INVESTIGATING KNOWN CANCER SUSCEPTIBILITY GENES IN BLACK SOUTH AFRICAN BREAST CANCER INDIVIDUALS

Reabetswe Pitere

A Dissertation submitted to the Faculty of Health Sciences, University of the Witwatersrand, in fulfilment of the requirements for the degree of Master of Science (Medicine) in Human Genetics

Johannesburg, 2018

DECLARATION

I, Reabetswe Pitere, declare that this Dissertation is my own, unaided work. It is being submitted for the degree Master of Science (Medicine) in Human Genetics at the University of the Witwatersrand, Johannesburg. It has not been submitted before for any degree or examination at any other University.

Reabetswe Pitere

06th day of September 2018 in Johannesburg

DEDICATION

I dedicate this work to my parents, Retlaroma and Motoloki Pitere. Thank you for sacrificing your all for me to get to this point. Thank you for your unfailing love, the words of encouragement, the great lessons, the constant support and for being my pillars of strength at all times.

To my little brother, Mpho and my daughter Olerato, thank you for reminding me to smile at the small things in life and bringing out the best in me. I hope that this will encourage you to always strive for success and reach your ultimate goals irrespective of the challenges you may come across. To my partner Musa, thank you for walking this road with me and for being a constant cheerleader and a shoulder to cry on.

To the rest of my family and friends, thank you all.

PRESENTATIONS ARISING FROM THIS RESEARCH PROJECT

1. Reabetswe Pitere, Tasha Wainstein, Fiona Baine, Amanda Krause. "***Investigating cancer susceptibility genes in black South African breast cancer individuals***". Oral presentation at 17th Biennial Congress of the Southern African Society for Human Genetics (Durban, 13th-16th August, 2017).

2. Reabetswe Pitere, Tasha Wainstein, Fiona Baine, Amanda Krause. "***Investigating known cancