risk to Population Average Risk		
rsID	Minor Allele	MAF	MAF	Chi-square (df=1)	V-squared (df=1)	Yates (df=1)	p<0.01	p<0.001	Odds Ratio	Lower Limt OR (95% C.I.)	Upper Limit OR (95% C.I.)	Relative Risk	Lower Limt RR (95% C.I.)	Upper Limit RR (95% C.I.)	RR(AA)*	RR(Aa)**	RR(aa)***
rs11571659	T	0.121212	0.006	0.0000	0.0000	0.0000	Significant	Significant	22.4828	5.8019	87.6047	0.8333	0.5388	1.2888	21.5087	0.9567	0.0426
rs11571831	A	0.030303	0.026	0.8548	0.8549	0.8241			1.1514	0.2540	5.2519	0.2512	0.0677	0.9315	0.6162	0.5351	0.4648
rs144848	C	0.060606	0.098	0..3322	0.3326	0.4567			0.5968	0.2080	1.7197	0.1995	0.0756	0.5261	0.2239	0.3752	0.6287
rs169547	T	0.015152	0.096	0.0000	0.0000	0.0000	Significant	Significant	615.4255	83.4750	4574.4167	0.5826	0.0817	4.1532	92.2211	0.1498	0.0002
rs1801426	G	0.090909	0.126	0.4130	0.4134	0.5364			0.6935	0.2875	1.6788	0.2107	0.0947	0.4689	0.2854	0.4115	0.5934
rs2219594	A	0.030303	0.004	0.0177	0.0178	0.1106			7.6563	1.0600	55.7473	0.6155	0.2249	1.6845	6.7711	0.8844	0.1155
rs45574331	C	0.015152	0.026	0.5825	0.5828	0.8960			0.5669	0.0729	4.4422	0.1909	0.0285	1.2795	0.2051	0.3619	0.6384
rs4987047	T	0.045455	0.035	0.6546	0.6549	0.9245			1.3305	0.3792	4.6923	0.2671	0.0917	0.7777	0.7593	0.5707	0.4289
rs55639415	T	0.030303	0.008	0.1010	0.1013	0.3151			3.8125	0.6845	21.3828	0.4493	0.1416	1.4257	3.0197	0.7921	0.2078
rs55689095	G	0.015152	0.002	0.0940	0.0943	0.5634			7.5538	0.4668	123.6339	0.6169	0.1514	2.5133	6.6706	0.8831	0.1169
rs55933907	A	0.015152	0.002	0.0940	0.0943	0.5634			7.5538	0.4668	123.6339	0.6169	0.1514	2.5133	6.6706	0.8831	0.1169
rs80358765	G	0.075758	0.002	0.0000	0.0000	0.0000	Significant	Significant	40.2459	4.6252	353.3066	0.9438	0.6146	1.4495	39.2640	0.9756	0.0242
rs80359052	A	0.045455	0.002	0.0001	0.0001	0.0016	Significant		23.3810	2.3954	230.3462	0.8637	0.4685	1.5923	22.4200	0.9589	0.0410
rs1799955	G	0.060606	0.159	0.0349	0.0350	0.0543			0.3424	0.1211	0.9726	0.1790	0.0670	0.4787	0.0888	0.2594	0.7575
rs1801406	G	0.075758	0.165	0.0604	0.0607	0.0899			0.4159	0.1621	1.0714	0.1874	0.0775	0.4528	0.1241	0.2985	0.7177
rs206075	A	0.030303	0.110	0.0437	0.0439	0.0720			0.2535	0.0603	1.0713	0.1632	0.0411	0.6488	0.0517	0.2041	0.8051
rs206076	C	0.060606	0.110	0.0000	0.0000	0.0000	Significant	Significant	125.7222	43.9997	360.7741	0.5435	0.2019	1.4631	49.7103	0.3954	0.0031
rs34351119	G	0.030303	0.026	0.8548	0.8549	0.8241			1.1514	0.2540	5.2519	0.2512	0.0677	0.9315	0.6162	0.5351	0.4648
rs36060526	C	0.030303	0.026	0.8548	0.8549	0.8241			1.1514	0.2540	5.2519	0.2512	0.0677	0.9315	0.6162	0.5351	0.4648
rs543304	C	0.227273	0.197	0.5662	0.5666	0.6818			1.1977	0.6462	2.2255	0.2483	0.1451	0.4248	0.6477	0.5408	0.4516
rs61754138	A	0.015152	0.020	0.7765	0.7767	0.8512			0.7415	0.0934	5.9378	0.2097	0.0319	1.3782	0.3158	0.4258	0.5743
rs11571623	T	0.015152	0.028	0.5303	0.5307	0.8241			0.5253	0.0679	4.0951	0.1864	0.0277	1.2551	0.1810	0.3445	0.6559
rs11571673	A	0.121212	0.008	0.0000	0.0000	0.0000	Significant	Significant	16.8276	4.9150	57.9035	0.7729	0.4839	1.2345	15.8676	0.9430	0.0560
rs11571743	G	0.015152	0.043	0.2805	0.2809	0.4578			0.3451	0.0456	2.6301	0.1667	0.0242	1.1469	0.0886	0.2568	0.7444
rs11571744	T	0.075758	0.055	0.4933	0.4937	0.6868			1.4117	0.5241	3.8178	0.2722	0.1176	0.6299	0.8253	0.5846	0.4141
rs11571823	A	0.106061	0.006	0.0000	0.0000	0.0000	Significant	Significant	19.3390	4.8690	77.2457	0.8077	0.5038	1.2948	18.3760	0.9502	0.0491
rs1799943	A	0.045455	0.049	0.9059	0.9060	0.8515			0.9286	0.2718	3.1888	0.2298	0.0771	0.6846	0.4472	0.4816	0.5186
rs206080	C	0.060606	0.890	0.2192	0.2197	0.3107			1.9110	0.6688	5.4837	0.1930	0.0729	0.5110	0.7287	0.3813	0.1995
rs76874770	T	0.045455	0.022	0.2599	0.2604	0.4793			2.0823	0.5656	7.7070	0.3301	0.1179	0.9242	1.4060	0.6752	0.3243
rs81002804	G	0.045455	0.002	0.0001	0.0001	0.0016	Significant		23.3810	2.3954	230.3462	0.8637	0.4685	1.5923	22.4200	0.9589	0.0410
rs81002819	G	0.045455	0.004	0.0008	0.0008	0.0079	Significant		11.6667	1.9124	71.6987	0.7139	0.3364	1.5151	10.7438	0.9209	0.0789
rs81002838	A	0.015152	0.002	0.0940	0.0943	0.5634			7.5538	0.4668	123.6339	0.6169	0.1514	2.5133	6.6706	0.8831	0.1169
rs81002852	A	0.030303	0.006	0.0501	0.0503	0.2063			5.0938	0.8352	31.2960	0.5157	0.1720	1.5462	4.2572	0.8358	0.1641
rs9534262	T	0.484848	0.441	0.2559	0.2563	0.3152			1.3465	0.8049	2.2571	0.2397	0.1524	0.3770	0.7239	0.5376	0.3993

Table 3.7. Statistical comparison of *BRCA1* SNV allelic frequencies between the Control group and the 1KG group.

Allele Reference		Omni5 Data	1KG Data	Chi-square test					Odds Ratio (95% C.I.)			Relative Risk (95% C.I.)			Relative Risk to Population Average Risk		
rsID	A1	MAF	MAF	Chi-square (df=1)	V-squared (df=1)	Yates (df=1)	p<0.01	p<0.001	Odds Ratio	Lower Limt OR (95% C.I.)	Upper Limit OR (95% C.I.)	Relative Risk	Lower Limt RR (95% C.I.)	Upper Limit RR (95% C.I.)	RR(AA)*	RR(Aa)**	RR(aa)***
rs16941	G	0.069	0.128	0.0292	0.0294	0.0409			0.5059	0.2715	0.9450	0.4608	0.2766	0.7677	0.1713	0.3387	0.6695
rs16942	G	0.117	0.215	0.0033	0.0033	0.0047	Significant		0.4825	0.2944	0.7923	0.4726	0.3165	0.7057	0.1609	0.3335	0.6911
rs1799966	G	0.117	0.220	0.0024	0.0024	0.0034	Significant		0.4711	0.2877	0.7731	0.4710	0.3152	0.7038	0.1548	0.3286	0.6975
rs2227945	G	0.074	0.049	0.1921	0.1925	0.2637			1.5690	0.7935	3.1110	0.6395	0.4138	0.9881	0.9570	0.6099	0.3887
rs799917	C	0.074	0.136	0.0000	0.0000	0.0000	Significant	Significant	0.5104	0.2794	0.9345	0.4633	0.2831	0.7582	0.1741	0.3410	0.6681
rs1060915	C	0.069	0.128	0.0292	0.0294	0.0409			0.5059	0.2715	0.9450	0.4608	0.2766	0.7677	0.1713	0.3387	0.6695
rs16940	C	0.069	0.128	0.0292	0.0294	0.0409			0.5059	0.2715	0.9450	0.4608	0.2766	0.7677	0.1713	0.3387	0.6695
rs1799949	T	0.117	0.215	0.0033	0.0033	0.0047	Significant		0.4825	0.2944	0.7923	0.4726	0.3165	0.7057	0.1609	0.3335	0.6911
rs28897679	G	0.532	0.018	0.0000	0.0000	0.0000	Significant	Significant	60.9812	29.7144	125.5163	1.0715	0.8771	1.3091	58.8536	0.9651	0.0158
rs8176316	A	0.027	0.016	0.3682	0.3686	0.5579			1.6712	0.5396	5.1992	0.6568	0.3265	1.3210	1.0454	0.6255	0.3743
rs80358035	G	0.043	0.008	0.0023	0.0023	0.0065	Significant		5.4218	1.6129	18.3155	0.9361	0.6156	1.4235	4.5762	0.8440	0.1557
rs55946644	C	0.022	0.020	0.8937	0.8938	0.8646			1.0832	0.3355	3.5146	0.5554	0.2403	1.2836	0.5632	0.5200	0.4800

*Relative risk to population average risk for homozygotes of the minor allele

**Relative risk to population average risk for heterozygotes

***Relative risk to population average risk for homozygotes of the major allele

Table 3.8. Statistical comparison of *BRCA2* SNV allelic frequencies between the Control group and the 1KG group.

Allele Reference		Omni5 Data	1KG Data	Chi-square test					Odds Ratio (95% C.I.)			Relative Risk (95% C.I.)			Relative Risk to Population Average Risk		
rsID	A1	MAF	MAF	Chi-square (df=1)	V-squared (df=1)	Yates (df=1)	p<0.01	p<0.001	Odds Ratio	Lower Limt OR (95% C.I.)	Upper Limit OR (95% C.I.)	Relative Risk	Lower Limt RR (95% C.I.)	Upper Limit RR (95% C.I.)	RR(AA)*	RR(Aa)**	RR(aa)***
rs11571831	A	0.085	0.026	0.0007	0.0007	0.0015	Significant		3.4277	1.6154	7.2955	0.8159	0.5737	1.1604	2.6492	0.7729	0.2255
rs144848	C	0.059	0.098	0.1057	0.1059	0.1427			0.5749	0.2918	1.1355	0.4715	0.2727	0.8150	0.2107	0.3665	0.6376
rs169547	T	0.037	0.096	0.0119	0.0120	0.0185			0.3661	0.1624	0.8280	0.4188	0.2075	0.8450	0.0987	0.2696	0.7363
rs1801426	G	0.117	0.126	0.7499	0.7500	0.8505			0.9190	0.5472	1.5465	0.5404	0.3690	0.7915	0.4407	0.4795	0.5218
rs2219594	A	0.011	0.004	0.3161	0.3165	0.6586			2.6348	0.3685	18.9922	0.7752	0.2888	2.0810	1.9099	0.7249	0.2751
rs45574331	C	0.085	0.026	0.0007	0.0007	0.0015	Significant		3.4277	1.6154	7.2955	0.8159	0.5737	1.1604	2.6492	0.7729	0.2255
rs4987047	T	0.048	0.035	0.4026	0.4030	0.5400			1.4208	0.6219	3.2570	0.6176	0.3588	1.0631	0.8335	0.5866	0.4129
rs55639415	T	0.016	0.008	0.3658	0.3661	0.6314			1.9787	0.4387	8.9803	0.7035	0.2965	1.6693	1.3143	0.6642	0.3357
rs1799955	G	0.085	0.159	0.0131	0.0132	0.0184			0.4938	0.2802	0.8721	0.4637	0.2917	0.7373	0.1653	0.3348	0.6781
rs1801406	G	0.085	0.165	0.0080	0.0080	0.0114			0.4720	0.2683	0.8324	0.4600	0.2889	0.7323	0.1536	0.3253	0.6892
rs206075	A	0.053	0.110	0.0000	0.0000	0.0000	Significant	Significant	0.4557	0.2270	0.9174	0.4452	0.2486	0.7973	0.1436	0.3151	0.6915
rs34351119	G	0.090	0.026	0.0003	0.0003	0.0006	Significant	Significant	3.6633	1.7427	7.7240	0.8298	0.5916	1.1638	2.8725	0.7841	0.2141
rs543304	C	0.213	0.197	0.6499	0.6501	0.7286			1.1008	0.7277	1.6680	0.5646	0.4204	0.7581	0.5746	0.5219	0.4741
rs61754138	A	0.005	0.020	0.1654	0.1657	0.2949			0.2577	0.0328	2.0447	0.3704	0.0569	2.4101	0.0528	0.2050	0.7953
rs11571623	T	0.027	0.028	0.8953	0.8954	0.8977			0.9330	0.3314	2.6382	0.5400	0.2520	1.1575	0.4504	0.4827	0.5174
rs11571743	G	0.070	0.043	0.1475	0.1478	0.2109			1.6853	0.8258	3.4493	0.6510	0.4168	1.0167	1.0564	0.6268	0.3719
rs11571744	T	0.081	0.055	0.2143	0.2146	0.2877			1.5108	0.7847	2.9165	0.6260	0.4091	0.9580	0.9077	0.6008	0.3977
rs1799943	A	0.032	0.049	0.3381	0.3385	0.4538			0.6428	0.2585	1.6042	0.4800	0.2322	0.9923	0.2517	0.3916	0.6093
rs76874770	T	0.037	0.022	0.2798	0.2801	0.4158			1.6909	0.6455	4.4467	0.6623	0.3663	1.1975	1.0622	0.6282	0.3715
rs9534262	T	0.463	0.441	0.0246	0.0247	0.0306			1.0918	0.7791	1.5321	0.5548	0.4349	0.7078	0.5598	0.5128	0.4697

*Relative risk to population average risk for homozygotes of the minor allele

**Relative risk to population average risk for heterozygotes

***Relative risk to population average risk for homozygotes of the major allele

3.3.6. Control group data versus 1KG group data (*BRCA2*)

Twenty out of 34 identified SNVs within the *BRCA2* gene were comparable between the Control group and the 1KG group. Two allele frequencies out of twenty tested were statistically different at the $p<0.01$ level and two allele frequency tested statistically different at both the $P<0.01$ and $p<0.001$ levels (Table 3.8.).

Seven alleles may confer increased risk for the development of breast cancer. Only three allele frequency differences (out of the seven) were statistically significant. The allele rs34351119 was of the most interest, this allele's frequency difference was significantly between the two groups and it conferred an increase of 2.87 fold in risk of developing breast cancer when compared to the 1KG group data.

3.3.7. Novel alleles and alleles not present in the Control group and 1KG group

Amongst all the SNVs found in the BRCA cohort, the allele frequencies for 27 different SNVs were not available as they were either novel variants discovered in this study or they were not included in the Human Omni 5 BeadChip design as the Omni 5 BeadChip utilized common/frequent tag SNPs. Of the 27 SNVs, 6 were novel SNVs in this study and 21 SNVs were not included in the Human Omni 5 BeadChip design (5 in *BRCA1*, 16 in *BRCA2*). Of these 21 SNVs, 15 allelic frequencies were available from the 1KG group data set, and the remaining 6 SNVs were not genotyped in the 1KG project.

The allele frequencies of these twenty seven SNVs in an ethnical-matched control group will be determined as a part of a future study. Due to financial constrains experienced in this project, these alleles were not screened to determine the allele frequencies in a control group.

3.3.8. Allele co-segregation pattern identification

SNVs found in the *BRCA1* and *BRCA2* gene within the Control group were studied to determine patterns of the co-segregation of alleles. Nine alleles in the *BRCA1* gene were informative for analysis. A total of 4 distinguishable patterns of allele co-segregation were identified in the *BRCA1* gene. Twelve alleles in the *BRCA2* gene were informative for this analysis. A total of 6 distinguishable patterns of allele co-segregation were identified in the *BRCA2* gene. Table 3.9a. shows the identified patterns for the *BRCA1* gene and Table 3.9b. shows the identified patterns for the *BRCA2* gene for each of the 94 samples screened by May et al. (2013).

The allele co-segregation patterns were not studied in the BRCA Cohort as not enough informative alleles were identified within the cohort. The observed frequencies of each identified co-segregation pattern are tabulated in Table 3.10.

Table 3.9. Distinguishable allele co-segregation patterns identified in the *BRCA1* & *BRCA2* gene in the Control group. a). *BRCA1* allele co-segregation patterns identified using 9 informative alleles. Four distinguishable allele co-segregation patterns are present, the H1 (yellow) combination is most common in *BRCA1* (12.76%). b). *BRCA2* allele co-segregation patterns identified using 12 informative alleles. Six distinguishable allele co-segregation patterns are present, the H7 (green) combination is most common in *BRCA2* (21.28%). # indicates the sample ID. Blank blocks indicates the sample's allele is wildtype for both loci. All informative allele are heterozygous

a. BRCA1 (Sample ID #01–#47)

rsID	c.DNA	#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27	#28	#29	#30	#31	#32	#33	#34	#35	#36	#37	#38	#39	#40	#41	#42	#43	#44	#45	#46	#47
rs 8176316	c.5531-10C>A				CA																																											
rs 55946644	c.5469+8T>C																																													TC		
rs 80358035	c.5050-20A>G										AG					AG				AG													AG							AG				AG				
rs 1799966	c.4900A>G							AG				AG		AG	AG				AG														AG										AG			AG	AG	
rs 1060915	c.4308T>C							TC							TC				TC														TC														TC	
rs 16942	c.3548A>G							AG				AG		AG	AG				AG														AG										AG			AG	AG	
rs 2227945	c.3418A>G				AG				AG		AG			AG		AG				AG													AG							AG				AG				
rs 16941	c.3113A>G							AG							AG				AG														AG														AG	
rs 16940	c.2311T>C							TC							TC				TC														TC														TC	
rs 1799949	c.2082C>T							CT				CT		CT	CT				CT														CT										CT			CT	CT	

a. BRCA1 (Sample ID #48–#94)

rsID	c.DNA	#48	#49	#50	#51	#52	#53	#54	#55	#56	#57	#58	#59	#60	#61	#62	#63	#64	#65	#66	#67	#68	#69	#70	#71	#72	#73	#74	#75	#76	#77	#78	#79	#80	#81	#82	#83	#84	#85	#86	#87	#88	#89	#90	#91	#92	#93	#94
rs 8176316	c.5531-10C>A	CA		CA									CA							CA																												
rs 55946644	c.5469+8T>C											TC																			TC																	TC
rs 80358035	c.5050-20A>G																																				AG		AG									
rs 1799966	c.4900A>G											AG			AG				GG					AG	AG		AG	AG			AG					AG		AG								AG		AG
rs 1060915	c.4308T>C														TC				TC					TC			TC	TC								TC										TC		
rs 16942	c.3548A>G											AG			AG				GG					AG	AG		AG	AG			AG					AG		AG								AG		AG
rs 2227945	c.3418A>G																					AG				AG						AG					AG		AG									
rs 16941	c.3113A>G														AG				AG					AG			AG	AG								AG										AG		
rs 16940	c.2311T>C														TC				CC					TC			TC	TC								TC										TC		
rs 1799949	c.2082C>T											CT			CT				CT					CT	CT		CT	CT			CT					CT		CT								CT		CT

b. BRCA2 (Sample ID #01–#47)

rsID	c.DNA	#01	#02	#03	#04	#05	#06	#07	#08	#09	#10	#11	#12	#13	#14	#15	#16	#17	#18	#19	#20	#21	#22	#23	#24	#25	#26	#27	#28	#29	#30	#31	#32	#33	#34	#35	#36	#37	#38	#39	#40	#41	#42	#43	#44	#45	#46	#47
rs 11571623	c.517-19C>T											CT														CT																						
rs 144848	c.1114A>C				AC																AC								AC												AC		AC					
rs 1801406	c.3396A>G						AG		AG						AG			AG								AG																			AG			
rs 543304	c.3807T>C		TC	TC		TC		TC												CC		TC	TC				TC				TC			CC	TC	TC	TC		TC	TC					TC	TC		
rs 34351119	c.4563A>G					AG		AG					AG	AG	AG	AG							AG	AG				AG				AG						AG										
rs 45574331	c.7017G>C	GG				GC		GC					GC	GC	GC	GC							GC					GC				GC						GC										
rs 1799955	c.7242A>G						AG		AG						AG			AG								AG																			AG			
rs 169547	c.7397C>T											CT				CT										CT																						
rs 11571744	c.8487+47C>T			CT																TT		CT													CT				CT	CT						CT		
rs 4987047	c.8830A>T				AT																AT								AT												AT							
rs 11571831	c.9730G>A					GA		GA					GA	GA	GA	GA							GA					GA				GA						GA										
rs 1801426	c.10234A>G		AG			AG		AG	AG														AG				AG				AG			AG			AG								AG			

b. BRCA2 (Sample ID #48–#94)

rsID	c.DNA	#48	#49	#50	#51	#52	#53	#54	#55	#56	#57	#58	#59	#60	#61	#62	#63	#64	#65	#66	#67	#68	#69	#70	#71	#72	#73	#74	#75	#76	#77	#78	#79	#80	#81	#82	#83	#84	#85	#86	#87	#88	#89	#90	#91	#92	#93	#94
rs 11571623	c.517-19C>T									CT																CT																			CT			
rs 144848	c.1114A>C		AC							AC																								AC		AC		AC					AC					
rs 1801406	c.3396A>G	AG																AG			AG						AG							AG					AG		AG			AG		AG		AG
rs 543304	c.3807T>C	TC						TC	TC			CC		TC	TC		TC		CC						CT			TC	TC		TC	TC			TC	TC	TC		TC				TC					TC
rs 34351119	c.4563A>G						AG						AG	AG	AG																							AG								AG		
rs 45574331	c.7017G>C						GC						GC	GC	GC																							GC								GC		
rs 1799955	c.7242A>G	AG																AG			AG						AG							AG					AG		AG			AG		AG		AG
rs 169547	c.7397C>T									CT																TT																			CT			
rs 11571744	c.8487+47C>T	CT										CT																	CT		CT				CT		CT											CT
rs 4987047	c.8830A>T		AT							AT																								AT		AT							AT					
rs 11571831	c.9730G>A						GA						GA	GA	GA																							GA								GA		
rs 1801426	c.10234A>G							AG	AG			AG		AG			AG		GG						AG							AG				AG			AG				AG					

Table 3.10. Observed frequencies of the distinguishable allele co-segregation patterns in the *BRCA1* and *BRCA2* gene in the Control group. The colours in the table below corresponds with the colours of allele co-segregation patterns in Table 3.9.

Co-segregation pattern	Gene	Colour	Observed Frequency (n=94)	Observed Frequency in Percentage
H1	*BRCA1*	Yellow	12	12.77%
H2	*BRCA1*	Green	4	4.26%
H3	*BRCA1*	Blue	5	5.32%
H4	*BRCA1*	Pink	8	8.51%
H5	*BRCA2*	Purple	16	17.02%
H6	*BRCA2*	Teal	5	5.32%
H7	*BRCA2*	Green	20	21.28%
H8	*BRCA2*	Orange	9	9.57%
H9	*BRCA2*	Grey	14	14.89%
H10	*BRCA2*	Yellow	16	17.02%

3.4. MLPA analysis data

The BRCA cohort was screened for large gene/exon duplications and/or deletions in the *BRCA1* and *BRCA2* genes using MLPA. *BRCA1* MLPA analysis was performed using the SALSA MLPA P002-C2 (Lot. 0811) probemix (MRC-Holland, the Netherlands) and *BRCA2* MLPA analysis was performed using the SALSA MLPA P090-A4 (Lot. A4-0712) probemix (MRC-Holland, the Netherlands).

All 33 patients within the BRCA cohort tested negative for large gene/exon duplications and/or deletions in either *BRCA1* or *BRCA2*.

The only result of significance is that the exon 5 probes of the *BRCA1* P002-C2 kit showed a large degree of binding variability amongst the patients of the BRCA cohort. It was bioinformatically determined that no SNVs reported in this study were present within the binding domains of the probes. Probe binding variability maybe due to the quality of the sample DNA used for the MLPA analysis. Certain probes are more sensitive to DNA quality. Variation in DNA quality and quantity may affect probe binding efficiency. Alternatively other chemical inhibitors present within the solution may interfere with probe binding. This have been observed previously in the Molecular Diagnostic Laboratory, NHLS (Essop F, personal communication). Lastly the cause of probe binding variability may be due to problems with the probes themselves. Such problems may arise during probe production and/or probe storage.

4. DISCUSSION

The genetic aetiology of HBC in Caucasian populations has been extensively studied. Current findings and evidence indicate that mutations in the genes *BRCA1 and BRCA2* account for a significant portion of inherited breast cancer in the Caucasian population. While 90% of all breast cancer cases are sporadic, only 10% are hereditary in nature. Amongst the 10% of hereditary breast cancer, *BRCA1* and *BRCA2* mutations account for up to 15% of the cases (Walsh & King, 2007). The prevalence of *BRCA* mutations differs amongst populations of different ethnicity (Petrucelli et al., 1998). By definition, a confirmed *BRCA* mutation indicates HBOC. HBOC differs from HBC as the causality of HBC may be non-*BRCA* related. From a clinical perspective, distinction between HBC and HBOC would have a significant impact on treatment and management strategies. The management of HBOC is different to that of sporadic breast cancer. A more frequent and stringent surveillance programme to monitor and identify possible development of cancer is required (Petrucelli et al., 1998). Risk reduction prophylactic surgeries must also be considered in patients diagnosed with HBOC.

South Africa is a country heavily burdened by infectious diseases, such as HIV and TB, and the enormity of these issues often overshadow the fact that this country is also significantly burdened by non-communicable diseases such as cancer. Breast cancer is the most prevalent cancer amongst all females in South Africa, and it is the second most prevalent cancer amongst black females in South Africa after cervical cancer (Ferlay et al., 2010; CANSA, 2014). Studies aimed at elucidating the causes of cancers and the early detection of breast cancer amongst the female populations in South Africa should take high priority. Although the cause of 90% of all breast cancers cases are sporadic in nature, the remaining 10% have underlying genetic predispositions that may be detected early, which would have significant impact on treatment and management of the disease. Risk reduction options such as intense surveillance for early disease development, prophylactic surgeries and drug therapies will become available to patients with identified genetic mutations and predispositions.

A recently published large scale genotyping study conducted by the COGS (Collaborative Oncological Gene-Environment Study) consortium has identified numerous alleles conferring significantly increased risk for the development of breast cancer. This study used DNA samples taken from breast cancer patients and control patients of European ancestry (Michailidou et al., 2013). The findings of this study have a significant impact on identifying the genetic risk factors that contribute to the development of breast cancer in European patients. This findings suggests that genetic risk factors or susceptibility factors or genes do exist. Information from this dataset may be useful as a starting point to investigate the genetic risk factors that may be present within patients of other ancestries.

This can be achieved by screening for the presence and the frequency at which these genetic risk factors exist in other populations. Individuals of Southern African ancestry may benefit from the analysis of novel genetic risk factors associated with the development of breast cancer that have been identified in the COGS in European patients, as these genetic risk factors have never been investigated in females of African descent. The COGS also illustrates the point that investigating the genetic aetiology of breast cancer on a genome-wide level is important as a tool, in addition to the traditional targeted approach of screening high risk candidate genes. This approach would also allow for the identification of other disease related variants not described in European populations.

This preliminary study aimed to gain a better understanding of the genetic aetiology underlying breast cancer in South African black women. The study design followed the targeted gene approach, screening the *BRCA1* & *BRCA2* genes for genetic variants. This approach was used to determine whether *BRCA1* and *BRCA2* gene mutations are the predominant cause of inherited breast cancer within a high risk, black cohort (as they are for high risk Caucasoids) and to determine the percentage of *BRCA* mutations present within this cohort. In contrast to European and American populations, little work on the genetics of HBC has been done in the South African black population.

Thirty three South African black females matching the predefined selection criteria (selection criteria outlined in section 2.2.1.) were recruited for this study. The mean age of disease diagnosis was 39 years of age. The youngest patient was 27 years of age, and the oldest patient was 50 years of age. Thirteen patients had a family history of cancer (i.e. one or more relatives, first, second or third degree relatives, of the patient had been diagnosed with a form of cancer). Six patients had TNBC. One patient had bilateral breast cancer.

Using a modified risk calculation method proposed by Wainstein (2011), 22 patients out of the 33 were assigned a high risk category and eleven patients fell into a moderate risk category. All patients were of South African black ethnicity and were recruited from the Breast and Plastic Clinic at the Chris Hani Baragwanath Hospital, Soweto, Johannesburg, South Africa.

Due to the nature of the clinic being focused on breast disease and plastic surgery, patients were not accessed for ovarian cancer as the primary disease diagnosis. The occurrence of both breast cancer and ovarian cancer within the same family is highly indicative of *BRCA* gene mutations (Petrucelli et al., 1998). Ovarian cancer data contribution would have aided our patient selection process by enriching for likely *BRCA* mutation carriers. Due to lack of ovarian cancer data from the participants, we were not able to comment on the prevalence of *BRCA* gene mutations in families with both breast and ovarian cancer.

Sanger sequencing chemistry was utilized for this project. This chemistry is routinely used for diagnostic testing services and all equipment required for Sanger sequencing was readily available within the Molecular Diagnostic Laboratory, Division of Human Genetics, NHLS Braamfontein. The inclusion of universal sequencing tag primers allowed for a batch sequencing setup and was very effective in streamlining the Sanger sequencing workflow. The universal sequencing tag design facilitated rapid sequencing setup and clean-up in the 96 well plate format thus greatly reducing the hands-on time required. This application allowed us to complete this study within our set timeframe and greatly reduced the cost of this project with respect to reagents and consumables. This application was also successfully adopted and incorporated into other tests conducted within the laboratory.

All coding regions and intron/exon boundaries of *BRCA1* and *BRCA2* were successfully sequenced for all 33 patients. An average read coverage of 2X was achieved for all target regions. Forward and reverse sequencing of the target region was carried out for each amplicon. This was done to allow the separation of true sequence variants from sequencing artefacts. The amplicons containing any detected sequence variants were PCR amplified again and re-sequenced, to validate the presence of the sequence variant. Three confirmed or likely pathogenic mutations were identified: one novel frameshift mutation in *BRCA1* and two different previously reported pathogenic mutations (on HGMD) in *BRCA2*.

MLPA analysis for *BRCA1* and *BRCA2* was also established in our laboratory and successfully performed for all patients. All patients were screened for large gene/exon deletions and/or duplications. No patient tested positive for a large gene/exon deletion and/or duplication.

As an additional objective, genotype data from the BRCA cohort were statistically compared with the genotype data from a Control group and the 1KG group. This was done to determine whether any similarities and/or differences with risk associations of specific genetic alleles between populations. And to look into if the frequencies of SNVs and SNPs within the *BRCA* genes differ in populations from different parts of Africa. This objective was added to the study, as during the course of generating sequencing data for the *BRCA1* and *BRCA2* gene, a large amount of genetic variation, in the form of single nucleotide variants/polymorphisms, was detected in both *BRCA1* and *BRCA2* amongst the BRCA cohort. Increased genetic variation level in normal black individuals is a phenomenon well described in the literature (discussed in May et al., 2013). The Control group is ethnically matched to the BRCA cohort, and the 1KG group is predominantly of African origin, not specifically southern African. 1KG genotyping were generated from low coverage whole genome sequencing and the Control group genotyping was done on SNP array. The difference in technology

could be a limiting factor in interpreting results as the SNP array were designed to only include "common" SNPs. The absence of a SNP on a SNP array does not necessarily mean that the SNP is absent from the sample population.

This chapter will discuss in detail the findings of this study and the possible implications these findings may have. The limitations of this study and recommendations arising from this study will also be discussed. This chapter will follow the order of information flow set out in the previous chapter.

4.1. Mutation screening

From the data generated from this study, mutations in the *BRCA* genes account for approximately 10% (3/33) of breast cancer cases from our high risk cohort. By definition, a positive *BRCA* gene mutation is the criteria for HBOC diagnosis, although ovarian cancer data is lacking for the patient cohort. Despite the strong evidence of the hereditability of the cancers as determined by family history, *BRCA* gene mutation prevalence is not higher than the international literature citing of 0.5% - 9.3% for *BRCA1* and 1.3% - 12.5% for *BRCA2* (Peto et al., 1999; Group, 2000; John et al., 2007; Han et al., 2013). Therefore *BRCA* mutations are not the predominant cause of breast cancer in this high-risk, young black cohort, although they do make a notable contribution. Predictive testing should be offered to all the relatives of a patient where a confirmed *BRCA* mutation was identified.

4.1.1. Cohort selection parameters

Upon conception of this study, the patient selection criteria were based on a set of indicators that potentially indicated or suggested an increased risk of inherited *BRCA* mutations. These criteria included young age at diagnosis of disease; family history of breast and/or related cancers; triple negative breast cancer status; breast cancer bilaterality and breast cancer risk categorization. We hypothesized that by preselecting patients based on these indicators, we could potentially enrich for *BRCA* mutation positive patients and increase our rate of mutation detection within the *BRCA* genes. This decision had two implications on the outcome of this study: firstly, the mutation pick-up rate within this cohort would be a bias and an over representation of the actual percentage of *BRCA* mutation positive breast cancer patients within the overall South African black breast cancer disease population. Secondly, if this set of indicators were to successfully enrich the rate of detection of *BRCA* mutations within this cohort then potentially this set of indicators could be employed by routine diagnostic laboratories to pre-screen patients prior to recommending *BRCA* mutation screening, as *BRCA* sequencing is very expensive. To fully evaluate both implications, the prevalence of *BRCA* mutations within an unselected breast cancer cohort of matching ethnicity would need to

be determined. *BRCA* mutation screening in an unselected breast cancer cohort could be the next strategic step in our research efforts but does not form part of this study due to limited resources.

A summary of the patients' against the set of indicators can be found in Table 2.1. Age at diagnosis was the primary indicator. Within our study, the mean age at disease diagnosis was 39 years of age (Ranging from 27 years to 50 years). This is considerably younger than the age of disease onset of sporadic invasive breast cancer diagnosed in European populations, where the largest portion of disease occurs within the age range of 50 to 69 years of age (Cancer Research UK, 2013). The younger age at disease diagnosis amongst the African population has been a topic of interest and concern over the past decade. Two explanations have been put forward: firstly, the high prevalence of young breast cancers is due to the underlying genetic predispositions that may exist within the African population so that African females are more susceptible to developing breast cancer; or secondly, the high prevalence of young breast cancers in Africa may be a consequence of the cone-shaped population structure of developing countries and third world countries: the high fertility rate and high infant mortality rate results in a skewed population age distribution with the majority of the population being young adults thus leading to the presentation of more young breast cancer (Akarolo-Anthony et al., 2010). With regards to disease outcome, young breast cancers occurring before the age of 40 years are known to be associated with adverse outcomes, however the underlining molecular mechanisms are not well understood (Anders et al., 2009).

Results of this study indicate that age at diagnosis may be important but it is not particularly informative with regard to the presence of *BRCA* mutations. The three patients in this cohort that tested positive for *BRCA* mutations were aged 30, 36 and 42. Statistically, correlation between age at diagnosis and presence of *BRCA* mutations cannot be made due to the small cohort size.

Family history was a more useful indicator of the potential heritable component of cancer. Positive history of breast cancer within a family may indicate germline predisposition to breast cancer and/or associated cancers but does not necessarily indicate that the germline predisposition exists in either of the *BRCA* genes. In this study, 2/3 mutation-positive patients had a positive family history of breast or related cancers through the maternal lineage.

TNBC status is known to be an indicator of *BRCA1* mutation status (Lee et al., 2011). Within our cohort two of the three patients with pathogenic mutations have TNBC and one of these patients carries a *BRCA1* mutation. The small cohort size limits us from making any further interpretations of tumour receptor status and its relation to *BRCA* gene mutations within the South African cohort as

the number of patients and the number of *BRCA* mutation-positive patients is too small to draw any statistically significant correlation.

Breast cancer laterality was the 4th indicator used in this study. Breast cancer patients who carry germline *BRCA* gene mutations are more likely to develop contralateral breast cancer as opposed to patients who are *BRCA* mutation negative. The risk for *BRCA* mutation carriers of developing a second primary breast cancer is approximately 30% at ten years. That is approximately four-fold higher than the general population risk (Metcalfe et al., 2004; Gronwald et al., 2006). Considering that bilateral breast cancer occurrence is much higher in *BRCA* mutation positive patients, patients presenting with bilateral breast cancer are at higher risk of carrying a *BRCA* mutation. However cancer laterality was not informative in this small study. All three *BRCA* mutation positive patients presented with unilateral breast cancer. The only patient (BRCA026) with bilateral breast cancer in this cohort tested *BRCA* mutation negative.

Risk categorization for the BRCA cohort was done by Wainstein (2011). The risk profiles of breast cancer patients were evaluated using a combination of pedigree analysis, Claus Table analysis, the Manchester scoring system and the Tyler-Cuzick model. The Wainstein study was undertaken to evaluate standard international (Caucasian population-based) tools currently used to categorise breast cancer patients into risk categories and how they perform with regards to *BRCA* mutation prediction in a black South African population. The scores from all four systems were then standardized to produce a composite risk for each patient. The BRCA cohort comprised patients of either “high” risk or “moderate” risk. All three patients who tested positive for *BRCA* mutations had been classified as high risk breast cancer patients. This represents 13.63% (3/22) of all high risk patients within our cohort. Thus, the modified risk analysis method may be useful for predicting the presence of hereditary predispositions for larger cohorts but within our cohort, statistical correlations were not possible.

4.1.2. Mechanism of disease

Genetic mutations within the *BRCA* genes have been extensively studied across different population groups and catalogued. To date (February 2015), 14911 sequence variant entries for *BRCA1* and 14822 sequence variant entries for *BRCA2* have been reported in the BIC database (Szabo et al., 2000a). This list is by no means comprehensive. The function of the variants can be grouped into large categories:

- Known or likely to be pathogenic, deleterious, and disease related
- Neutral variants without significant clinical significance
- Variants of unknown significance

Pathogenic mutations are variants that impact on the function or amount of the protein product of the *BRCA1* and *BRCA2* genes. Mutations impacting the protein product can be classified into five classes:

1. Truncation mutations, reduced length of final product or depletion of mRNA through NMD.
2. Missense mutations, amino acid codon alterations that lead to loss-of-function of the full length protein product.
3. Regulatory region mutations, affecting sequences upstream of the gene of interest.
4. Disruption of splice sites within the gene or creation of novel splice sites.
5. Whole exon and/or partial gene deletions or duplications.

Mutations within the regulatory regions lead to loss of gene expression, thus loss of protein expression or altered protein products due to alternative mRNA splicing.

4.1.2.1. Truncation mutations

Truncating mutations produce truncated/aberrant mRNA transcripts through the introduction of premature stop codons. The premature stop codon can be introduced by means of single nucleotide mutations (point mutations) and/or small nucleotide insertions, deletions, and/or duplications. Large gene/exon deletions and/or duplications can also introduce truncating mutations. Truncation mutation can also cause exon skipping by the creation of abolishing of criterial splice sites.

Truncating mutations heavily impact the quantity of fully functional protein produced. This dosage loss occurs through the production of truncated proteins, or loss of gene expression through NMD of truncated mRNA transcripts. Often truncated proteins are subject to degradation as well, as the truncated proteins are often unstable (Larsson et al., 2008).

Two of the three mutation positive patients carried truncating mutations. Both mutations are likely to be deleterious.

The mutation found in BRCA007 generated a stop codon 15 codons downstream of the mutation site in *BRCA1,* thus producing a truncated mRNA transcript. Polymerase slippage may be the mechanism that caused this error as the mutation is a single base A duplication within a poly-A region. Predictive testing should be offered to these relatives to determine if they carry the p.Asn144Lysfs*15 mutation.

Update: this mutation was also reported by Feliubadalo et al. (2013), and the data was published after the analysis of the data from our study.

Patient BRCA014 carries a single nucleotide point mutation that generates a stop codon at the site of the mutation (p.Trp194*) in *BRCA2*. This mutation produces a truncated mRNA transcript, thus depleting the dosage of the full mRNA transcript available for translation. It is likely the truncated mRNA will be subjected to NMD due to its very short length (Chang, Imam, & Wilkinson, 2007). A strong family history of cancer is present in this family. Predictive testing should be offered to relatives to determine if they carry the p.Trp194* mutation.

Carriers of *BRCA1* or *BRCA2* mutations have been shown to be haplo-insufficient for the corresponding proteins (Moynahan, 2002; Cousineau & Belmaaza, 2007; King et al., 2007). Mutations that produce a truncated RNA transcript will greatly increase the risk of developing breast cancer in mutation carriers as the mRNA/protein produced from one copy of a functional *BRCA* gene is insufficient in maintaining cellular integrity. Loss of heterozygosity may also occur whereby, through homologous recombination, the one remaining functional copy of the *BRCA* gene is lost, alternatively a second mutation could arise somatically within the remaining copy of the *BRCA* gene (King et al., 2007). This phenomenon leads to the fulfilment of Knudson's two hit hypothesis for tumour suppressor genes, whereby the loss of both alleles lead to cancer formation (Knudson, 1971).

The identification of these two mutations allows us to offer related, at-risk, family members predictive screening. Tracking the inheritance of the mutation through the families and identifying other mutation-carriers within the family would allow the implementation of effective management and screening protocols and prophylactic surgeries to reduce the risk of developing breast cancer and/or ovarian cancer in the mutation carrying individuals.

Prophylactic oophorectomy may also be considered for the affected individuals. The exact risk for the development of ovarian cancer in *BRCA* mutation carriers is uncertain but ovarian cancers have been previously described in *BRCA* mutation carriers (Bolton et al., 2012). Bilateral salpingo-oophorectomy (removal of both ovaries and the Fallopian tubes) may offer 80-90% risk reduction for the development of ovarian cancer in *BRCA* mutation positive patients of European ancestry (Reviewed in Roukos & Briasoulis, 2007). But no comparable data exist for the Sub-Saharan black populations.

4.1.2.2. Deleterious missense mutations

One likely missense pathogenic mutation was identified: c.7712A>G in patient BRCA021 causes a codon change in the BRCA2 protein (p.Glu2571Gly). As described by Kote-Jarai, et al. (2011) the 2571th amino acid of the BRCA2 protein is be highly conserved and may not tolerate change.

Four *in Silico* tools used in this study predicted the change of amino acid codon 2571 as pathogenic and intolerable. This amino acid codon is highly conserved and it is also located in a critical DNA binding domain of the BRCA2 protein which spans from amino acid codon 2500 to 3098 (Easton, Deffenbaugh, et al., 2007). Alterations to the DNA-binding domain may cause the full length protein to be dysfunctional. This is a possible explanation for the predisposition to breast cancer for this patient.

The BRCA2 protein forms part of a multi protein interaction/complex (BRCA1-PALB2-BRCA2) that facilitates the localization of RAD51 to double strand DNA breaks (Xia et al., 2006). BRCA2 also functions as a loader for RAD51 recombinase protein (the BRCA2 protein physically directs the RAD51 recombinase protein onto the site of double stranded DNA breakage) in the process of telomere replication and capping (Badie et al., 2010). Disruption to the binding domain of the BRCA2 protein may greatly reduce the binding affinity of the BRCA1, PALB2 and RAD51 protein complex to DNA.

However, this mutation was predicated to be neutral by Karchin, et al. (2008) using their proposed *in Silico* method that analyses the potential effect a missense mutation has on the BRCA2 protein and generates a "Protein likelihood ratio". A high protein likelihood ratio is in favour of a loss-of-function of the protein given the amino acid change. This mutation (c.7712A>G) was predicted to be benign by protein likelihood ratio analysis, although no functional study (using either yeast cells or other cell lines) was performed to confirm the *in Silico* analysis result.

This is a common challenge faced when analysing DNA variants, especially rare variants. It is often difficult to evaluate the impact of missense mutations, both novel and previously reported, on the function of the protein, without conducting functional assays. Screening a control population to determine the allelic frequency of the mutation may provide some evidence and insights into its potential pathogenicity.

This variant had not been included in the design of the Omni5 Human BeadChip. This variant was detected in the 1KG cohort, with a minor allele frequency (MAF) of 0.002 for the G allele of c.7712A>G (rs55689095). The MAF is much lower than 1%, thus this allele may either be a rare African variant or a rare disease associated allele. The relative risk to average population risk associated with this allele for the homozygous minor G alleles is a 6.6 fold increase in risk for the development of breast cancer. This relative risk calculation is performed under the assumption that the average population risk is 1 (Lathrop, 1983). The likelihood of the presentation of the homozygotes carrying the G allele is very low.

It is necessary to consider all evidence available in combination in order to deduce the potential effect of this variant:

- Four out of 5 *in Silico* prediction tools (SIFT, PolyPhen_2, Align GVGD, MutationTaster & Protein likelihood ratio) predict this mutation to be deleterious
- The allele frequency of the mutant allele (G allele) in the African population is rare (less than 1 in 1000)
- The increase in relative risk associated with this mutation
- This mutation is located in the DNA-binding domain
- Previous evidence from other studies

We suggest that this mutation is likely to be deleterious. And it is important to follow up with the family and offer further testing of at risk relatives.

4.1.2.3. Regulatory mutations

Mutations within the promoter regions, 5'UTR and 3'UTR may adversely affect the expression of a gene. Disease-causing mutations within the UTRs of the *BRCA* genes are not well characterised or understood (Larsson, et al., 2008).

Mutations within the splice regulatory regions require careful consideration when interpreting the effects of the mutations. Mutations may abolish splice sites or activate cryptic splice sites, both will lead to the alternative splicing of the *BRCA* mRNA transcript.

No confirmed or potentially deleterious regulatory region mutations were detected within our BRCA cohort. Intronic sequence variants close to the exon/intron boundaries were detected within this BRCA cohort (see Table 3.1. and 3.2.). However through *in Silico* analysis, using a combination of algorithms that evaluated the mutations' potential impact on the splice sites and splice site regulatory regions, these sequences variants were deemed unlikely to have an effect on the splicing of the *BRCA* transcripts.

4.1.3. Novel sequence variants and their potential (non-pathogenic) consequences

In total, 5 novel variants were found in this study. These 5 variants are novel to the 1000 Genomes project phase I data (1KG cohort). Four of these variants were found in *BRCA1* and one variant was found in *BRCA2*. The pathogenic novel variant *BRCA1* c.431dupA was included here.

Of the 4 novel variants found in *BRCA1*, two are coding region variants and 2 are deep intronic variants (Table 3.1.). One novel variant found in *BRCA2* is located near (+10bp) the intron/exon boundary (Table 3.2.). All 5 novel variants were analysed bioinformatically using the analysis pipeline

described previously (section 2.3.4.3.2.) and they were predicted to have little or no clinical significance. None of the 5 variants were predicted to have a pathogenic effect on the BRCA proteins.

One limitation with regards to the interpretation of these novel variants in this study is that these variants were not screened for in a control cohort to determine the allelic frequencies of these variants in the general population. This was due to financial and resource limitations.

Further, *in Silico* analysis suggests these five novel variants are not of significance. Thus, no strong evidence to warrant further molecular investigation of these 5 novel variants exists. The limitations remain regarding our interpretation of these variants as we do not know if they are population variants or private variants. If our interpretations are wrong and if any of these 5 novel variants are associated with the disease, then we would expect to see the inheritance of these variants through the families of the affected individuals whereby all relatives with the disease would have the associated variants but at-risk relatives with the variants may not necessarily express the disease as this disease is not fully penetrant. Sequencing of disease affected and non-affected family members for the associated variants would thus provide insight and evidence as to the potential pathogenicity of these variants. Unfortunately, due to the design of the study, other relatives (affected or unaffected) were not available for analysis.

Alternatively, these 5 novel variants could be included as a part of a larger genotyping study used to determine the population allele frequency of the novel variants in a control cohort to evaluate their potential pathogenicity.

The five novel variants are described below:

4.1.3.1. c.3751T>G (p.Cys1251Gly) – BRCA1

A cysteine to glycine substitution is created by this variant located in exon 11 of *BRCA1* (Figure 3.8.). This variant is predicted to be benign by 3/4 *in Silico* prediction tools and predicted to be disease causing by 1/4 *in Silico* prediction tool. Although this amino acid codon is located in the largest exon of the *BRCA1*, containing the binding domain for RB, MYC, RAD50, & RAD51 (Clark et al, 2012), the codon and the nucleotide are weakly conserved. This is suggestive of either its tolerance to change in the protein, or that the codon may not be in a critical protein binding domain or that this codon has little impact on the structural integrity on the BRCA1 protein thus tolerating alterations to the codon.

4.1.3.2. c.306T>G (p.=) – BRCA1

This variant is located in *BRCA1* exon 7 (Figure 3.9.). This is a synonymous variant and does not cause a change in the amino acid codon. *In Silico* analysis tools predict this variant to be benign as no

change of amino acid is caused by this variant. This variant is located 5 nucleotides downstream of the 5' exon boundary of exon 7. Although it is not located at the traditional splice sites, it could potentially have an effect on the splicing efficiency of exon 7. There is also the possibility that this variant may activate a cryptic splice site at the mutation site and cause abnormal splicing of the *BRCA1* mRNA transcript. Numerous variants adversely affecting the splicing of the *BRCA1* & *BRCA2* mRNA have been previously described (Sanz et al., 2010; Wappenschmidt et al., 2012). There is no evidence from bioinformatic analysis to suggest that c.306A>C would activate a cryptic splice site. Alternatively, functional studies or RNA analysis would show whether this variant was affecting normal splicing. Determining the allele frequency of this variant in a control population would also aid in the interpretation of the significance of this variant.

4.1.3.3. c.5332+78C>T – BRCA1

This variant is located deep in intron 21 of *BRCA1* (Figure 3.10.). None of the *in Silico* analysis tools predicted this variant to have any effect on the *BRCA1* gene or its protein product. This variant is probably too far from the splice site boundaries to affect splicing of exon 22 but it is also known that deep intronic variants may activate cryptic splice sites (Anczuków et al., 2012). *In Silico* tools are limited in their abilities to predict the functions of deep intron variants but they can predict the activation of cryptic splice sites. Population allelic frequencies and RNA studies would offer more insight into the functional significance of this variant (Anczuków et al., 2012).

4.1.3.4. c.212+66A>G – BRCA1

This variant is located in intron 5 of *BRCA1* (Figure 3.11.). No evidence of splice site alteration or alternative splicing was presented by *in Silico* analysis. Similar to c.5332+78C>T discussed above, functional analysis of this variant is not possible with sequencing data alone. Allele frequency and RNA studies would be the logical next step in determining whether this variant is functional.

4.1.3.5. c.681+10T>G – BRCA2

The final novel variant is located close to the intron/exon boundary of intron 9 and exon 9 of *BRCA2* (Figure 3.12.). *In Silico* analysis suggests this variant does not alter the splice site of intron 9 since the variant falls outside of the splice site consensus sequences (Thanaraj & Clark, 2001). The function of this variant can be investigated in the future via RNA studies to determine its functional effect.

4.1.3.6. Previously described mutations within the African populations

The mutations described by Fackenthal, et al. (2012); Zhang, et al. (2012) and van der Merwe, et al. (2012) were not found in this study. The mutations described by Fackenthal, et al. (2012) and Zhang, et al. (2012) were identified in West African black populations. It was not surprising that these mutations were not found in the Southern African black populations. The mutations described by

van der Merwe, et al. (2012) were identified in Southern African black Xhosa population. It is possible that these mutations may be specific to the Xhosa-speaking population group. The Xhosa mutation was not found in this study. This may be due to the small sample size of the BRCA cohort. Its presence within the black populations cannot be ruled out. Future studies with larger sample sizes would be useful to evaluate the founder effect of the Xhosa mutation in other southern African black populations.

4.1.4. Known sequence variants

Fifty one previously reported single nucleotide variants were detected within the two *BRCA* genes of the BRCA cohort. The allele frequencies of these 51 SNVs were statistically evaluated and compared against the population allele frequencies from the Control group and the 1KG group. This additional analysis component was added to this study to maximize the usage of the sequence data generated by this study. This study recognized that due to the small size of the study cohorts, the statistical analysis performed has no power. Despite this short-coming, the data are presented below, in the hopes that the data will provide a hint as to the genotype associated risks within the *BRCA* genes of a black African cohort and justify further large scale molecular studies to uncover how risk alleles contribute to the development of breast cancer within the South African black population.

In addition to statistical comparisons, the genotype relative risks for the comparable alleles were also calculated. Genotype relative risk is defined as "the risk associated with a genotype relative to the genotype that does not have the risk variant (non-carrier)" (DecodeHealth, 2012). This is done to assess whether the minor alleles observed in the BRCA cohort are indeed "risk" alleles. Comparisons are made under the hypothesis that carriers carrying the "risk" allele will have an increase in the risk for developing cancer. Thus theoretically, if the minor allele is indeed a true risk allele, then homozygosity for the minor allele would confer more risk than a heterozygote. Further, a heterozygote allele will confer more risk than the homozygotes of the major (non-risk) allele. This hypothesis is formed under the assumption that the minor allele is the risk conferring allele. The homozygote minor (risk) allele is shown as "AA", heterozygotes are shown as "Aa" and homozygote wildtype (non-risk) allele is shown as "aa". With the risk allele being the allele of interest, it is therefore represented with "A".

4.1.4.1. In Silico analysis of known SNVs

Prior to statistically evaluating the allele frequencies of the known SNVs, all 51 SNVs were evaluated using the *in Silico* analysis pipeline outlined in Figure 2.2. This was done to bioinformatically determine if any of the SNVs had potential adverse effects on the BRCA protein and on the individuals that are harbouring these SNVs.

Amongst the 51 SNVs found within the BRCA cohort, 21 were missense variants, 13 were synonymous variants, and 17 were intronic variants. All 51 variants were classified as "unlikely to be pathogenic" or benign by the *in Silico* analysis pipeline used in this study (section 2.3.4.3.2.), agreeing with published reports that they are non-pathogenic population variants.

The likely pathogenic missense mutation c.7721A>G in *BRCA2* was observed within the 1KG group. The MAF is 0.002 in the 1KG group. We can speculate that this variant is not a polymorphism but possibly a rare disease-associated allele for the common complex disease of breast cancer (Schork, Murray, Frazer, & Topol, 2009). This variant is present in a population level at a MAF of 0.002, meaning out of every 1000 alleles sampled, this allele would appears twice. Thus 1 out of every 250 people sampled from a given population would carry this allele. This allele may follow an incomplete penetrance disease model, in keeping with the incomplete penetrant nature of *BRCA1* and *BRCA2* mutations in breast cancer studies (Evans et al., 2008). However, this is only speculation at this stage as there is insufficient evidence to fully prove or disprove this theory.

4.1.4.2. Statistical comparison of the allele frequencies

Statistical comparison of the allele frequencies were done for the following datasets:

- *BRCA1* SNVs:
    - BRCA cohort allele data versus Control group allele data (BRCA vs. Control)
    - BRCA cohort allele data versus 1KG group allele data (BRCA vs. 1KG)
    - Control group allele data versus 1KG group allele data (Control vs. 1KG)
- *BRCA2* SNVs:
    - BRCA cohort allele data versus Control group allele data (BRCA vs. Control)
    - BRCA cohort allele data versus 1KG group allele data (BRCA vs. 1KG)
    - Control group allele data versus 1KG group data (Control vs. 1KG)

Two different population groups were used as reference for comparison because recent studies have shown that a large degree of genetic diversity exist for populations within and across Africa (Schuster et al., 2010; Gurdasani et al., 2015). Using African populations from western or eastern Africa as the genetic reference background for a southern African population is not always appropriate. The comparisons made in the study attempted to highlight this issue with the SNVs identified in the *BRCA* genes. Within the *BRCA* genes alone, a discrepancy is already apparent when the two different population cohorts were used as population reference. Thus, ideally, ethnically matching populations must be used when evaluating and statistically comparing alleles frequencies between a disease cohort and a disease-free control cohort.

4.1.4.2.1. *BRCA1* – BRCA vs. Control

Of the 12 comparable alleles, only one allele (rs28897679, A>G) was statistically different between the two cohorts. Harbouring the homozygous major allele (wild type allele) appeared to confer an "increased" relative genotype risk for developing breast cancer.

The risk increase is estimated at 1.36 fold more if an individual is homozygous for the wild type allele (Table 3.3.). This observation is consistent with the hypothesis of "Common Disease, Common Variant (CDCV)", which implies that due to the low penetrant nature of the common disease associated alleles, the allele frequency of the disease allele in the general population may be much higher than expected (Schork et al., 2009). However this observation opposes the disease model of "Common Disease, Rare Variant (CDRV)", whereby rare genetic mutations with high penetrance drive the development of disease (Schork et al., 2009). There is no way to ascertain which disease model this variant aligns to as the cohort size is too small. Small cohort size may have skewed the true allele frequency of this allele.

4.1.4.2.2. *BRCA1* – BRCA vs. 1KG

When the 1KG group was used as the population reference, a total of 13 alleles were comparable. Three alleles (rs16942 A>G, rs1799966 A>G & rs1799949 C>T) were statistically different between the two groups. None of these three alleles conferred changes in relative genotype risk. Two additional alleles (rs55815649 G>A & rs80358035 A>G) conferred increased relative genotype risk for the homozygous minor allele but the observed frequencies for these two alleles were not statistically different between the two cohorts (Table 3.5.).

It should be noted that different African populations vary genetically between each other, but overall cluster together when compared to populations of European ancestry. Therefore in the absence of an ethnically matching group, this African Super group could be a second best alternative to use for population based comparisons.

4.1.4.2.3. *BRCA1* – Control vs. 1KG

When we compared the two different population groups against each other, amongst the 12 comparable alleles, 6 allele (50%) frequencies were statistically different (Table 3.7.) This again highlighted the large degree of normal genetic diversity within the *BRCA1* gene and reinforced the observation that a large degree of genetic diversity exists between different African populations. This observed genetic diversity in *BRCA1* cautions against the use of other African populations as population references when studying this gene in sub-Saharan African populations.

Although these alleles may be naturally occurring genetic polymorphisms within the sub-Saharan African population, the increased risk association may suggest that a population based disease susceptibility may be present within the sub-Sahara African population.

4.1.4.2.4. *BRCA2* – BRCA vs. Control

Amongst the 20 comparable alleles present in the *BRCA2* gene, only one allele was statistically different. The allele rs169547 C>T was almost exclusively observed in the minor allele form (T) within the BRCA cohort, and the major allele (C) was predominantly observed in the Control group. This allele may be a risk conferring allele and confers increased risk for the development of breast cancer. Genotype relative risk calculations indicated that the homozygous minor allele (TT) of rs169547 conferred a 504 (five hundred and four) fold increase of risk for developing breast cancer (Table 3.4.).

This was an interesting finding but it may well be wrong. However, ascertaining the true impact of this allele on disease would be very difficult without further functional/RNA analysis. This result would need to be validated before any follow-up studies is undertaken.

The statistical significance of this result needs to be verified in a larger cohort. It is difficult to determine whether this risk association is a true genotype risk association or simply a result of sampling bias. Alternatively, the 'T' allele may simply be the common allele in this population (although if this were the case, the Control group data would be expected to reflect this).

This allele was not further studied due to two considerations: firstly, the high frequency of a risk conferring allele within a population that conferred a risk increase of over five hundred fold would have translated to significantly more cases of breast cancer within this population, in comparison to other populations, questioning whether this is a "true" finding or not. The difference in incidence of breast cancer South African black populations is different from that of the European populations but the degree of difference does not suggest an underlying increase of 500 fold. Secondly, and the more likely explanation, this allele may be an ancestral allele within the *BRCA* gene. The presence of this ancestral allele may be predominantly present in the South African black population. However this allele was not observed in the Control group. It is also possible that genotyping error or chip design flaws exist on the Human Omni 5 BeadChip. This allele frequency may be the result of genotyping error and/or reaction artefacts. The risk calculation methodology did not account for this phenomenon. Thus the 500 fold increase in risk may be the result of calculation errors due to the limitation of the methodology, as the incorrect reference allele was used to calculate the risks conferred.

4.1.4.2.5. *BRCA2* – BRCA vs. 1KG

When the 1KG group was used as the population reference, 34 alleles were comparable. Amongst the 34 alleles, 9 alleles were statistically different and 16 alleles conferred a change in genotype relative risk (increases in relation to average population risk) to the development of breast cancer (Table 3.7.).

Once again this study demonstrated that if ethnically-unmatched populations were used as the population reference, false positive statistical associations and false positive risk associations may present themselves. This is purely due to population diversity and not disease association.

4.1.4.2.6. *BRCA2* – Control vs. 1KG

Four out of the 20 comparable alleles within the *BRCA2* gene were statistically different between the two reference population cohorts. The degree of genetic diversity within *BRCA2* is lower than that observed for *BRCA1*. However the genetic diversity observed in *BRCA2* reinforces the notion that normal genetic variation does indeed exist between African populations.

When combined, a 35% allelic difference is observed in the *BRCA1* and *BRCA2* genes between the sub-Saharan African population and the 1000 Genomes Super African population. This infers that 1 out of 3 allele frequencies observed will be different between the populations. Further studies are required to confirm if this observation is true across the human genome. Therefore the use of ethnically matched populations for genetic studies done in Africa is extremely important.

4.1.4.3. Limitations to interpreting experimental data

Through the data generated in this study, two alleles of interest, but with unknown significance, were flagged: one in *BRCA1* and one in *BRCA2*. However, due to the very small sample size of both the BRCA cohort and the Control cohort, there was no confidence in interpreting the statistical association as true associations since this study has no power. In order to identify true statistical associations, a larger sample size is needed.

The impact and effects of 19 known SNVs were not statistically evaluated as these SNVs were not screened in the Control group. Therefore it cannot be determined whether any of these SNVs are important determining risk factors for the development of breast cancer.

Sporadic breast cancer is a multifactorial disease that is genetically heterogeneous. Recently, a study has identified numerous disease-risk associated alleles across the genome that would predispose individuals to the development of breast cancer (Michailidou et al., 2013). Michailidou, et al. (2013) reinforces the notion that genetic predispositions (here evaluated) exist in the general population and having a combination of a set of risk alleles would greatly increase one's chances of developing

breast cancer. Since this research was conducted using individuals of European ancestry, the degree of extrapolation of the data when evaluating individuals from an African ancestry, especially sub-Saharan African ancestry, is limited.

Having one risk allele only does not necessarily confer sufficient risk or even change the risk profile that would drive cancer formation, but having multiple risk alleles may certainly increase the chances of cancer formation (Michailidou et al., 2013). It is not known what values the risk alleles identified in this study have in predicting the overall risk profile of an individual. The role these risk alleles have on the penetrance are also unclear. It is likely that the combinatory effect of all the risk factors may play a role in disease causality in both sporadic and hereditary breast cancer but no conclusions can be drawn from this study to support this thinking. Nevertheless, the findings from this study warrant further, large scale, investigation into the genetic risk profiles of African black individuals as within the small BRCA cohort and Control cohort, signals of genotype associated risk may be evident.

4.1.4.4. Co-segregated alleles identified in the BRCA1 & BRCA2 gene

The large amount of SNV data generated by May et al. (2013) was useful in screening for common patterns of co-segregation of alleles with the *BRCA* genes. Informative alleles identified in both *BRCA1* & *BRCA2* genes were used to compile common patterns of allele co-segregation for the two genes. Due to the relatively small size of the *BRCA* genes (in comparison to whole chromosomes), it is unlikely that meiotic recombination would have occurred within the *BRCA* genes. Identifying common patterns of co-segregated alleles may allow researchers to track the inheritance and origin of the genetic background on which the *BRCA* genes are located. Another reason for the addition of this component to the study was to determine whether it would be possible to identify common breast cancer predisposing haplotypes from the SNV data. It was impossible to ascertain the phase of the alleles thus it was impossible to determine the haplotypes present within the *BRCA* genes. Only common patterns of allele co-segregation were determined from the SNV data.

Informative alleles were present in both *BRCA* genes within the Control group (Table 3.10a.). Common patterns of allele co-segregation were present in both genes. Four common co-segregation patterns were present in *BRCA1* and 6 patterns were present in *BRCA2*. This suggests the possible origins of the *BRCA* gene sequences are diverse, and supports the notion that the African populations are ancient and therefore genetically diversity in nature (May et al., 2013).

Being able to identify common co-segregation patterns allows for the tracking of the “disease” gene through a family, thus allowing the identification of potential disease risk without identifying the actual causative genetic mutation. However, caution needs to be applied when this method is used

for diagnostic testing, as it is entirely possible that the causative mutation is totally unrelated to the *BRCA* genes and is on separate chromosomes.

Not all the informative alleles used to identify common co-segregation patterns in the Control group were present in the BRCA cohort. As a result the co-segregation pattern was not evaluated in the BRCA cohort and the due to the lack of common allele data between the Control group and the BRCA cohort, these two cohorts' data were not overlaid and compared. It may be interesting to determine if the co-segregation patterns are presenting in the BRCA cohort as this may be useful to track the "disease" or high-risk chromosome.

4.1.5. MLPA results

The last component of our study was to evaluate whether large gene or exon deletions and /or duplications contributed to the causality of breast cancer in this BRCA cohort. Large gene or exon deletions and/or duplications account for less than 10% of all *BCRA1* mutations and in *BRCA2,* large rearrangements account for an even smaller percentage of mutations (Engert et al., 2008b). However, for the sake of completeness, this study aimed to rule out large gene rearrangements as the causative mutation within the BRCA cohort, therefore the presence of large gene/exon deletions and/or duplication of/in *BRCA1* and *BRCA2* were investigated in the BRCA cohort using MLPA analysis.

No large gene/exon alterations of *BRCA1* and *BRCA2* were found amongst the BRCA cohort. This supports the findings available on most other populations in the literature. Since our cohort was comprised of only 33 individuals, it was not unexpected that no large gene /exon alterations were identified (Engert et al., 2008b).

4.2. Study limitations

This study was subjected to a number of limitations and confounding factors. This section will discuss these limitations and their potential impact on the outcomes and conclusions drawn from our results.

4.2.1. Cohort size

The first limitation is the cohort size of the BRCA cohort. Due to financial constraints, only 33 patients were screened for *BRCA* gene mutations. The high cost of Sanger sequencing makes screening large cohorts unfeasible. A next generation sequencing platform was not available in our laboratory at the time of the study.

One of the objectives of this study was to evaluate whether possible founder mutations existed within the Sub-Saharan African black breast cancer patients. The small cohort size made evaluating

founder mutations extremely difficult. Three non-recurring mutations in *BRCA1* and *BRCA2* contributed to 10% of disease causality within our BRCA cohort. This study was unable to classify any of the 3 mutations found in this study as either a founder mutation or a family specific mutation or a sporadic mutation.

Small cohort sizes also greatly limit the strength of the statistical analysis performed. The Sub-Saharan African Control cohort comprised 96 individuals, and the BRCA cohort sample compromised of 33 individuals, leading to the statistical analysis having no power. The significant statistical associations cannot be deemed as true associations but as mere speculations and evidence that would promote further investigation using much larger case verses control cohort studies.

4.2.2. Number of genes studied

The genetic heterogeneity of breast cancer is well established (Andersen, 1996; Easton, et al., 2007b; Walsh & King, 2007). By screening only two genes, other genetic causes such as genetic mutations in *TP53*, *CHEK2*, *RAD51*, *RAD50* and *PALB2*, just to name a few, cannot be ruled out (Walsh & King, 2007).

To gain a more comprehensive understanding of the genetic contributions to breast cancer, next generation sequencing technology can be employed to screen the exome or genome of the patients for mutations that may be causative for hereditary breast cancer.

4.2.3. Sequencing only the coding regions

In this study, only the coding regions and intron/exon boundaries of *BRCA1* and *BRCA2* were sequenced. The contribution to disease causality of mutations in the non-coding regions (deep intronic regions, 5' UTRs and 3'UTR) or the epigenetic modifications potentially present in either of the *BRCA* genes were not investigated.

Regulatory regions and deep intronic regions may have an effect on the expression of the gene. Alternatively, deep intronic cryptic splice sites may also be present within the gene. However the percentage of overall mutations in these regions is likely to be very low in other populations.

4.2.4. Novel SNVs and unscreened known SNPs

A total of 5 novel SNVs' and 19 known SNVs' were not screened in addition to the allelic data provided by May et al. (2013) in the sub- Saharan African Control group so their allelic frequencies in the general population could not be determined. This is largely due to financial constraints.

No phenotypic data were available both the Control group and the 1Kg group. This limits the usefulness of the genotype data as a comparison for disease risk studies. In addition, the Control group were not age-matched or gender-matched to the BRCA cohort.

In Silico analysis suggested that these variants were not of significant functional importance but without population frequency data, the pathogenicity of these variants cannot be completely ruled out.

These variants would be included as a part of a larger panel of variants to be screened in a future study that would look at the genome-wide association of risk alleles in breast cancer.

4.2.5. Environmental risk factors and genetic risk factors

This study investigated the possible genetic causal and risk factors within the *BRCA1* and *BRCA2* genes within the BRCA cohort. However, environmental risk factors contributing to disease causality were not investigated so their contribution to disease causality cannot be ruled out. Both environmental risk factors and genetic risk factors may play a role in disease causality. It would have been interesting to investigate the environmental risk factors in conjunction with genetic risk factors.

The low mean age of disease onset (39 years) within the BRCA cohort may be a reflection of the population structure within South Africa and not predisposition to disease. However, numerous studies have suggested that African populations are more predisposed to developing breast cancer and that population structures present in African countries may be masking the predisposition (Anders et al., 2009; Akarolo-Anthony et al., 2010; Amirikia et al., 2011). This study was not able to address the contribution of environmental risks to the development of breast or whether genetic susceptibility is the main contributor to disease.

Environmental risk factor data (such as lifestyle data) for the BRCA cohort were not available. Therefore, environmental risk's contribution to disease development was not evaluated. The cohort selection was based on risk calculated by considering age, tumour grade, tumour characteristics and family history. This was done to minimize the impact of environmental risk factors contributing to disease and to enrich possible existing genetic risk factors. However, environmental risks cannot be ruled out completely.

4.3. Future prospects

This study indicated that *BRCA* mutations contribute to causality of disease in approximately 10% of cases within this high risk breast cancer cohort. The causes of disease for the remaining 90% of breast cancer cases are still unknown.

This study proposes two study models to further investigate the genetic aetiology of breast cancer in South African black individuals:

4.3.1. Next generation sequencing

With the advances in DNA sequencing technologies in the recent years, the full genome or exome of cancer patients can be sequenced at relatively low cost. The genome and/or exome sequencing of these cancer patients would enable researchers to investigate more fully the genetic causalities of disease.

Exome sequencing may be a good starting point. Exome sequencing is much more cost effective than whole genome sequencing and the volume of data generated would be more manageable than whole genome sequence data. This study proposes, within reasonable financial considerations, sequencing the exomes of 100 high risk breast cancer patients to identify genetic mutations that may be causative for breast cancer. A recent publication indicated good mutation detection rates (up to 27% of all cases) for picking-up of genetic mutations causative for ovarian cancer using NGS (Walsh et al., 2010). A similar strategy can be applied in the current scenario to look for predisposing mutations implicated in the development of breast cancer in young black South African patients.

Cancer gene panels are also a viable option to explore. In comparison to exome sequencing, cancer gene panels are designed to sequence known cancer genes. This approach will be more targeted thus reduce the burden of analysis, and at a reduced cost. The BROCA cancer gene panel is one such example. This panel sequences the exomic region and exon-flanking regions of 60 known cancer causing genes (http://tests.labmed.washington.edu/BROCA).

4.3.2. GWAS

A large scale genome wide association study published recently has identified numerous risk associated alleles and disease susceptibility alleles for breast cancer (Michailidou et al., 2013). This large scale study was conducted on Caucasian populations. A similar study identified very different susceptibility and risk associated alleles for the development of breast cancers in West African black populations (Huo et al., 2012). The presentation of different risk and susceptibility alleles across different ethnic populations suggests that population ethnicity plays a very important role in determining disease susceptibility.

Genetic diversity amongst African populations is well established. The genetic susceptibility loci within the sub-Saharan African population may well be different to those of the West African population. A Genome Wide Association Study may shed light onto the genetic susceptibility loci that may be present in the Sub-Saharan African population for the development of breast cancer.

Again, within reasonable financial considerations, a cohort of breast cancer patients should be genotyped using a genome wide SNP array to identity susceptibility loci, which would then require follow up on any significant loci identified with a larger cohort.

5. CONCLUSION

This study found that genetic mutations within the *BRCA1* & *BRCA2* genes cause approximately 10% (no confidence interval can be given due to the small sample size) of the breast cancer cases within this high risk, early onset South African black cohort. The percentage of *BRCA* mutations' contribution to the development of breast cancer in this black cohort is not more or less than the percentages reported for other ethnic groups studied worldwide to date. No founder mutation was identified in this cohort, although the sample size tested was small.

The results from this study will immediately benefit the 10% of the cohort for whom a causative mutation has been identified. Their at-risk family members can undergo targeted genetic mutation screening to determine whether they are *BRCA* mutation carriers. The mutation positive patients can benefit from more stringent and proactive surveillance programs and possibly prophylactic surgery.

The sub-Saharan black populations are genetically diverse and distinct. The use of ethnically matched control groups for population allele frequency comparison based studies is essential. This allows for the accurate comparison between case group versus the control group. The use of Western African, Central African or African American black populations as the reference populations for South African black populations in genetic studies is not ideal and may often yield inaccurate data.

This study looked at a narrow aspect of the genetic aetiology of high risk, early onset breast cancer. The causality of breast cancer in the remaining 90% of this cohort remains to be determined. The genetic aetiology of high risk, early onset breast cancer appears to be complex and diverse. A traditional candidate gene, targeted approach may not be effective in uncovering the complete genetic contribution to disease. New technologies allowing for the sequencing of whole genomes and exomes, at relatively low cost, would yield data that are much more informative and useful.

This study has set a foundation for future genetic studies that will investigate the genetic causality of early onset breast cancer in South African black individuals. Breast cancer is the most prevalent cancer in South Africa, and the second most prevalent cancer in the world – a better understanding of its aetiology would greatly benefit the population at large.

6. LITERATURE CITED

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

Adank, M. A., Jonker, M. A., Kluijt, I., van Mil, S. E., Oldenburg, R. A., Mooi, W. J., … Waisfisz, Q. (2011). CHEK2*1100delC homozygosity is associated with a high breast cancer risk in women. *Journal of Medical Genetics*, *48*(12), 860–3.

Adzhubei, I. A., Schmidt, S., Peshkin, L., Ramensky, V. E., Gerasimova, A., Bork, P., … Sunyaev, S. R. (2010). A method and server for predicting damaging missense mutations. *Nature Methods*, *7*(4), 248–9.

Akarolo-Anthony, S. N., Ogundiran, T. O., & Adebamowo, C. A. (2010). Emerging breast cancer epidemic: evidence from Africa. *Breast Cancer Research : BCR*, *12 Suppl 4*, S8.

Amirikia, K. C., Mills, P., Bush, J., & Newman, L. A. (2011). Higher population-based incidence rates of triple-negative breast cancer among young African-American women : Implications for breast cancer screening recommendations. *Cancer*, *117*(12), 2747–53.

Anand, P., Kunnumakkara, A. B., Kunnumakara, A. B., Sundaram, C., Harikumar, K. B., Tharakan, S. T., … Aggarwal, B. B. (2008). Cancer is a preventable disease that requires major lifestyle changes. *Pharmaceutical Research*, *25*(9), 2097–116.

Anczuków, O., Buisson, M., Léoné, M., Coutanson, C., Lasset, C., Calender, A., … Mazoyer, S. (2012). BRCA2 deep intronic mutation causing activation of a cryptic exon: opening toward a new preventive therapeutic strategy. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *18*(18), 4903–9.

Anders, C. K., Johnson, R., Litton, J., Phillips, M., & Bleyer, A. (2009). Breast cancer before age 40 years. *Seminars in Oncology*, *36*(3), 237–49.

Andersen, T. I. (1996). Genetic heterogeneity in breast cancer susceptibility. *Acta Oncologica (Stockholm, Sweden)*, *35*(4), 407–10.

Antoniou, A., Pharoah, P. D. P., Narod, S., Risch, H. A., Eyfjord, J. E., Hopper, J. L., … Easton, D. F. (2003). Average risks of breast and ovarian cancer associated with BRCA1 or BRCA2 mutations detected in case Series unselected for family history: a combined analysis of 22 studies. *American Journal of Human Genetics*, *72*(5), 1117–30.

Badie, S., Escandell, J. M., Bouwman, P., Carlos, A. R., Thanasoula, M., Gallardo, M. M., … Tarsounas, M. (2010). BRCA2 acts as a RAD51 loader to facilitate telomere replication and capping. *Nature Structural & Molecular Biology*, *17*(12), 1461–9.

Bartlett, J. M. S., & Stirling, D. (2003). A short history of the polymerase chain reaction. *Methods in Molecular Biology (Clifton, N.J.)*, *226*, 3–6.

Bolton, K. L., Chenevix-Trench, G., Goh, C., Sadetzki, S., Ramus, S. J., Karlan, B. Y., … Pharoah, P. D. P. (2012). Association between BRCA1 and BRCA2 mutations and survival in women with invasive epithelial ovarian cancer. *JAMA : The Journal of the American Medical Association*, *307*(4), 382–90.

Boulton, S. J. (2006). Cellular functions of the BRCA tumour-suppressor proteins. *Biochemical Society Transactions*, *34*(Pt 5), 633–45.

Carey, L. A., Perou, C. M., Livasy, C. A., Dressler, L. G., Cowan, D., Conway, K., … Millikan, R. C. (2006). Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. *JAMA : The Journal of the American Medical Association*, *295*(21), 2492–502.

Census 2011. (2012). Census 2011 Statistical release – P0301.4. Statistics South Africa. Pretoria, South Africa.

Chang, Y.-F., Imam, J. S., & Wilkinson, M. F. (2007). The nonsense-mediated decay RNA surveillance pathway. *Annual Review of Biochemistry*, *76*, 51–74.

Clark, S. L., Rodriguez, A. M., Snyder, R. R., Hankins, G. D. V, & Boehning, D. (2012). Structure-Function Of The Tumor Suppressor BRCA1. *Computational and Structural Biotechnology Journal*, *1*(1).

Cousineau, I., & Belmaaza, A. (2007). BRCA1 haploinsufficiency, but not heterozygosity for a BRCA1-truncating mutation, deregulates homologous recombination. *Cell Cycle (Georgetown, Tex.)*, *6*(8), 962–71.

Desmet, F.-O., Hamroun, D., Lalande, M., Collod-Béroud, G., Claustres, M., & Béroud, C. (2009). Human Splicing Finder: an online bioinformatics tool to predict splicing signals. *Nucleic Acids Research*, *37*(9), e67.

Easton, D. F., Deffenbaugh, A. M., Pruss, D., Frye, C., Wenstrup, R. J., Allen-Brady, K., … Goldgar, D. E. (2007a). A systematic genetic assessment of 1,433 sequence variants of unknown clinical significance in the BRCA1 and BRCA2 breast cancer-predisposition genes. *American Journal of Human Genetics*, *81*(5), 873–83.

Easton, D. F., Pooley, K. A., Dunning, A. M., Pharoah, P. D. P., Thompson, D., Ballinger, D. G., … Ponder, B. A. J. (2007b). Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, *447*(7148), 1087–93.

Engert, S., Wappenschmidt, B., Betz, B., Kast, K., Kutsche, M., Hellebrand, H., … Meindl, A. (2008a). MLPA screening in the BRCA1 gene from 1,506 German hereditary breast cancer cases: novel deletions, frequent involvement of exon 17, and occurrence in single early-onset cases. *Human Mutation*, *29*(7), 948–58.

Evans, D. G., Shenton, A., Woodward, E., Lalloo, F., Howell, A., & Maher, E. R. (2008). Penetrance estimates for BRCA1 and BRCA2 based on genetic testing in a Clinical Cancer Genetics service setting: risks of breast/ovarian cancer quoted should reflect the cancer burden in the family. *BMC Cancer*, *8*, 155. doi:10.1186/1471-2407-8-155

Fackenthal, J. D., Zhang, J., Zhang, B., Zheng, Y., Hagos, F., Burrill, D. R., … Olopade, O. I. (2012). High prevalence of BRCA1 and BRCA2 mutations in unselected Nigerian breast cancer patients. *International Journal of Cancer. Journal International Du Cancer*, *131*(5), 1114–23.

Feliubadalo, L., Lopez-Doriga, A., Castellsague, E., del Valle, J., Menedez, M., Tornero, E., … Lazaro, C. (2013). Next-generation sequencing meets genetic diagnostics: development of a comprehensive workflow for the analysis of *BRCA1* and *BRCA2* genes. *European Journal of Human Genetics, 21,* 864-870.

Ferlay, J., Shin, H., Bray, F., Forman, D., Mathers, C., & Parkin, D. (2010). GLOBOCAN 2008 v.1.2, Cancer Incidence and Mortality Worldwide: IARC CancerBase No.10 [Internet]. Lyon, France: International Agency for Research on Cancer.

Fijnerman, R. J. (2005). Genetic predisposition to sporadic cancer: how to handle major effects of minor genes? *Cell Oncology, 27*(5-6), 281-292.

Fokkema, I. F. A. C., Taschner, P. E. M., Schaafsma, G. C. P., Celli, J., Laros, J. F. J., & den Dunnen, J. T. (2011). LOVD v.2.0: the next generation in gene variant databases. *Human Mutation*, *32*(5), 557–63.

Frankenberg-Schwager, M., Becker, M., Garg, I., Pralle, E., Wolf, H. & Frankenburg, D. (2008). The role of nonhomologous DNA end joining , conservative homologous recombination, and single-strand annealing in the cell cycle-dependent repair of DNA double-strand breaks induced by H(2)O(2) in mammalian cells. *Radiation Research, 170*(6), 784-793.

Gao, Q., Neuhausen, S., Cummings, S., Luce, M., & Olopade, O. I. (1997). Recurrent germ-line BRCA1 mutations in extended African American families with early-onset breast cancer. *American Journal of Human Genetics*, *60*(5), 1233–6.

Gonzalez-Angulo, A. M., Timms, K. M., Liu, S., Chen, H., Litton, J. K., Potter, J., … Meric-Bernstam, F. (2011). Incidence and outcome of BRCA mutations in unselected patients with triple receptor-negative breast cancer. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *17*(5), 1082–9.

Gracia-Aznarez, F. J., Fernandez, V., Pita, G., Peterlongo, P., Dominguez, O., de la Hoya, M., … Benitez, J. (2013). Whole exome sequencing suggests much of non-BRCA1/BRCA2 familial breast cancer is due to moderate and low penetrance susceptibility alleles. *PloS One*, *8*(2), e55681.

Gronwald, J., Tung, N., Foulkes, W. D., Offit, K., Gershoni, R., Daly, M., … Narod, S. A. (2006). Tamoxifen and contralateral breast cancer in BRCA1 and BRCA2 carriers: an update. *International Journal of Cancer. Journal International Du Cancer*, *118*(9), 2281–4.

Group, A. B. C. S. (2000). Prevalence and penetrance of BRCA1 and BRCA2 mutations in a population-based series of breast cancer cases. Anglian Breast Cancer Study Group. *British Journal of Cancer*, *83*(10), 1301–8.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., … Sandhu, M. S. (2015). The African Genome Variation Project shapes medical genetics in Africa. *Nature, 517,* 327-332.

Haffty, B. G., Silber, A., Matloff, E., Chung, J., & Lannin, D. (2006). Racial differences in the incidence of BRCA1 and BRCA2 mutations in a cohort of early onset breast cancer patients: African American compared to white women. *Journal of Medical Genetics*, *43*(2), 133–7.

Hall, J. M., Lee, M. K., Newman, B., Morrow, J. E., Anderson, L. A., Huey, B., & King, M. C. (1990). Linkage of early-onset familial breast cancer to chromosome 17q21. *Science (New York, N.Y.)*, *250*(4988), 1684–9.

Han, S.-A., Kim, S.-W., Kang, E., Park, S. K., Ahn, S.-H., Lee, M. H., … Jung, S. H. (2013). The prevalence of BRCA mutations among familial breast cancer patients in Korea: results of the Korean Hereditary Breast Cancer study. *Familial Cancer*, *12*(1), 75–81.

Hogervorst, F. B. L., Nederlof, P. M., Gille, J. J. P., McElgunn, C. J., Grippeling, M., Pruntel, R., … Pals, G. (2003). Large genomic deletions and duplications in the BRCA1 gene identified by a novel quantitative method. *Cancer Research*, *63*(7), 1449–53.

Howlett, N. G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., De Die-Smulders, C., … D'Andrea, A. D. (2002). Biallelic inactivation of BRCA2 in Fanconi anemia. *Science (New York, N.Y.)*, *297*(5581), 606–9.

Huo, D., Zheng, Y., Ogundiran, T. O., Adebamowo, C., Nathanson, K. L., Domchek, S. M., … Olopade, O. I. (2012). Evaluation of 19 susceptibility loci of breast cancer in women of African ancestry. *Carcinogenesis*, *33*(4), 835–40.

Jensen, R. A., Thompson, M. E., Jetton, T. L., Szabo, C. I., van der Meer, R., Helou, B., … Holt, J. T. (1996). BRCA1 is secreted and exhibits properties of a granin. *Nature Genetics*, *12*(3), 303–8.

John, E. M., Miron, A., Gong, G., Phipps, A. I., Felberg, A., Li, F. P., … Whittemore, A. S. (2007). Prevalence of pathogenic BRCA1 mutation carriers in 5 US racial/ethnic groups. *JAMA : The Journal of the American Medical Association*, *298*(24), 2869–76.

Karchin, R., Agarwal, M., Sali, A., Couch, F., & Beattie, M. S. (2008). Classifying Variants of Undetermined Significance in BRCA2 with protein likelihood ratios. *Cancer Informatics*, *6*, 203–16.

King, T. A., Li, W., Brogi, E., Yee, C. J., Gemignani, M. L., Olvera, N., … Boyd, J. (2007). Heterogenic loss of the wild-type BRCA allele in human breast tumorigenesis. *Annals of Surgical Oncology*, *14*(9), 2510–8.

Kinzler, K. W., & Vogelstein, B. (1997). Cancer-susceptibility genes. Gatekeepers and caretakers. *Nature*, *386*(6627), 761, 763.

Knudson, A. G. (1971). Mutation and cancer: statistical study of retinoblastoma. *Proceedings of the National Academy of Sciences of the United States of America*, *68*(4), 820–3.

Kote-Jarai, Z., Leongamornlert, D., Saunders, E., Tymrakiewicz, M., Castro, E., Mahmud, N., … Eeles, R. (2011a). BRCA2 is a moderate penetrance gene contributing to young-onset prostate cancer: implications for genetic testing in prostate cancer patients. *British Journal of Cancer*, *105*(8), 1230–4.

Kuusisto, K. M., Bebel, A., Vihinen, M., Schleutker, J., & Sallinen, S.-L. (2011). Screening for BRCA1, BRCA2, CHEK2, PALB2, BRIP1, RAD50, and CDH1 mutations in high-risk Finnish BRCA1/2-founder mutation-negative breast and/or ovarian cancer individuals. *Breast Cancer Research : BCR*, *13*(1), R20.

Lacroix, M., & Leclercq, G. (2005). The "portrait" of hereditary breast cancer. *Breast Cancer Research and Treatment*, *89*(3), 297–304.

Lathrop, G. M. (1983). Estimating genotype relative risks. *Tissue Antigens*, *22*(2), 160–6.

Lee, E., McKean-Cowdin, R., Ma, H., Spicer, D. V, Van Den Berg, D., Bernstein, L., & Ursin, G. (2011). Characteristics of triple-negative breast cancer in patients with a BRCA1 mutation: results from a population-based study of young women. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *29*(33), 4373–80.

Liu, Y. & West, S. C. (2002). Distinct functions of BRCA1 and BRCA2 in double-strand break repair. *Breast Cancer Research, 4*(1), 9-13.

Lu, H., Guo, X., Meng, X., Liu, J., Allen, C., Wray, J., Nickoloff, J. A. & Shen, Z. (2005). The BRCA2-interacting protein BCCIP functions in RAD51 and BRCA2 focus formation and homologous recombination repair. *Molecular Cell Biology, 25*(5), 1949-1957.

Mattocks, C. J., Watkins, G., Ward, D., Janssens, T., Bosgoed, E. A. J., van der Donk, K., … Matthijs, G. (2010). Interlaboratory diagnostic validation of conformation-sensitive capillary electrophoresis for mutation scanning. *Clinical Chemistry*, *56*(4), 593–602.

May, A., Hazelhurst, S., Li, Y., Norris, S. A., Govind, N., Tikly, M., … Ramsay, M. (2013). Genetic diversity in black South Africans from Soweto. *BMC Genomics*, *14*, 644.

McLaren, W., Pritchard, B., Rios, D., Chen, Y., Flicek, P., & Cunningham, F. (2010). Deriving the consequences of genomic variants with the Ensembl API and SNP Effect Predictor. *Bioinformatics (Oxford, England)*, *26*(16), 2069–70.

Mefford, H. C., Baumbach, L., Panguluri, R. C., Whitfield-Broome, C., Szabo, C., Smith, S., … Arena, F. (1999). Evidence for a BRCA1 founder mutation in families of West African ancestry. *American Journal of Human Genetics*, *65*(2), 575–8.

Metcalfe, K., Lynch, H. T., Ghadirian, P., Tung, N., Olivotto, I., Warner, E., … Narod, S. A. (2004). Contralateral breast cancer in BRCA1 and BRCA2 mutation carriers. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *22*(12), 2328–35.

Michailidou, K., Hall, P., Gonzalez-Neira, A., Ghoussaini, M., Dennis, J., Milne, R. L., … Easton, D. F. (2013). Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*, *45*(4), 353–61, 361e1–2.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., … Ding, W. (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. *Science (New York, N.Y.)*, *266*(5182), 66–71.

Miller, S. A., Dykes, D. D., & Polesky, H. F. (1988). A simple salting out procedure for extracting DNA from human nucleated cells. *Nucleic Acids Research*, *16*(3), 1215–1215.

Moynahan, M. E. (2002). The cancer connection: BRCA1 and BRCA2 tumor suppression in mice and humans. *Oncogene*, *21*(58), 8994–9007.

Neuhausen, S. L. (2000). Founder populations and their uses for breast cancer genetics. *Breast Cancer Research : BCR*, *2*(2), 77–81.

Ng, P. C., & Henikoff, S. (2001). Predicting deleterious amino acid substitutions. *Genome Research*, *11*(5), 863–74.

Pardo, B., Gomez-Gonzalez, B. & Aquilera, A. (2009). DNA rrepair in mammalian cells: DNA double-strand break repair: how to fix a broken relationship. *Cellular and Molecular Life Sciences, 66*(6), 1039-1056.

Pertea, M., Lin, X., & Salzberg, S. L. (2001). GeneSplicer: a new computational method for splice site prediction. *Nucleic Acids Research*, *29*(5), 1185–90.

Peto, J., Collins, N., Barfoot, R., Seal, S., Warren, W., Rahman, N., … Stratton, M. R. (1999). Prevalence of BRCA1 and BRCA2 Gene Mutations in Patients With Early-Onset Breast Cancer. *JNCI Journal of the National Cancer Institute*, *91*(11), 943–949.

Reese, M. G., Eeckman, F. H., Kulp, D., & Haussler, D. (1997). Improved splice site detection in Genie. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *4*(3), 311–23.

Reeves, M. D., Yawitch, T. M., van der Merwe, N. C., van den Berg, H. J., Dreyer, G., & van Rensburg, E. J. (2004). BRCA1 mutations in South African breast and/or ovarian cancer families: evidence of a novel founder mutation in Afrikaner families. *International Journal of Cancer. Journal International Du Cancer*, *110*(5), 677–82.

Reiner, A. S., John, E. M., Brooks, J. D., Lynch, C. F., Bernstein, L., Mellemkjær, L., … Bernstein, J. L. (2013). Risk of asynchronous contralateral breast cancer in noncarriers of BRCA1 and BRCA2 mutations with a family history of breast cancer: a report from the Women's Environmental Cancer and Radiation Epidemiology Study. *Journal of Clinical Oncology : Official Journal of the American Society of Clinical Oncology*, *31*(4), 433–9.

Richter, L., Norris, S., Pettifor, J., Yach, D. & Cameron, N. (2007). Cohort profile: Mandela's children: The 1990 birth to twenty study in South Africa. *International Journal of Epidemiology, 36*(3), 504-511.

Roukos, D.H. & Briasoulis, E. (2007). Individualized preventive and therapeutic management of hereditary breast ovarian cancer syndrome. *Nature Clinical Practice: Oncology,* 4(10), 578-90

Ruiz de Garibay, G., Gutiérrez-Enríquez, S., Garre, P., Bonache, S., Romero, A., Palomo, L., … de la Hoya, M. (2012). Characterization of four novel BRCA2 large genomic rearrangements in Spanish breast/ovarian cancer families: review of the literature, and reevaluation of the genetic mechanisms involved in their origin. *Breast Cancer Research and Treatment*, *133*(1), 273–83.

Sanger, F., Nicklen, S., & Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences of the United States of America*, *74*(12), 5463–7.

Sanz, D. J., Acedo, A., Infante, M., Durán, M., Pérez-Cabornero, L., Esteban-Cardeñosa, E., … Velasco, E. A. (2010). A high proportion of DNA variants of BRCA1 and BRCA2 is associated with aberrant splicing in breast/ovarian cancer patients. *Clinical Cancer Research : An Official Journal of the American Association for Cancer Research*, *16*(6), 1957–67.

Sharan, S. K. & Bradley, A. (1998). Functional characterization of BRCA1 and BRCA2: clues from their interacting proteins. *Journal of Mammary gland and Neoplasia, 3*(4), 413-421.

Schork, N. J., Murray, S. S., Frazer, K. A., & Topol, E. J. (2009). Common vs. rare allele hypotheses for complex diseases. *Current Opinion in Genetics & Development*, *19*(3), 212–9.

Shrivastav, M., De Haro, L. P., & Nickoloff, J. A. (2008). Regulation of DNA double-strand break repair pathway choice. *Cell Research, 18,* 134-147.

Schuster, S. C., Miller, W., Ratan, A., Tomsho, L. P., Giardine, B., Kasson, L. R., … Hayes, V. M. (2010). Complete Khoisan and Bantu genomes from southern Africa. *Nature*, *463*(7283), 943–7.

Schwarz, J. M., Rödelsperger, C., Schuelke, M., & Seelow, D. (2010). MutationTaster evaluates disease-causing potential of sequence alterations. *Nature Methods*, *7*(8), 575–6.

Singh, K. K., Shukla, P. C., Quan, A., Desjardins, J.-F., Lovren, F., Pan, Y., … Verma, S. (2012). BRCA2 protein deficiency exaggerates doxorubicin-induced cardiomyocyte apoptosis and cardiac failure. *The Journal of Biological Chemistry*, *287*(9), 6604–14.

Szabo, C., Masiello, A., Ryan, J. F., & Brody, L. C. (2000). The breast cancer information core: database design, structure, and scope. *Human Mutation*, *16*(2), 123–31.

Tavtigian, S. V, Deffenbaugh, A. M., Yin, L., Judkins, T., Scholl, T., Samollow, P. B., … Thomas, A. (2006). Comprehensive statistical study of 452 BRCA1 missense substitutions with classification of eight recurrent substitutions as neutral. *Journal of Medical Genetics*, *43*(4), 295–305.

Tavtigian, S. V, Simard, J., Rommens, J., Couch, F., Shattuck-Eidens, D., Neuhausen, S., … Goldgar, D. E. (1996). The complete BRCA2 gene and mutations in chromosome 13q-linked kindreds. *Nature Genetics*, *12*(3), 333–7.

Thanaraj, T. A., & Clark, F. (2001). Human GC-AG alternative intron isoforms with weak donor sites show enhanced consensus at acceptor exon positions. *Nucleic Acids Research*, *29*(12), 2581–93.

Turner, N., Tutt, A. & Ashworth, A. (2004). Hallmarks of 'BRCAness' in sporadic cancers. *Nature Review Genetics, 4*(10), 814-819.

Van der Groep, P., Bouter, A., van der Zanden, R., Siccama, I., Menko, F. H., van Kalken, C., … van Diest, P. J. (2006). Distinction between hereditary and sporadic breast cancer on the basis of clinicopathological data. *Journal of Clinical Pathology, 59*(6), 611-617.

Van der Groep, P., van der Wall, E., & van Diest, P. J. (2011). Pathology of hereditary breast cancer. *Cellular Oncology (Dordrecht)*, *34*(2), 71–88.

Van der Merwe, N. C., Hamel, N., Schneider, S.-R., Apffelstaedt, J. P., Wijnen, J. T., & Foulkes, W. D. (2012). A founder BRCA2 mutation in non-Afrikaner breast cancer patients of the Western Cape of South Africa. *Clinical Genetics*, *81*(2), 179–84.

Venkitaraman, A. R. (2004). Tracing the network connecting BRCA and Fanconi anaemia proteins. *Nature Reviews. Cancer*, *4*(4), 266–76.

Walsh, T., & King, M.-C. (2007). Ten genes for inherited breast cancer. *Cancer Cell*, *11*(2), 103–5.

Walsh, T., Lee, M. K., Casadei, S., Thornton, A. M., Stray, S. M., Pennil, C., … King, M.-C. (2010). Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(28), 12629–33.

Wainstein, T. (2011). Family History and Risk Assessment in Black South African Women with Breast Cancer. MSc Dissertation. Univerisity of Witwatersrand, Johannesburg, South Africa.

Wang, Y., Cortez, D., Tazdi, P., Neff, N., Elledge, S.J., & Qin, J. (2000). BASC, a super complex of BRCA1-associated proteins involved in the recognition and repair of aberrant DNA structures. *Genes & Development, 14*(8), 927-39.

Wappenschmidt, B., Becker, A. A., Hauke, J., Weber, U., Engert, S., Köhler, J., … Schmutzler, R. K. (2012). Analysis of 30 putative BRCA1 splicing mutations in hereditary breast and ovarian cancer families identifies exonic splice site mutations that escape in silico prediction. *PloS One*, *7*(12), e50800.

Weinberg, R.A. (2014). The biology of Cancer, second edition. Garland Science, Taylor & Francis Group, New York, USA. 553-55

Welcsh, P. L., & King, M. C. (2001). BRCA1 and BRCA2 and the genetics of breast and ovarian cancer. *Human Molecular Genetics*, *10*(7), 705–13.

Wildeman, M., van Ophuizen, E., den Dunnen, J. T., & Taschner, P. E. M. (2008). Improving sequence variant descriptions in mutation databases and literature using the Mutalyzer sequence variation nomenclature checker. *Human Mutation*, *29*(1), 6–13.

Woodward, A. M., Davis, T. A., Silva, A. G. S., Kirk, J. A., & Leary, J. A. (2005). Large genomic rearrangements of both BRCA2 and BRCA1 are a feature of the inherited breast/ovarian cancer phenotype in selected families. *Journal of Medical Genetics*, *42*(5), e31.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., et al. (1995). Identification of the breast cancer susceptibility gene *BRCA*2. *Nature,* 378, 789-792.

Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., … Livingston, D. M. (2006). Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Molecular Cell*, *22*(6), 719–29.

Yates, F. (1934). Contingency Tables Involving Small Numbers and the χ 2 Test. *Supplement to the Journal of the Royal Statistical Society*, *1*(2), 217.

Yeo, G., & Burge, C. B. (2004). Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *Journal of Computational Biology : A Journal of Computational Molecular Cell Biology*, *11*(2-3), 377–94.

Yoshida, K. & Miki, Y. (2004). Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage. *Cancer Science, 95*(11), 866-871.

Zhang, J., Fackenthal, J. D., Zheng, Y., Huo, D., Hou, N., Niu, Q., … Olopade, O. I. (2012). Recurrent BRCA1 and BRCA2 mutations in breast cancer patients of African ancestry. *Breast Cancer Research and Treatment*, *134*(2), 889–94.

6.1. Electronic resources

Cancer Research UK. (2013). Breast Cancer. Retrieved on June, 01, 2015, from http://www.cancerresearchuk.org/about-cancer/type/breast-cancer/

CANSA. (2014). South African Cancer Statistics. Retrieved on June, 01, 2015, from http://www.cansa.org.za/south-african-cancer-statistics/

DecodeHealth. (2012). Genotype Relative Risk. Retrieved on June, 01, 2015, from http://www.decodehealth.com/glossary#genotypeRR

Howlader, N., Noone, A. M., Krapcho, M., Garshell, J., Miller, D., Altekuse, S. F., Kosary, C. L., … Cronin K. A. (eds). SEER Cancer Statistics Review, 1975-2011, National Cancer Institute. Besthesda, MD. http://seer.cancer.gov/csr/1975_2011/, based on November 2013 SEER data submission, posted to the SEER website, April 2014. Updated September 10, 2014. Retrieved on June, 01, 2015.

National Cancer Institute (NCI). (2012). Defining Cancer. National Cancer Institute at the National Institutes of Health. Retrieved on June, 01, 2015, from http://www.cancer.gov/cancertopics/cancerlibrary/what-is-cancer

National Cancer Registry (NCR). (2012). Incidence of Histologically Diagnosed Cancer in South Africa, 2004. Retrieved on June, 01, 2015, from http://www.nioh.ac.za/?page=cancer_statistics&id=163

NCCN. (2014a). NCCN Clinical Practice Guidelines in Oncology, Breast Cancer. Retrieved on June, 01, 2015, from http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site

NCCN. (2014b). NCCN Clinical Practice Guidelines in Oncology, Genetic/Familial High-Risk Assessment: Breast and Ovarian. Retrieved on June, 01, 2015, from http://www.nccn.org/professionals/physician_gls/f_guidelines.asp#site

NCI Dictionary of Cancer Terms. (2015). NCI Dictionary of Cancer TermsL Sporadic cancer. NIH National Cancer Institute. Retrieved on June, 01, 2015, from http://www.cancer.gov/publications/dictionaries/cancer-terms?cdrid=339347

Larrsson, N., Borg, A., Hodgson, S., Sinilnikova, O., Loman, N., McDevitt, T., Muller-Reible, C., Krisoffersson, U. (2008). EMQN Best Practice Guidelines for Molecular Genetics Analysis in Hereditary Breast/Ovarian Cancer. Retrieved on June, 01, 2015, from http://www.emqn.org/emqn/digitalAssets/0/232_EMQNBRCAguidelines0908.pdf

Petrucelli, N., Daly, M.B., Feldman, G.L., (1998). *BRCA1* and *BRCA2* Hereditary Breast and Ovarian Cancer. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2014. Retrieved on June, 01, 2015.

WHO. (2012). Health Topic: Cancer. Retrieved on June, 01, 2015, from http://www.who.int/topics/cancer/en/

7. APPENDICES

7.1. Appendix A – Ethics certificate M110922

.UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG

Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Mr Wenlong Chen

CLEARANCE CERTIFICATE	M110922
PROJECT	The molecular Aetiology of Inherited Breast Cancer in the South African Black Population
INVESTIGATORS	Mr Wenlong Chen
DEPARTMENT	Division of Human Genetics
DATE CONSIDERED	30/09/2011
DECISION OF THE COMMITTEE*	Approved unconditionally

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 30/09/2011 **CHAIRPERSON** .....
(Professor P E Cleaton Jones)

*Guidelines for written 'informed consent' attached where applicable

cc: Supervisor : Dr Robyn Kerr

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

7.2. Appendix B – Ethics certificate M10745

UNIVERSITY OF THE WITWATERSRAND, JOHANNESBURG
Division of the Deputy Registrar (Research)

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)
R14/49 Professor M Ramsay et al

CLEARANCE CERTIFICATE	**M10745**
PROJECT	Genetic Diseases Related Studies in Southern African Populations: Control Samples (Previously M050706)
INVESTIGATORS	Professor M Ramsay et al.
DEPARTMENT	Division of Human Genetics
DATE CONSIDERED	06/08/2010
DECISION OF THE COMMITTEE*	Renewal approved

Unless otherwise specified this ethical clearance is valid for 5 years and may be renewed upon application.

DATE 06/08/2010 **CHAIRPERSON** .....
(Professor PE Cleaton-Jones)

*Guidelines for written 'informed consent' attached where applicable
cc: Supervisor : Prof A Krause

DECLARATION OF INVESTIGATOR(S)

To be completed in duplicate and **ONE COPY** returned to the Secretary at Room 10004, 10th Floor, Senate House, University.
I/We fully understand the conditions under which I am/we are authorized to carry out the abovementioned research and I/we guarantee to ensure compliance with these conditions. Should any departure to be contemplated from the research procedure as approved I/we undertake to resubmit the protocol to the Committee. **I agree to a completion of a yearly progress report.**
PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES...

7.3. Appendix C – Recipes

Reagents and Solutions

0.5M EDTA (pH 8)

93.05g EDTA

500mL dH_2O

10X TBE Buffer (pH 8)

108g Tris

55g Boric Acid

7.44g EDTA

Makes up to 1 litre with dH_2O

0.8 % Agarose Gel

400ml 1x TBE

3.2g Molecular Grade Agarose

12 µL EtBr (10mg/mL)

1.5 % Agarose Gel

400ml 1x TBE

6g Molecular Grade Agarose

12 µL EtBr (10mg/mL)

1x TE Buffer

10 mL 1M Tris-HCL (pH8)

2mL 0.5M EDTA

Add ddH_2O to make up to 1L

1M Tris-HCL (pH7.4)

121.1g Tris Base

Add ddH_2O to make up to 1L

70% Ethanol

700mL Absolute Ethanol

250mL dH2O

BRCA1 PCR MIX

See Appendix E

BRCA2 PCR MIX

See Appendix E

BIGDYE V3.1. CYCLE SEQUENCING MIX

See Appendix E

7.4. Appendix D – BRCA primer lists

The BRCA primers were obtained from the research protocol described by Mattocks, et al. (2010). The primer list was adapted from the Supplemental Data 2 of the original publication by Mattocks, et al. (2010).

Notes on primer tables.

1. Primer reference is composed of [gene]_[fragment]_[orientation]_[tag]_[design iteration]. BC1=*BRCA1* and BC2=*BRCA2*, fwd=forward primer, rvs=reverse primer.

2. Locations of all the primers are given as the most 5' base on the coding strand. This means that for forward primers the 5' end is given and for reverse primers the 3' end is given.

3. The gene specific primer sequence only has been given. Primers work equally well with or without tags.

4. In some cases a filler is used to prevent hair-pin looping of the primer if the tag is used. Tagged primers are composed 5'[tag]-[filler]-[gene specific primer sequence]3'

5. Fragment length includes filler and tag sequences

BRCA1 Primer set – BC1

Location	17q21-q24
Reference sequence (cDNA)	NM_007294.2 Nucleotide 1 = base 1
Design basis (genomic DNA)*	NCBI36:17:38440001:38560000:-1
Latest SNP check	23/9/2008 (dbSNP build 129)
OMIM	113705

* NCBI build No : chromosome no : from : to : orientation w.r.t chromosome [1=p to q, -1=q to p]

Primer reference[1]	Location[2]	Gene specific primer sequence (5'-3')[3]	Filler[4]	Fragment length[5]
BC1_02_fwd_US1_04	1-133	ATTCTAAAACCTTCCAAATCTTAAA		345
BC1_02_rvs_US2_01	80+76	CATGTCTTTTCTTCCCTAGTATG		345
BC1_03_fwd_US2_01	81-127	ACTTGAGGCCTTATGTTGACTC		362
BC1_03_rvs_US1_01	134+125	CTCTCTGAGAAAGAATGAAATGG		362
BC1_05_fwd_US2_01	135-107	CTTAAGGGCAGTTGTGAGATTAT		439

BC1_05_rvs_US1_01	212+198	GAAGCATTAGAGAAAGGCAGTAA		439
BC1_06_fwd_US2_02	213-157	CACGGTTTATACAGATGTCAATG		446
BC1_06_rvs_US1_01	301+147	TGTGAGACCAGTGGGAGTAA		446
BC1_07_fwd_US1_01	302-81	GGGTTTCTCTTGGTTTCTTTGATT		293
BC1_07_rvs_US2_02	441+7	AAGAAGAAGAAGAAGAAGAAAACAAATGG		293
BC1_08_fwd_US2_01	442-79	TGGTGTCAAGTTTCTCTTCAGG		361
BC1_08_rvs_US1_02	547+120	ATAACTCACCATAGGGCTCATAA		361
BC1_09_fwd_US1_03	548-55	CCCTTTTAATTAAGAAAACTTTTATTG		275
BC1_09_rvs_US2_03	593+117	CCTATATAACAAACTGCACATAC	A	275
BC1_10_fwd_US1_01	594-178	CCAGCAACCATTTCATTTC		445
BC1_10_rvs_US2_01	670+135	ATACCACGACATTTGACAGAGA		445
BC1_11A_fwd_US1_01	671-152	TGAGCTACATCTTCAGTATACTTGGT		473
BC1_11A_rvs_US2_01	937	GTTATGTTGGCTCCTTGCTAA		473
BC1_11B_fwd_US1_01	845	CATTACAGCATGAGAACAGCAGT		440
BC1_11B_rvs_US2_01	1228	TTTAGAACGTCCAATACATCAGC		440
BC1_11C_fwd_US2_02	1159	TCCAGAAGTGATGAACTGTTAGGT		441
BC1_11C_rvs_US1_02	1542	TCTGCTTTCTTGATAAAATCCTCA		441
BC1_11D_fwd_US1_01	1476	AATACAAGAGCGTCCCCTCAC		472
BC1_11D_rvs_US2_01	1888	TCAGTACAATTAGGTGGGCTTAGATT		472
BC1_11E_fwd_US1_01	1797	TATCCACAATTCAAAAGCACCTA		301
BC1_11E_rvs_US2_02	2040	GTCTGTTCATTTGGCTTGTTACTC		301
BC1_11F_fwd_US2_03	1961	AGTACAACCAAATGCCAGTCAGG		506
BC1_11F_rvs_US1_02	2411	GGGTTTTCAAATGCTGCACACT		506
BC1_11G_fwd_US2_01	2319	TGGTACTGATTATGGCACTCA		415
BC1_11G_rvs_US1_01	2677	GTGACTTTTGGACTTTGTTTCTT		415
BC1_11H_fwd_US1_01	2626	GGAAATGCAGAAGAGGAATGTG		379
BC1_11H_rvs_US2_01	2948	AACARATGACTTGATGGGAAAAA		379
BC1_11I_fwd_US2_01	2878	GGCAACGAAACTGGACTCA		428
BC1_11I_rvs_US1_01	3249	GTTGCAAAACCCCTAATCTAAGC		428
BC1_11J_fwd_US2_01	3206	AAGCAGAACTAGGTAGAAACAGA		500
BC1_11J_rvs_US1_01	3653	GGAAGCTCTTCATCCTCAC		500
BC1_11K_fwd_US2_01	3547	ARAGGAGAGCTTAGCAGGAGTC		416
BC1_11K_rvs_US1_01	3906	TATTTGCAGTCAAGTCTTCCAAT		416
BC1_11L_fwd_US2_02	3818	AGGTAATATTGGCAAAGGCATC		341
BC1_11L_rvs_US1_02	4096+10	GGCAAACACAAAAACCTGG		341
BC1_12_fwd_US2_01	4097-83	TGCCAATGAGAAGAAAAAGAC	A	307
BC1_12_rvs_US1_02	4185+80	CAAAGAGATGATGTCAGCAAA		307
BC1_13_fwd_US2_03	4186-152	GGTTAGATTAAAAGGTGTTCAGC		463
BC1_13_rvs_US1_02	4357+84	CAAGGATCATAAAATGTTGGAG		463
BC1_14_fwd_US2_02	4358-53	TCACTATCAGAACAAAGCAGTAAAG		258
BC1_14_rvs_US1_02	4484+23	AAAGATGTCAGATACCACAGCA		258
BC1_15_fwd_US1_01	4485-63	SAAGTATGATTTGTCCTTTCACA		397
BC1_15_rvs_US2_01	4675+85	GTGTCCATGATAGACTAGTACATCT		397
BC1_16_fwd_US1_02	4676-82	CTTTTTAATTCTTAACAGAGACCAG		486
BC1_16_rvs_US2_02	4986+36	CTCTTTCCAGAATGTTGTTAAGTC		486
BC1_17_fwd_US1_02	4987-56	TCTGAGCTGTGTGCTAGAGGT		320
BC1_17_rvs_US2_03	5074+39	GCCTCATGTGGTTTTATGC	A	236

BC1_18_fwd_US1_01	5075-137	TGTCACCAGGGGTTTTAGAATC		292
BC1_18_rvs_US2_02	5152+23	TGAGGTGTTAAAGGGAGGAGG		292
BC1_19_fwd_US1_02	5153-87	CTTTTGTGAATCGCTGACCTCT	CT	283
BC1_19_rvs_US2_02	5193+97	ATCATGGAAAATTTGTGCATTGT		283
BC1_20_fwd_US1_01	5194-46	GCTTCTCTTTCTCTTATCCTGATG		398
BC1_20_rvs_US2_01	5277+213	AATCCAAATTACACAGCCTCTC		398
BC1_21_fwd_US1_01	5278-46	CCTTCTCTCCATTCCCCTGTCC		278
BC1_21_rvs_US2_01	5332+122	TTTGGAGAGTGGTAGAGAAATA		278
BC1_22_fwd_US1_01	5333-139	ATTGTCCTTTGGAGCAGAGAG		410
BC1_22_rvs_US2_01	5406+143	TCCTTACCCATCCCTTACAGA		410
BC1_23_fwd_US1_03	5407-67	GATGAAGTGACAGTTCCAGTAGTC		279
BC1_23_rvs_US2_03	5467+94	TTCAACAAATATTTAAAATGTGCC		279
BC1_24_fwd_US2_02	5468-76	CCTGGAGTCGATTGATTAGAGC	A	335
BC1_24_rvs_US1_02	5669	GAAGGACTGAAGAGTGAGAGGAG		335

BRCA2 Primer set – BC2

Location	13q12-q13
Reference sequence (cDNA)	NM_000059.3 1 = base 1
Design basis (genomic DNA)*	NCBI36:13:31780001:31880000:1
Latest SNP check	23/9/2008 (dbSNP build 129)
OMIM	600185

* NCBI build No : chromosome no : from : to : orientation w.r.t chromosome [1=p to q, -1=q to p]

Primer reference[1]	Location[2]	Gene specific primer sequence (5'-3')[3]	Filler[4]	Fragment length[5]
BC2_02_fwd_US1_01	1-103	TACCTCAGTCACATAATAAGGA		284
BC2_02_rvs_US2_01	67+63	CGTACTGGGTTTTTAGCA		284
BC2_03_fwd_US2_01	68-73	ATCTTTAACTGTTCTGGGTCAC		430
BC2_03_rvs_US1_01	316+52	CTAAGATTTTAACACAGGTTTGC		430
BC2_04_fwd_US1_01	317-123	ACAACTCCCTATACATTCTCATTC		368
BC2_04_rvs_US2_01	425+76	GCCAAAATATTAGCATAAAAATCA	TTA	368
BC2_05_fwd_US2_01	426-118	ATCTAAAAGTAGTATTCCAACAAT		405
BC2_05_rvs_US1_01	516+53	GGCAAAGGTATAACGCTAT		405
BC2_07_fwd_US2_01	516+89	GTACCTAGCATTCTGCCTCATAC		321
BC2_07_rvs_US1_01	631+22	ACCTCATCTGCTCTTTCTTGTAA		321
BC2_08_fwd_US1_02	632-171	CACTGTGTTGATTGACCTTTCTA		311
BC2_08_rvs_US2_02	681+37	RGCATTCCAAAATTGTTAGC		311
BC2_09_fwd_US1_01	682-116	AGGTTGATTGCAGATAACTGAAAT		294
BC2_09_rvs_US2_02	793+10	AAACCTGTAGTTCAACTAAACAG		294
BC2_10A_fwd_US2_01	794-44	TGGCTTATAAAATATTAATGTGCTTC		378
BC2_10A_rvs_US1_03	1088	GCTACATTTGAATCTAATGGATCAG		397
BC2_10B_fwd_US1_03	1002	TGAAGCAAACGCTGATGAATGTG		501
BC2_10B_rvs_US2_02	1447	CAGATATTGCCTGCTTTACTGC		501
BC2_10C_fwd_US2_01	1364	CRGAGAAGCCATTAAATGAGGA		409
BC2_10C_rvs_US1_02	1717	TTAAACCTGCATTCTTCAAAGC		409
BC2_10D_fwd_US1_01	1633	GTTTGCTCACAGAAGGAGGAC		341
BC2_10D_rvs_US2_02	1909+4	CTATTTACAAAAAAAAAAAAGACAGAGG		341
BC2_11A_fwd_US1_01	1910-46	TGATTGATGGTACTTTAATTTTGTC		433

BC2_11A_rvs_US2_01	2238	TTGAAAGTCAGTATCACTGTATTCC		433
BC2_11B_fwd_US1_01	2172	AAAAGTTTCAGATATAAAAGAAGAGGTC		504
BC2_11B_rvs_US2_01	2618	CTTCAGAGTCTGGATTGACAGTTA		504
BC2_11C_fwd_US2_02	2517	CATGAGAGTAGCATCACCTTC		509
BC2_11C_rvs_US1_01	2972	TAAGAGTCCTGCCCATTTGT		509
BC2_11D_fwd_US1_01	2913	AAAATCGGACATCTCCTTGA		498
BC2_11D_rvs_US2_01	3354	CAAACTGACTTCCTGATTCTTCT		498
BC2_11E_fwd_US1_01	3281	AGCAGGATTTTAATTCAAACCATAA		462
BC2_11E_rvs_US2_02	3709	CTCAATATCACTAAACAGTTTCACAGC		489
BC2_11F_fwd_US1_01	3629	ATGAAAATGAAGTGGGGTTTAG		478
BC2_11F_rvs_US2_01	4050	ATAGCAAGTCCGTTTCATCTTTA		478
BC2_11G_fwd_US2_01	3995	ATAACTTAGAATTTGATGGCAGTG	A	467
BC2_11G_rvs_US1_01	4403	CCATTTTGTTCTTTCTTATGTCAG		467
BC2_11H_fwd_US2_01	4291	GCAAGTGGGAAAAATATTAGTG	A	506
BC2_11H_rvs_US1_01	4741	AGCTGTGATCTCAATGGTCTC		506
BC2_11I_fwd_US2_01	4663	GAAATCACCAGTTTTAGCCATC		495
BC2_11I_rvs_US1_01	5099	GCAGTATTTATTCTTTCTGGTTGAC		495
BC2_11J_fwd_US2_01	5000	CAGCCTTAGCTTTTTACACAAGTT	A	505
BC2_11J_rvs_US1_01	5448	TTTATTTTTGCAGGGTGAAGAG		505
BC2_11K_fwd_US2_01	5312	GTATTGAGCCAGTATTGAAGAATG		492
BC2_11K_rvs_US1_01	5747	GTCAGCAAAAACCTTATGTGAAT		492
BC2_11L_fwd_US2_01	5660	CGAAAATTATGGCAGGTTGT		501
BC2_11L_rvs_US1_01	6104	TTGGGATATTAAATGTTCTGGAG		501
BC2_11M_fwd_US1_01	6049	AAAAGTAACGAACATTCAGACCA		490
BC2_11M_rvs_US2_01	6481	TCCTAATACCAACTGTTGTTTGTC		490
BC2_11N_fwd_US1_01	6334	AGAAACCCAGAGCACTGTGTAA		475
BC2_11N_rvs_US2_01	6752	TTCGGGACATGTAAAAAGAGAAT		475
BC2_11O_fwd_US1_02	6573	AATGGAAATTGGTAAAACTGAAAC		345
BC2_11O_rvs_US2_02	6841+22	AAATAGTGATTGGCAACACGA		345
BC2_12_fwd_US2_01	6842-208	TTTAAAGTGGTCAAAACAGAACA	A	473
BC2_12_rvs_US1_01	6937+115	AGGTCCTTGATTAGGCACAG		473
BC2_13_fwd_US1_01	6938-263	CACCTTTGAGTTATGATGGTTA		423
BC2_13_rvs_US2_01	7007+33	AAACGAGACTTTTCTCATACTGTA		423
BC2_14A_fwd_US1_01	7008-94	CCATGTAGCAAATGAGGGTC		332
BC2_14A_rvs_US2_01	7189	TTGGTTGGTCTGCCTGTAGTAAT		332
BC2_14B_fwd_US1_01	7115	CAAGCAATTTAGCAGTTTCAGGA		464
BC2_14B_rvs_US2_01	7435+87	GGCTTTAAAATTACCACCACCAA		464
BC2_15_fwd_US2_01	7436-55	GGTTGTGCTTTTTAAATTTCAAT		314
BC2_15_rvs_US1_01	7617+22	CACTCTGTCATAAAAGCCATCA		314
BC2_16_fwd_US1_01	7618-97	GGTAAATTCAGTTTTGGTTTGTTAT		388
BC2_16_rvs_US2_01	7805+46	CAGTTAAGAGAAGAAAGAGGGATG		388
BC2_17_fwd_US1_01	7806-205	CCATGCTCAGCAATGAAGTT		490
BC2_17_rvs_US2_01	7976+56	GACATGGAAGTCACAGACTACACA	A	490
BC2_18A_fwd_US2_01	7977-59	AAACAGTGGAATTCTAGAGTCACA		422
BC2_18A_rvs_US1_01	8286	CAAGAGGTGTACAGGCATCA		422
BC2_18B_fwd_US2_01	8197	GATCCTCCCCTCTTAGCTGTCTTA		346
BC2_18B_rvs_US1_01	8331+157	CCTCCCAAAAACTGCACAAAA		346
BC2_19_fwd_US2_01	8332-97	TAAGGCAGTTCTAGAAGAATGAA		368
BC2_19_rvs_US1_01	8487+61	GAAACTCCATCTCAAACAAAC		368
BC2_20_fwd_US2_01	8488-67	GGCCTGATACAATTAACTTGAA	AT	441
BC2_20_rvs_US1_01	8632+172	AGACCTGATATTTCTGTCCCTT		441
BC2_21_fwd_US1_01	8633-103	ATCTTTAAATCTCCCTTCTTTGG		362
BC2_21_rvs_US2_01	8754+82	CAGCATTTCAACATACTCCTTC		362
BC2_22_fwd_US2_02	8755-44	TTTGTTCTGATTGCTTTTTATTCC		355
BC2_22_rvs_US1_02	8953+50	GGATTTATAATCATTTTGTTAGTAAGGTC		355
BC2_23_fwd_US1_02	8954-50	AAATGATAATCACTTCTTCCATTGC		330
BC2_23_rvs_US2_02	9118-36	AAAACAAAAATTCAACATATGGAGA		330
BC2_24_fwd_US1_01	9117+37	AGTGCTTGTTAGTTTATGGAATC		371

BC2_24_rvs_US2_01	9256+121	GAGGTTCAAAGAGGCTTACTT		371
BC2_25_fwd_US1_02	9257-81	ATATTAGAGTTTCCTTTCTTGCATC		447
BC2_25_rvs_US2_02	9501+62	AATTTGTATAAAAGCTATTTCCTTGA		447
BC2_26_fwd_US1_01	9502-132	GGTATCACATTTAGGGTTTTTCA		456
BC2_26_rvs_US2_01	9648+119	CTATACTTACAGGAGCCACATAACA		456
BC2_27A_fwd_US1_01	9649-130	TTAGGAGTTAGGGGAGGGAGACT		424
BC2_27A_rvs_US2_01	9889	CCTTGGTGGCTGAAATGC	CT	424
BC2_27B_fwd_US2_01	9762	AGGGGAGAAAGAGATTGATGACC		346
BC2_27B_rvs_US1_01	10051	CCTGTTGAACCAGACAAAAGAGC		346
BC2_27C_fwd_US2_01	10002	GGAAAGTAATTCAATAGCTGACG		387
BC2_27C_rvs_US1_01	10331	TGCAACATAAGTACTAATGTGTGG		387

7.5. Appendix E – Protocols

7.5.1. BRCA1 PCR protocol

DATE:		BENCH DECONTAMINATED:		PERFORMED BY:		
SAMPLE PREPARATION						
SAMPLE ID	**STOCK DNA CONC. (ng/µl)**	**260/280**	**DNA (µl)**	**ddH2O (µl)**	**RESULTS**	
1.						
2.						
3.						
4.						
5.						
6.						
7.						
8.						
9.						
10.						
11.						
12.						
13.						
14.						
15.						
16.						
17.						
18.						
19.						
20.						
PCR SETUP						
Fragment Number:						
PCR COMPONENTS	**STOCK CONC.**	**FINAL CONC.**	**BATCH/LOT NO.**	**1x MIX**	**__ x MIX**	
DNA	-	50 ng/µl	-	1.00 µl	-	
5x Reaction Buffer	5 x	1 x		2.50 µl	µl	
Primer Mix	10 pmol/µl	0.1 pmol/µl		0.50 µl	µl	
ddH2O	-	-	-	20.80 µl	µl	
MyTaq HS Poly.	5 U/µl	1 U		0.20 µl	µl	
TOTAL				**25 µl**	**24 µl/rxn**	

Thermocycler:		**Band Pattern Reference:**
Program:		bp

PCR Conditions:			**POST PCR CLEANUP PROTOCOL**
95 °C	2 minutes	1 cycle	1. Add ddH2O to PCR products to make 100 µl
95 °C	25 seconds	40 cycles	2. Transfer PCR Products to Purple plate (PCR96)
61 °C	30 seconds		3. Vacuum for 5 minutes @ >15inch.hg
72 °C	60 seconds		4. Blot&Wipe bottom of plate on paper to dry off
72 °C	5 minutes	1 cycle	5. Add ddH2O to each well
15 °C	Hold		6. Vacuum for 4.5 minutes @ >15inch.hg
Run 5 µl of PCR products on 1.5% Agarose gel @ 150 volts in 50 cm tank fo 30 minu			7. Blot/Wipe bottom of plate on paper to dry off
			8. Resuspend sample with 20 µl of ddH2O
			9. Pipette each sample up/down for 20x to mix
POST PCR CLEANUP			10. Transfer products into 96 well V-bottom plate
Use Montage PCR96 plate for Post PCR Cleanup			

7.5.2. BRCA2 PCR protocol

DATE:		BENCH DECONTAMINATED:		PERFORMED BY:		
SAMPLE PREPARATION						
SAMPLE ID	**STOCK DNA CONC. (ng/µl)**	**260/280**	**DNA (µl)**	**ddH2O (µl)**	**RESULTS**	
1.						
2.						
3.						
4.						
5.						
6.						
7.						
8.						
9.						
10.						
11.						
12.						
13.						
14.						
15.						
16.						
17.						
18.						
19.						
20.						
PCR SETUP						
Fragment Number:						
PCR COMPONENTS	**STOCK CONC.**	**FINAL CONC.**	**BATCH/LOT NO.**	**1x MIX**	**17 (x 2) x MIX (x3)**	
DNA	-	50 ng/µl	-	1.00 µl	-	
KAPA ReadyMix	2 x	1 x		12.00 µl	204 (102 x 2) µl	
Primer Mix	10 pmol/µl	0.1 pmol/µl		0.50 µl	8.5 µl	
ddH2O	-	-	-	11.50 µl	195.5 µl	
TOTAL				**25 µl**	**24 µl/rxn**	
Thermocycler:				**Band Pattern Reference:**		
Program:				bp		
PCR Conditions:				**POST PCR CLEANUP PROTOCOL**		
95 °C	2 minutes	1 cycle		1. Add ddH2O to PCR products to make 100 µl		
95 °C	25 seconds	40 cycles		2. Transfer PCR Products to Purple plate (PCR96)		
61 °C	30 seconds			3. Vacuum for 7 minutes @ >15inch.hg		
72 °C	60 seconds			4. Blot&Wipe bottom of plate on paper to dry off		
72 °C	5 minutes	1 cycle		5. Add ddH2O to each well		
15 °C	Hold			6. Vacuum for 5 minutes @ >15inch.hg		
Run 5 µl of PCR products on 1.5% Agarose gel @ 150 volts in 50 cm tank fo 30 min				7. Blot/Wipe bottom of plate on paper to dry off		
				8. Resuspend sample with 20 µl of ddH2O		
				9. Pipette each sample up/down for 20x to mix		
POST PCR CLEANUP				10. Transfer products into 96 well V-bottom plate		
Use Montage PCR96 plate for Post PCR Cleanup						

7.5.3. Cycle sequencing protocol

BIGDYE V.3.1. CYCLE SEQUENCING SETUP						
COMPONENTS	**STOCK CONC.**	**FINAL CONC.**	**BATCH/LOT NO.**	**1x MIX**	**13 x MIX (x 8)**	
Cleaned PCR Produc	-	-	-	4 µl	-	
BigDye v.3.1	-	-		1 µl	12.00 µl	
BigDye Buffer	5 x	-		1.5 µl	19.50 µl	
Primer US ( )	10 pmol/µl			1 µl	13.00 µl	
ddH2O	-	-	-	4.5 µl	60.00 µl	
TOTAL				**10 µl**	**8 µl/rxn**	

Cycle Sequencing Conditions:				**POST BIGDYE CLEANUP PROTOCOL**	
95 °C	30 seconds	25 cycles		1. Add 20 µl of injection solution to product, mix	
50 °C	15 seconds			2. Transfer diluted products onto purple plate	
60 °C	2 minutes			3. Vacuum for 2minutes @ >15inch.hg	
24 °C	Hold			5. Blot&wipe bottom of plate dry, NB!!!	
				6. Add 20 µl of injection solution to each well	
				7. Vacuum for 1-2 minutes @ >15inch.hg	
POST BIGDYE CLEANUP				8. Blot&wipe bottom of plate dry NB!!!	
Use Montage PCR96 plates				9. Repeat Step to 6 to 8	
Keep sample in Dark bag				10. Pipette each sample up/down for 20x to mix	
Added 5 µl of Injection Solution to blank wells				11. Transfer products to injection plate	
PLATE & SEQUENCER SETUP					
Plate Preparation				**Sequencer Preparation**	
Load 20 µl of resuspended product onto 96 well plate				Add plate under Plate manager	
Denature sample for 2 minutes @ 95 °C				Label samples	
Place plate on ice before loading				Instrument protocol - Z_Seq POP7-36 Ultra (400 bp to 500 b	

7.5.4. MLPA protcol for BRCA

General protocol for MLPA - One tube procedure

Probe kit:		**Bench decontaminated:**	
Date:		**Performed by:**	

The protocol takes place over two days.

DAY 1:

1. DNA sample preparation

- Dilute DNA samples to 40 ng/µl
- Dilute DNA in 1x TE buffer
- Allow DNA to resuspend for 45 minutes

2. Hybridisation protocol

- Add 5 µl of diluted DNA to 0.2 ml PCR Strip tube
- Denature DNA using the Denaturation protocol

Denaturation protocol		
98 °C	5 minutes	x1 cycle
25 °C	hold	

- Prepare Hybridisation mix (Hybe Mix)
- Vortex reagents briefly before use

Hybe Mix			
Reagents	**Batch number**	**1x volume**	**x Mix**
SALSA probe mix (black cap)		1.5 µl	µl
MLPA buffer (yellow cap)		1.5 µl	µl
Total	-	3.0 µl	3.0 µl/rxn

Mix thoroughly and centrifuge. Extra Hybe Mix can be frozen and reused.

- Remove tubes from thermal cycler
- Add 3.0 µl of Hybe Mix to each tube
- Mix by pipetting
- Cap tube with new strip cap
- Start Hybridisation protocol:

Hybridisation protocol		
95 °C	1 minute	x1 cycle
60 °C	18 hours (12 to 26 hours)	x1 cycle
54 °C	hold	

- End of DAY 1

DAY 2:

3. Ligation protocol

- Prepare a Ligase-65 mastermix using Buffer A and B
- Vortex Buffer A and B before use
- Note to prepare Ligase-65 Mix on ice

Ligase-65 Mix

Reagent	Batch number	1x volume	x Mix
Ligase-65 Buffer A (transparent cap)		3.0 µl	µl
Ligase-65 Buffer B (white cap)		3.0 µl	µl
ddH_2O	-	25.0 µl	µl
Mix thoroughly and centrifuge			
Ligase-65 (green cap)		1.0 µl	µl
Mix by PIPETTING and centrifuge. Extra Ligase-65 mix can be frozen and reused			
Hybridisation products	-	8.0 µl	-
Total	-	40.0 µl	32.0 µl/rxn

- Do not remove 0.2 ml PCR Strip tubes from thermal cycler
- Add 32 µl of Ligase-65 mix to each reaction while at 54 °C
- Mix by pipetting
- Cap tube with new strip cap
- Start Ligation protocol

Ligation protocol

54 °C	15 minutes	x1 cycle
98 °C	5 minutes	x1 cycle
20 °C	hold	

- Remove 0.2 PCR Strip tubes from thermal cycler
- Proceed to PCR setup or Ligated products can be stored at 4 °C

4. PCR protocol (One tube)

- Note to prepare PCR mix on ice
- Vortex & centrifuge the SALSA PCR mix before use
- Warm the Polymerase for 10 seconds at body temperature to reduce viscosity

Polymerase Mix

Reagents	Batch number	1x volume	x Mix
SALSA PCR primer mix (brown cap)*		2.0 µl	µl
ddH2O	-	7.5 µl	µl
SALSA polymerase (orange cap)		0.5 µl	µl
Mix by PIPETTING and centrifuge. Extra Polymerase Mix can be frozen and reused.			
MLPA ligated products	-	40.0 µl	-
Total	-	50.0 µl	10.0 µl/rxn

FAM labelled primers. For One tube reaction, primer mix has the MRC Holland logo on the label

- At room temperature, add 10 µl of Polymerase Mix to each tube
- Mix by pipetting & centrifuge
- Cap tube with new strip cap
- Start PCR protocol

PCR protocol

95 °C	30 seconds	x 35 cycles
60 °C	30 seconds	
72 °C	1 minute	
72 °C	20 minutes	x 1 cycle
4 °C	hold	

- PCR products can be stored for one week at 4 °C. For longer period storage, store PCR products at -25 °C to -15 °C.
- Store PCR products in a black bag of cover with foil to prevent loss-of fluorescence

5. Genetic Analyzer protocol

- Prepare LIZ-500 marker mix
- Load 1 µl of PCR products with 0.5 µl LIZ-500 marker mix and 8.5 µl Hi-Dye

LIZ-500 marker mix			
Reagents	**Batch number**	**1 Run (16 samples)**	**3 Runs (48 samples)**
ABI LIZ-500 Marker		7.0 µl	10.0 µl
Hi-Dye		146.0 µl	350.0 µl
Mix and centrifuge.Extra LIZ-500 marker mix can be frozen and reused.			

- Run products on the ABI 3130xl genetic analyzer
- GeneMapper - Generic application be used for data collection and Coffalyser analysis

Name of run: **Date:**

7.6. Appendix F – MLPA kit probe list

The MLPA kit probe list were obtained from MRC Holland (www.mlpa.com).

7.6.1. P002-C2 BRCA1

Notes

* LPO is the 5' half of the probe, RPO is the 3' half of the probe.
* The two sequences are directly adjacent and together they form the complete hybridising sequence of the probe.
* A small number of probes consist of three oligonucleotides and have a SP (spanning oligo).
* The first nucleotides of the LPO may not be complementary to the target DNA and is part of the stuffer sequence.
* Locations are based on hg18 / mapview build 36. Mapview location consist of the chromosome no. + distance (nt) from the p-telomere to the start of the probe sequence.
* Observed length of the fragments depend on the software, sequencer type and molecular weigth markers used.
* Exon numbering might be different from the literature.
* Please let us know if you find SNPs under one of these sequences that affects the probe signal: info@mlpa.com

length	number	ref. probe	gene	genbank exon	mut. spec.	mapview	chr. pos.	LPO	SP	RPO	start	end
127	00797-L00093	TRUE	IL4	1	no	05-132.037612	05q31.1	CATTGTCACTGCAAATCGACACCTAT		TAATGGGTCTCACCTCCCAACTGCTTCCCCCT	132037612	132037671
136	06452-L05978	TRUE	NRSN1	4	no	06-024.253896	06p22.2	CTCTGGACATGTACAAGCTGGCAGGA		GCTGTTCTCTTCTGCATTGGAGGCACGTCCATGGCAGG	24253896	24253960
148	00763-L00268	FALSE	BRCA1	1a	no	17-038.530812	17q21.31	CCTTTACCCAGAGCAGAGGGTGA		AGGCCTCCTGAGCGCAGGGGCCCAGTTATCTGAGA	38530812	38530870
157	00764-L00269	FALSE	BRCA1	1b	no	17-038.530571	17q21.31	TGAGGATCAGGAAGGGGGCACTGA		GTGTCCGTGGGGGAATCCTCGTGATAGGAACTGG	38530571	38530629
166	00765-L00270	FALSE	BRCA1	2	no	17-038.529585	17q21.31	ACAGAAAGAAATGGATTTATCTGCTC		TTCGCGTTGAAGAAGTACAAAATGTCATTAATGCTATG	38529585	38529649

175	00826-L00341	FALSE	BRCA1	3	no	17-038.521268	17q21.31	CTGCTAGTCTGGAGTTGATCAAGGAACCTG		TCTCCACAAAGTGTGACCACATATTTTGCAA	38521268	38521329
184	00767-L00272	FALSE	BRCA1	5	no	17-038.512011	17q21.31	TTTTGCATGCTGAAACTTCTCAACCA		GAAGAAAGGGCCTTCACAGTGTCCTTTATGTAAGAATG	38512011	38512075
198	02946-L03265	TRUE	CFTR	2	no	07-116.931557	07q31.2	CGATTGTCAGCAGAATCAACAGAAGGGATTTGGTATA		TGTCTGACAATTCCAGGCGCTGTCTGTATCCTTTCCTCAAA	116931557	116931634
208	00827-L00342	FALSE	BRCA1	6	no	17-038.510430	17q21.31	GGAGCCTACAAGAAAGTACGAGATTTAGTC		AACTTGTTGAAGAGCTATTGAAAATCATTTGTGCTTTTCA	38510430	38510497
216	00769-L00274	FALSE	BRCA1	7	no	17-038.509662	17q21.31	TGGGCTACAGAAACCGTGCCAAA		AGACTTCTACAGAGTGAACCCGAAAATCCTTCCTTGGT	38509662	38509723
226	01004-L00569	FALSE	BRCA1	8	no	17-038.505340	17q21.31	GTGTCCAACTCTCTAACCTTGGAACTGTG		AGAACTCTGAGGACAAAGCAGCGGATACAACCTCAAAA	38505340	38505407
236	01005-L00581	FALSE	BRCA1	9	no	17-038.502751	17q21.31	CGATCTGATTCTTCTGAAGATACCGTTAATAAGGC		AACTTATTGCAGGTGAGTCAAAGAGAACCTTTGTCTATGAAGCTGGT	38502751	38502832
244	00772-L00277	FALSE	BRCA1	10	no	17-038.501402	17q21.31	GTGTGGGAGATCAAGAATTGTTACAAATC		ACCCCTCAAGGAACCAGGGATGAAATCAGTTTGGA	38501402	38501466
256	00518-L00098	TRUE	IL1A	7	no	02-113.249258	02q13	GATGCCTGAGATACCCAAAACCAT		CACAGGTAGTGAGACCAACCTCCTCTTCTTC	113249258	113249313
266	00830-L00345	FALSE	BRCA1	11	no	17-038.500273	17q21.31	GGAACACCACTGAGAAGCGTGCAGCT		GAGAGGCATCCAGAAAAGTATCAGGGTAGTTCTGTTTC	38500273	38500334
277	00774-L00279	FALSE	BRCA1	11	no	17-038.497460	17q21.31	GAGCTTAGCAGGAGTCCTAGCCCTTT		CACCCATACACATTTGGCTCAGGGTTACCGAAGAG	38497460	38497521
285	00775-L00280	FALSE	BRCA1	12	no	17-038.496486	17q21.31	GAGTGAAACAAGCGTCTCTGAAGACTG		CTCAGGGCTATCCTCTCAGAGTGACATTTTAACCACTCAG	38496486	38496553
295	02603-L02074	FALSE	BRCA1	13	no	17-038.488025	17q21.31	GCTCCAGCAGGAAATGGCTGAACT		AGAAGCTGTGTTAGAACAGCATGGGAGCCAGCCTTCT	38488025	38488086
305	00833-L00349	FALSE	BRCA1	14	no	17-038.482059	17q21.31	GCTATAAGCCAGAATCCAGAAGGCCTT		TCTGCTGACAAGTTTGAGGTGTCTGCAGATAGTTCTACCA	38482059	38482124
316	00495-L00303	TRUE	LRMP	1	no	12-025.152177	12p12.3	CAAGAGGACTCATGGACGTCTCTAGAACA		TATCTTGTGGCCATTTACCAGACTCCGACAC	25152177	25152240
328	00778-L00347	FALSE	BRCA1	15	no	17-038.479958	17q21.31	GATGCACAGTTGCTCTGGGAGTCT		TCAGAATAGAAACTACCCATCTCAAGAGGAGCTCATTAAG	38479958	38480021
337	00779-L00003	FALSE	BRCA1	16	no	17-038.476716	17q21.31	CCCTTACCTGGAATCTGGAATCAG		CCTCTTCTCTGATGACCCTGAATCTGATCCTTCT	38476716	38476774
346	00780-L00283	FALSE	BRCA1	17	no	17-038.473167	17q21.31	TGCTCGTGTACAAGTTTGCCAGAAAAC		ACCACATCACTTTAACTAATCTAATTACTGAAGAGACTACTCA	38473167	38473237

355	00781-L00284	FALSE	BRCA1	18	no	17-038.469436	17q21.31	CTGCAGATGCTGAGTTTGTGTGTGAA		CGGACACTGAAATATTTTCTAGGAATTGCGGGAGGAAA	38469436	38469500
364	00782-L00285	FALSE	BRCA1	19	no	17-038.468858	17q21.31	TCTCTTTAGGGGTGACCCAGTCTATT		AAAGAAAGAAAAATGCTGAATGAGGTAAGTACTTGATGTTA	38468858	38468925
374	00655-L00304	TRUE	IL2	1	no	04-123.596999	4q27	TAACCTCAACTCCTGCCACAATGTACAGG		ATGCAACTCCTGTCTTGCATTGCACTAAGTCTTGCACT	123596999	123597063
389	00783-L00356	FALSE	BRCA1	20	no	17-038.462606	17q21.31	GGTCAGAGGAGATGTGGTCAATGGAA		GAAACCACCAAGGTCCAAAGCGAGCAAGAGAATCC	38462606	38462666
399	00784-L12004	FALSE	BRCA1	21	no	17-038.456598	17q21.31	CTCTTCAGGGGGCTAGAAATCTGTTGC		TATGGGCCCTTCACCAACATGCCCACAGGTAAGAG	38456598	38456659
407	00785-L00288	FALSE	BRCA1	22	no	17-038.454659	17q21.31	CAGCTGTGTGGTGCTTCTGTGGTGAA		GGAGCTTTCATCATTCACCCTTGGCACAGTAA	38454659	38454717
415	00786-L00289	FALSE	BRCA1	23	no	17-038.453196	17q21.31	TCCAGGGTGTCCACCCAATTG		TGGTTGTGCAGCCAGATGCCTGGACAGAGGACAA	38453196	38453251
427	02831-L13862	FALSE	BRCA1	24	no	17-038.450726	17q21.31	CCTGCAGAGAGTCAGACCCTTCAATGGAAGGA		GAGTGCTTGGGATCGATTATGTGACTTAAAGTCAGAATAGTCCT	38450726	38450799
436	00596-L00083	TRUE	LMO2	5	no	11-033.837614	11p13	TCCTGTGACAAGCGGATTCGT		GCCTATGAGATGACAATGCGGGTGAAAGACAAAG	33837614	33837669
445	04074-L03710	TRUE	NF1	23	no	17-026.581419	17q11.2	CGGTCATCTCTCCTTGCCATCATTACTTCAACTAAT		TGACACAGTTTCGTTTTTATTTGAATTGCATGGACCATATTCCC	26581419	26581498
454	00673-L00117	TRUE	CTNNB1	3	no	03-041.241033	03p22.1	GTTGGACATGGCCATGGAACC		AGACAGAAAAGCGGCTGTTAGTCACTGGCAG	41241033	41241085
463	11283-L12001	FALSE	BRCA1	13	no	17-038.487745	17q21.31	CCTTGGTTCTTGTCCCTGCTCACAACTAAT		ATACCAGTCAGAGGGACCCAAGGCAGTCATTCATGTTGTCATC	38487745	38487817

7.6.2. P090 BRCA2 A4-0712

Notes

* LPO is the 5' half of the probe, RPO is the 3' half of the probe.
* The two sequences are directly adjacent and together they form the complete hybridising sequence of the probe.
* A small number of probes consist of three oligonucleotides and have a SP (spanning oligo).
* The first nucleotides of the LPO may not be complementary to the target DNA and is part of the stuffer sequence.
* Locations are based on hg18 / mapview build 36. Mapview location consist of the chromosome no. + distance (nt) from the p-telomere to the start of the probe sequence.
* Observed length of the fragments depend on the software, sequencer type and molecular weigth markers used.
* Exon numbering might be different from the literature.
* Please let us know if you find SNPs under one of these sequences that affects the probe signal: info@mlpa.com

length	number	ref. probe	gene	genbank exon	mut. spec.	mapview	chr. pos.	LPO	SP	RPO	start	end
130	00797-L00463	TRUE	IL4	1	no	05-132.037610	05q31.1	CTACATTGTCACTGCAAATCGACACCTAT		TAATGGGTCTCACCTCCCAACTGCTTCCCCCT	132037610	132037671
137	02283-L12281	FALSE	BRCA2	1	no	13-031.787593	13q13.1	GGTCACGTGGCCAGCGCGGGCTT		GTGGCGCGAGCTTCTGAAACTAGGCGGCAGAGGCGGAG	31787593	31787655
148	02285-L13497	FALSE	BRCA2	1	no	13-031.787801	13q13.1	TGGGTTAGTGGTGGTGGTAGTGGGTT		GGGACGAGCGCGTCTTCCGCAGTCCCAGTCCAGCGTGG	31787801	31787866
154	09297-L08066	FALSE	BRCA2	14	no	13-031.827123	13q13.1	CCAGGACATCCATTTTATCAAGTTTCTGCTACAAGA		AATGAAAAAATGAGACACTTGATTACTACAGGCAGACCAACCAAAG	31827123	31827203
160	02143-L09586	FALSE	FRY	61	no	13-031.767521	13q13.1	GCATGATCCGCAGGGCCCAGAGTTA		CCGAGTCCTCACTACTTTTCTTCCAGACTCCAGTGTTTCT	31767521	31767582
166	02486-L01985	FALSE	BRCA2	2	no	13-031.788599	13q13.1	GCCTGCTTTGTTGCAGCGTGTCTTAA		AAATTTCAAAAAATGTTGGCCTCTCTTTGGATCCAATAGGC	31788599	31788665
172	08898-L09587	FALSE	BRCA2	3	no	13-031.791539	13q13.1	GTACATTCATTAATTGTGTCATGCTGGGCAAATCA		GTCTCTCTGGCCTCTTTTTCCTCACTCGAAAAATGGAGACGATG	31791539	31791614
178	01599-L10642	FALSE	BRCA2	3	no	13-031.791362	13q13.1	GTCAGCTGGCTTCAACTCCAATAATATTCAA		AGAGCAAGGGCTGACTCTGCCGCTGTACCAATCTCCTGTAAAAGAAT	31791362	31791438
184	01217-L00694	TRUE	KLKB1	3	no	04-187.390325	04q35.2	CTTCCATGTACACCCCAAATGCCCAATACT		GCCAGATGAGGTGCACATTCCACCCAAGGTGTTTGCTATTCAG	187390325	187390398
191	09812-L10643	FALSE	BRCA2	23	no	13-031.851892	13q13.1	TCTGAGTATTTGGCGTCCATCATCAGATTT		ATATTCTCTGTTAACAGAAGGAAAGAGATACAGAATTTATCATCTTG	31851892	31851969
197	01600-L04671	FALSE	BRCA2	4	no	13-031.797209	13q13.1	CGTTTCAGGAAGGAATGTTCCCAATAGTAGACAT		AAAAGTCTTCGCACAGTGAAAACTAAAATGGATCAAGCAGATGA	31797209	31797282

202	08265-L08128	FALSE	BRCA2	7	no	13-031.798677	13q13.1	GTGAGGTGGATCCTGATATGTCTTGGTCA		AGTTCTTTAGCTACACCACCCACCCTTAGTTCTACTGTGCT	31798677	31798748
211	02333-L01826	TRUE	PAH	10	no	12-101.762243	12q23.2	CTGCAAACAAGGAGACTCCATAAAGGCATA		TGGTGCTGGGCTCCTGTCATCCTTTGGTGAATTACAG	101762243	101762310
220	01602-L01184	FALSE	BRCA2	8	no	13-031.801290	13q13.1	CCCTCACAGCATCATCTGACTTTCCA		ACTCATTGTGGACAGTATTACCATAAAGTAATGATCACCAAGCCATA	31801290	31801362
229	01603-L01185	FALSE	BRCA2	9	no	13-031.803119	13q13.1	CTGTGACAGACAGTGAAAACACAAATCAA		AGAGAAGCTGCAAGTCATGGTAAGTCCTCTGTTTAGTTGAACTACAG	31803119	31803195
238	00517-L00097	TRUE	IL1RN	7	no	02-113.606702	02q13	GGCAGTTAACATCACTGACCTGAG		CGAGAACAGAAAGCAGGACAAGCGCTTCGCC	113606702	113606757
247	01604-L01186	FALSE	BRCA2	10	no	13-031.804736	13q13.1	GCCCTTTGAGAGTGGAAGTGACAAAA		TCTCCAAGGAAGTTGTACCGTCTTTGGCCTGTGAATGGTCTCAA	31804736	31804806
256	02279-L01770	FALSE	BRCA2	11	no	13-031.808425	13q13.1	GAAGCTGTTCACAGAATGATTCTGAAGAACCA		ACTTTGTCCTTAACTAGCTCTTTTGGGACAATTCTGAGGAA	31808425	31808498
265	02318-L01809	TRUE	LDLR	6	no	19-011.079136	19p13.2	GCAACATGGCTAGAGACTGCCGGG		ACTGGTCAGATGAACCCATCAAAGAGTGCGGTGAGTC	11079136	11079197
274	01606-L01188	FALSE	BRCA2	11	no	13-031.813228	13q13.1	CAAGTCATGCCACACATTCTCTTTTTACA		TGTCCCGAAAATGAGGAAATGGTTTTGTCAAATTCAAGAATTGGAAA	31813228	31813304
283	01607-L01189	FALSE	BRCA2	12	no	13-031.816481	13q13.1	GGTGTTTTAAAGTGGTCAAAACAGAACAAA		AATGTAATTGACATTGAAGACTGACTTTACTCTTTCAAACATTAGG	31816481	31816557
292	03018-L02458	TRUE	ACVRL1	6	no	12-050.594518	12q13.13	GGCTTGTGGCACGGTGAGAGTGT		GGCCGTCAAGATCTTCTCCTCGAGGGATGAACAGT	50594518	50594580
301	02280-L01771	FALSE	BRCA2	13	no	13-031.818973	13q13.1	CATGTCTTACCGAAAGGGTACACAGGTAA		TCGGCTCTAAAGAAACATGATGCATAAACAATCTTCGATCT	31818973	31819043
310	09809-L10257	FALSE	BRCA2	5	no	13-031.798240	13q13.1	CTGTTGTTCTACAATGTACACATGTAACACCACA		AAGAGATAAGTCAGGTATGATTAAAAACAATGCTTTTTATTCTTAGAA	31798240	31798321
319	09296-L11090	FALSE	BRCA2	27	no	13-031.871035	13q13.1	CCAAATTTACCTCAGCGTTTGTGTATCGGGCAAAAA		TCGTTTTGCCCGATTCCGTATTGGTATACTTTTGCTTCAG	31871035	31871101
326	01610-L01192	FALSE	BRCA2	15	no	13-031.828641	13q13.1	CCACAGCCAGGCAGTCTGTATCT		TGCAAAAACATCCACTCTGCCTCGAATCTCTCTGAAAGCAGCAGTAG	31828641	31828711
337	01611-L01193	FALSE	BRCA2	16	no	13-031.829982	13q13.1	GACTGGAAAAGGAATACAGTTGGCTGA		TGGTGGATGGCTCATACCCTCCAATGATGGAAAGGCTGGAAAAGAA	31829982	31830055
346	04585-L03983	FALSE	BRCA2	6	no	13-031.798360	13q13.1	CCAGGTAGAATATTTACCTTCACAAACTTTGGT		GTATGAAACAAACTCCCACATACCACTGGGGGTAAAAAAAGGG	31798360	31798436
355	02281-L01772	FALSE	BRCA2	17	no	13-031.834744	13q13.1	GAAGCACCCTTTCTGGGCTTAGGCATCTAT		TAGCAAATTCCTTAGGAAAGGCACATTCCATAGCTGCCAG	31834744	31834814
364	01613-L01195	FALSE	BRCA2	18	no	13-031.835565	13q13.1	GGCAGACTGACAGTTGGTCAGAAGATTAT		TCTTCATGGAGCAGAACTGGTGGGCTCTCCTGATGCCTGTACACCTC	31835565	31835641
373	02667-L04984	TRUE	ATM	23	no	11-107.655436	11q22.3	GCAGGACACGAAGGGAGATTCTTCCA		GGTTACTGAAAGCACTTCCTTTGAAGCTTCAGCAAACAGCT	107655436	107655500
382	01614-L01196	FALSE	BRCA2	19	no	13-031.842558	13q13.1	GCCTGCTCGCTGGTATACCAAACT		TGGATTCTTTCCTGACCCTAGACCTTTTCCTCTGCCCTTATCATCG	31842558	31842628
391	08266-L08129	FALSE	BRCA2	20	no	13-031.843090	13q13.1	CCAGTGGATGGAGAAGACATCATCTGGATTATA		CATATTTCGCAATGAAAGAGAGGAAGAAAAGGAAGCAGCA	31843090	31843161

400	00801-L00639	TRUE	DLEU1	2	no	13-049.576840	13q14.2	GATTCTACCAGAAAGGAATGAAGAACAGAAC		CTTCAGGAATTGAGTCACAATGCAGACAAATA	49576840	49576904
409	02069-L01970	FALSE	BRCA2	21	no	13-031.848828	13q13.1	CACCATCACGTGCACTAACAAGACAGCAA		GTTCGTGCTTTGCAAGATGGTGCAGAGCTTTATGAAGCAGTGAAGAA	31848828	31848903
418	01617-L01199	FALSE	BRCA2	22	no	13-031.851548	13q13.1	GGCCATGGAATCTGCTGAACAAAA		GGAACAAGGTTTATCAAGGGATGTCACAACCGTGTGGAAGTTGCGT	31851548	31851618
427	06942-L06522	TRUE	BEST1	10	no	11-061.486331	11q12.3	CCAATCAGGAGGACGAGGAGGAT		GCTCACGCTGGCATCATTGGCCGCTTCCTAGGCCTGCA	61486331	61486392
436	08267-L08130	FALSE	BRCA2	24	no	13-031.852214	13q13.1	CCTGTTTTTTTCACAACAGAAACGACAAAT		CCTATTAGGTCCACCTCAGAACAAGATGGCTGAAAGTCTG	31852214	31852283
445	08268-L08131	FALSE	BRCA2	25	no	13-031.867023	13q13.1	CCTGAGGGCCACTTTCAAGAGACATTCAA		CAAAATGAAAAATACTGTTGAGGTAAGGTTACTTTTCAGCATCACCA	31867023	31867097
454	02144-L01619	FALSE	N4BP2L1	3	no	13-031.879392	13q13.1	GAGGAATGGCATATCCCCCATTATTATTGA		TAATACCAACCTCCACGCCTGGGAAATGAAGCCCTATGCA	31879392	31879462
463	11984-L15346	FALSE	BRCA2	26	no	13-031.869059	13q13.1	CAGCAGAAAACAAGCTTATGCATATACTGCATGCAA		ATGATCCCAAGTGGTCCACCCCAACTAAAGACTGTACTTC	31869059	31869132
476	09293-L15678	FALSE	BRCA2	27	no	13-031.870386	13q13.1	CTCAGCCCAGATGACTTCAAAGTCTTGTAA		AGGGGAGAAAGAGATTGATGACCAAAAGAACTGCAAAAAGAGA	31870386	31870454
486	05028-L15679	TRUE	COL3A1	51	no	02-189.584681	02q32.2	CACCCTATGACATTGGTGGTCCTGATCAAGAATT		TGGTGTGGACGTTGGCCCTGTTTGCTTTTTATAAACCAAACTC	189584681	189584754

7.7. Appendix G – Individual patient results summary

Patients' results are listed below in a patient-by-patient format. *BRCA1* and *BRCA2* single nucleotide variants findings are presented together for each patient. Please note that several large exons within the *BRCA1* and *BRCA2* genes are amplified and sequenced several amplicons. The alphabetical letters after the exon numbers indicates the amplicon name.

7.7.1. BRCA001

BRCA001	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs80357192	c.2120G>A	Heterozygous	Missense	G/D	0.07	0.035	Benign	Unknown	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11J	rs2227945	c.3418A>G	Heterozygous	Missense	S/G	0.23	0.017	Benign	Unknown	No entry
BRCA001	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	Intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	
19	rs11571744	c.8487+47C>T	Heterozygous	Intronic	-	-	-	Unknown		

7.7.2. BRCA002

BRCA002	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA002	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
02	rs1799943	c.-26G>A	Heterozygous	Intronic	-	-	-	Unknown		CR068211; Disease causing mutation?
10B	rs144848	c.1114A>C	Heterozygous	Missense	N/H	0.13	0.116	Benign		
11E	rs1801406	c.3396A>G	Heterozygous	Synonymous	K/K	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
14B	rs1799955	c.7242A>G	Heterozygous	Synonymous	S/S	-	-	Benign	Clinically not important	
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	
22	rs4987047	c.8830A>T	Heterozygous	Missense	I/F	0	0.159	Unknown	CM050183; Disease Causing Mutation? ; conservec	

7.7.3. BRCA003

BRCA003	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs16940	c.2311T>C	Heterozygous	Synonmous	L/L	-	-	Benign	Not clinically significant	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11I	rs16941	c.3113A>G	Heterozygous	Missense	E/G	0.05	0.033	Unknown	Clinically insignificant/unknown	CM032861; Disease-associated polymorphism
11J	rs16942	c.3548A>G	Heterozygous	Missense	K/R	1	0.001	Benign	Not clinically significant	CM058359; Disease-associated polymorphism
11K	novel	c.3751T>G	Heterozygous	Missense	S/R	0.4	0.427	dicted Ben	No entry	No entry
13	rs1060915	c.4308T>C	Heterozygous	Synonmous	S/S	-	-	Benign	Clinically insignificant/unknown	No entry
16	rs1799966	c.4837A>G	Heterozygous	Missense	S/G	0.02	0.006	Unknown	Clinically insignificant/unknown	No entry for A>G; Only for A>T (CM053798)
BRCA003	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Homozygous	Synonymous	V/V	-	-	Benign	Clinically not important	
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
19	rs11571744	c.8487+47C>T	Heterozygous	intronic	-	-	-	Unknown		

7.7.4. BRCA004

BRCA004	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA004	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Homozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
19	rs11571744	c.8487+47C>T	Heterozygous	intronic	-	-	-	Unknown		

7.7.5. BRCA005

BRCA005	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11J	rs2227945	c.3418A>G	Heterozygous	Missense	S/G	0.23	0.017	dicted Ben	Unknown	No entry
BRCA005	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11E	rs1801406	c.3396A>G	Heterozygous	Synonymous	K/K	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11M	rs11571659	c.6412G>T	Heterozygous	Missense	V/F	0.03	0.01	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
14B	rs1799955	c.7242A>G	Heterozygous	Synonymous	S/S	-	-	Benign	Clinically not important	
16	rs81002819	c.7805+6C>G	Heterozygous	ntronic (Splice	-	-	-	Unknown		
17	rs9534262	c.7806-14T>G	Homozygous	Intronic	-	-	-	Unknown	Clinically not important	
26	rs11571823	c.9648+54G>A	Heterozygous	Intronic	-	-	-	Unknown		

7.7.6. BRCA006

BRCA006	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
07	novel	c.306T>G	Heterozygous	Synonymous	A/A	-	-	Benign	No entry	No entry
11G	rs799917	c.2612C>T	Heterozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
24	rs8176316	c.5468-10C>A	Heterozygous	Intronic	-	-	-	Unknown	Unknown	CS086718; Disease causing mutation?
BRCA006	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs80358765	c.5414A>G	Heterozygous	Missense	N/S	0.17	0.139	Benign		
11N	rs11571659	c.6412G>T	Heterozygous	Missense	V/P	0.03	0.01	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	
18A	rs80359052	c.8092G>A	Heterozygous	Missense	A/T	1	0.001	Unknown		
26	rs11571823	c.9648+54G>A	Heterozygous	Intronic	-	-	-	Unknown		
27C	rs1801426	c.10234A>G	Heterozygous	Missense	I/V	0.2	0.001	Unknown		

7.7.7. BRCA007

BRCA007	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
07	novel	c.428_431dulAp	Heterozygous	rtion/Duplica	.Asn144LysfsX1	-	-	Pathogenic	No entry	No entry
11F	rs16940	c.2311T>C	Heterozygous	Synonmous	L/L	-	-	Benign	Not clinically significant	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11I	rs16941	c.3113A>G	Heterozygous	Missense	E/G	0.05	0.033	Unknown	Clinically insignificant/unknown	CM032861; Disease-associated polymorphism
11J	rs16942	c.3548A>G	Heterozygous	Missense	K/R	1	0.001	Benign	Not clinically significant	CM058359; Disease-associated polymorphism
16	rs1799966	c.4837A>G	Heterozygous	Missense	S/G	0.02	0.006	Unknown	Clinically insignificant/unknown	No entry for A>G; Only for A>T (CM053798)
BRCA007	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
07	rs81002804	c.517-4C>G	Heterozygous	Intronic	-	-	-	Unknown		
08	novel	c.681+10T>G	Heterozygous	Intronic	-	-	-	Unknown		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs80358765	c.5414A>G	Heterozygous	Missense	N/S	0.17	0.139	Unknown		
11N	rs11571659	c.6412G>T	Heterozygous	Missense	V/F	0.03	0.01	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Homozygous	Intronic	-	-	-	Unknown	Clinically not important	
18A	rs80359052	c.8092G>A	Heterozygous	Missense	A/T	1	0.001	Unknown		
26	rs11571823	c.9648+54G>A	Heterozygous	Intronic	-	-	-	Unknown		

7.7.8. BRCA008

BRCA008	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11J	rs2227945	c.3418A>G	Heterozygous	Missense	S/G	0.23	0.017	Benign	Unknown	No entry
17	rs80358035	c.4987-20A>G	Heterozygous	Intronic	-	-	-	Unknown	Unknown	CS012668; Disease causing mutation?
BRCA008	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs55639415	c.5198C>T	Heterozygous	Missense	S/F	0.02	0.382	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.9. BRCA009

BRCA009	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Heterozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
15	rs55815649	c.4600G>A	Heterozygous	Missense	V/M	0.32	0.912	Unknown	Unknown	CM045533; Disease causing mutation?
BRCA009	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.10. BRCA010

BRCA010	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11D	rs111539978	c.1724A>G	Heterozygous	Missense	E/G	0.13	0.152	Benign	No entry	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA010	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Homozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.11. BRCA011

BRCA011	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA011	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
16	rs61754138	c.7626G>A	Heterozygous	Synonymous	T/T	-	-	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	
19	rs11571743	c.8487+19A>G	Heterozygous	Intronic	-	-	-	Unknown		

7.7.12. BRCA012

BRCA012	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA012	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
07	rs81002804	c.517-4C>G	Heterozygous	Intronic	-	-	-	Unknown		
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs80358765	c.5414A>G	Heterozygous	Missense	N/S	0.17	0.139	Unknown		
11M	rs11571659	c.6412G>T	Heterozygous	Missense	V/F	0.03	0.01	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	
19	rs11571744	c.8487+47C>T	Heterozygous	Intronic	-	-	-	Unknown		

7.7.13. BRCA013

BRCA013	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA013	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
02	rs76874770	c.-11C>T	Heterozygous	Intronic	-	-	-	Unknown		
11D	rs36060526	c.3264T>C	Heterozygous	Synonymous	P/P	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs34351119	c.5418A>G	Heterozygous	Synonymous	E/E	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>C	Homozygous	Intronic	-	-	-	Unknown		
27A	rs11571831	c.9730G>A	Heterozygous	Missense	V/I	0.41	0.007	Unknown		

7.7.14. BRCA014

BRCA014	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Heterozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA014	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
07	rs80358810	c.582G>A	Heterozygous	Stop Gained	W/*	-	-	Pathogenic	Clinically important	
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Homozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.15. BRCA015

BRCA015	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA015	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11E	rs1801406	c.3396A>G	Heterozygous	Synonymous	K/K	-	-	Benign	Clinically not important	
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
14B	rs1799955	c.7242A>G	Heterozygous	Synonymous	S/S	-	-	Benign	Clinically not important	
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.16. BRCA016

BRCA016	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA016	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
02	rs1799943	c.-26G>A	Heterozygous	Intronic	-	-	-	Unknown		CR068211; Disease causing mutation?
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.17. BRCA017

BRCA017	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11J	rs2227945	c.3418A>G	Heterozygous	Missense	S/G	0.23	0.017	Benign	Unknown	No entry
BRCA017	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
10B	rs144848	c.1114A>C	Heterozygous	Missense	N/H	0.13	0.116	Benign		
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
22	rs4987047	c.8830A>T	Heterozygous	Missense	I/F	0	0.159	Unknown		CM050183; Disease Causing Mutation?; Conserved
27C	rs1801426	c.10234A>G	Heterozygous	Missense	I/V	0.2	0.001	Unknown		

7.7.18. BRCA018

BRCA018	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11D	rs111539978	c.1724A>G	Heterozygous	Missense	E/G	0.13	0.152	Benign	No entry	No entry
11E	rs28897679	c.1971A>G	Heterozygous	Synonymous	Q/Q	-	-	Benign	Unknown	
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA018	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
07	rs81002804	c.517-4C>G	Heterozygous	Intronic	-	-	-	Unknown		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs80358765	c.5414A>G	Heterozygous	Missense	N/S	0.17	0.139	Unknown		
11M	rs11571659	c.6412G>T	Heterozygous	Missense	V/F	0.03	0.01	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Homozygous	Intronic	-	-	-	Unknown	Clinically not important	
26	rs11571823	c.9648+54G>A	Heterozygous	Intronic	-	-	-	Unknown		

7.7.19. BRCA019

BRCA019	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11E	rs55932871	c.1972A>G	Heterozygous	Missense	M/V	0.39	0.023	Benign	No entry	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA019	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign	Clinically not important	
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		

7.7.20. BRCA020

BRCA020	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs1799949	c.2082C>T	Heterozygous	Synonymous	S/S	-	-	Benign	Not clinically important	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11I	rs16941	c.3113A>G	Heterozygous	Missense	E/G	0.05	0.033	Unknown	Clinically insignificant/unknown	CM032861; Disease-associated polymorphism
11J	rs16942	c.3548A>G	Heterozygous	Missense	K/R	1	0.001	Benign	Not clinically significant	CM058359; Disease-associated polymorphism
13	rs1060915	c.4308T>C	Heterozygous	Synonymous	S/S	-	-	Benign	Not clinically important/unknown	No entry
16	rs1799966	c.4837A>G	Heterozygous	Missense	S/G	0.02	0.006	Unknown	Clinically insignificant/unknown	No entry for A>G; Only for A>T (CM053798)
BRCA020	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11E	rs1801406	c.3396A>G	Heterozygous	Synonymous	K/K	-	-	Unknown	Clinically not important	
11H	rs206075	c.4563A>G	Heterozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Heterozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Heterozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
14B	rs1799955	c.7242A>G	Heterozygous	Synonymous	S/S	-	-	Benign		

7.7.21. BRCA021

BRCA021	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA021	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11H	rs2219594	c.4681C>A	Heterozygous	Missense	H/N	0.25	0.159	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
11N	rs11571659	c.6412G>T	Heterozygous	Missense	V/F	0.03	0.01	Unknown		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
16	rs55689095	c.7712A>G	Heterozygous	Missense	E/G	0	0.874	cted Pathogenic		c.7711G>A(E/K) prostate cancer; CM118438
16/2	rs81002819	c.7805+6C>G	Heterozygous	Intronic	-	-	-	Unknown		
17	rs9534262	c.7806-14T>C	Homozygous	Intronic	-	-	-	Unknown		
19	rs81002838	c.8487+8G>A	Heterozygous	Intronic	-	-	-	Unknown		
26	rs11571823	c.9648+54G>A	Heterozygous	Intronic	-	-	-	Unknown		

7.7.22. BRCA022

BRCA022	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA022	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
07	rs11571623	c.517-19C>T	Heterozygous	Intronic	-	-	-	Unknown		
10B	rs144848	c.1114A>C	Heterozygous	Missense	N/H	0.13	0.116	Benign		
11H	rs206075	c.4563A>G	Heterozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Heterozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Heterozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Heterozygous	Missense	V/A	1	0	Benign		
20	rs201392123	2+131_8632+13	Heterozygous	Intronic	-	-	-	Unknown		

7.7.23. BRCA023

BRCA023	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Heterozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA023	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
02	rs1799943	c.-26G>A	Heterozygous	Intronic	-	-	-	Unknown		CR068211; Disease causing mutation?
11D	rs36060526	c.3264T>C	Heterozygous	Synonymous	P/P	-	-	Benign		
11E	rs1801406	c.3396A>G	Heterozygous	Synonymous	K/K	-	-	Benign	Clinically not important	
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs34351119	c.5418A>G	Heterozygous	Synonymous	E/E	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14A	rs45574331	c.7017G>C	Heterozygous	Missense	K/N	0.04	0.003	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Homozygous	Intronic	-	-	-	Unknown	Clinically not important	
27A	rs11571831	c.9730G>A	Heterozygous	Missense	V/I	0.41	0.007	Benign		

7.7.24. BRCA024

BRCA024	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA024	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign	Clinically not important	
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11H	rs2219594	c.4681C>A	Heterozygous	Missense	H/N	0.25	0.159	Benign		
11M	rs11571659	c.6412G>T	Heterozygous	Missense	V/F	0.03	0.01	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
14B	rs81002852	c.7435+6G>A	Heterozygous	Intronic	-	-	-	Unknown		
16	rs81002819	c.7805+6C>G	Heterozygous	Intronic	-	-	-	Unknown		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	
26	rs11571823	c.9648+54G>A	Heterozygous	Intronic	-	-	-	Unknown		
27C	rs1801426	c.10234A>G	Heterozygous	Missense	I/V	0.2	0.001	Benign		

7.7.25. BRCA025

BRCA025										
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA025	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
10B	rs144848	c.1114A>C	Heterozygous	Missense	N/H	0.13	0.116	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
14B	rs81002852	c.7435+6G>A	Heterozygous	Intronic	-	-	-	Unknown		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	
22	rs4987047	c.8830A>T	Heterozygous	Missense	I/F	0	0.159	Unknown		CM050183; Disease Causing Mutation?; Conserved

7.7.26. BRCA026

BRCA026	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA026	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.27. BRCA027

BRCA027	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA027	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
19	rs11571744	c.8487+47C>T	Heterozygous	Intronic	-	-	-	Unknown		

7.7.28. BRCA028

BRCA028	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA028	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs80358765	c.5414A>G	Heterozygous	Missense	N/S	0.17	0.139	Benign		
11M	rs11571659	c.6412G>T	Heterozygous	Missense	V/F	0.03	0.01	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
12	rs11571673	c.6842-73T>A	Heterozygous	Intronic	-	-	-	Unknown		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>C	Homozygous	Intronic	-	-	-	Unknown		
18A	rs80359052	c.8092G>A	Heterozygous	Missense	A/T	1	0.001	Unknown		
26	rs11571823	c.9648+54G>A	Heterozygous	Intronic	-	-	-	Unknown		

7.7.29. BRCA029

BRCA029	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA029	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Homozygous	Synonymous	V/V	-	-	Benign	Clinically not important	
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
27C	rs1801426	c.10234A>G	Homozygous	Missense	I/V	0.2	0.001	Benign		

7.7.30. BRCA030

BRCA030	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA030	BRCA2									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
02	rs76874770	c.-11C>T	Heterozygous	Intronic	-	-	-	Unknown		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>G	Heterozygous	Intronic	-	-	-	Unknown	Clinically not important	

7.7.31. BRCA031

BRCA031	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11D	rs111539978	c.1724A>G	Heterozygous	Missense	E/G	0.13	0.152	Benign	No entry	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
BRCA031	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
02	rs76874770	c.-11C>T	Heterozygous	5-UTR	-	-	-	Unknown		
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11J	rs55639415	c.5198C>T	Heterozygous	Missense	S/F	0.02	0.382	Unknown		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	Intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		

7.7.32. BRCA032

BRCA032	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
12	rs147448807	c.4113G>A	Heterozygous	Synonymous	G/G	-	-	Benign	No entry	No entry
21	Novel	c.5332+78C>T	Heterozygous	Intronic	-	-	-	Unknown	No entry	No entry
BRCA032	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11F	rs543304	c.3807T>C	Heterozygous	Synonymous	V/V	-	-	Benign		
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
13	rs206080	c.6938-120T>C	Homozygous	Intronic	-	-	-	Unknown		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
17	rs9534262	c.7806-14T>C	Heterozygous	Intronic	-	-	-	Unknown		
27C	rs1801426	c.10234A>G	Heterozygous	Missense	I/V	0.2	0.001	Benign		

7.7.33. BRCA033

BRCA033	BRCA1									
Exon	Variants	HGVS: NM_007294.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
05	Novel	c.212+66A>G	Heterozygous	Intronic	-	-	-	Unknown	No entry	No entry
11G	rs799917	c.2612C>T	Homozygous	Missense	P/L	1	0	Benign	Clinically insignificant/unknown	CM096315; Disease-associated polymorphism
11J	rs16942	c.3548A>G	Heterozygous	Missense	K/R	1	0.001	Benign	Not clinically significant	CM058359; Disease-associated polymorphism
16	rs1799966	c.4837A>G	Heterozygous	Missense	S/G	0.02	0.006	Unknown	Clinically insignificant/unknown	No entry for A>G; Only for A>T (CM053798)
22	rs55946644	c.5406+8T>C	Heterozygous	Intronic	-	-	-	Benign	Not clinically significant	No entry
BRCA033	**BRCA2**									
Exon	Variants	HGVS: NM_000059.3	State	Mutation Class	Protein Change	SIFT	PolyPhen	Mutation Class	BIC	HGMD
11H	rs206075	c.4563A>G	Homozygous	Synonymous	L/L	-	-	Benign		
11N	rs206076	c.6513G>C	Homozygous	Synonymous	V/V	-	-	Benign		
14B	rs169547	c.7397C>T	Homozygous	Missense	V/A	1	0	Benign		
24	rs55933907	c.9235G>A	Heterozygous	Missense	V/I	0.87	0.635	Unknown		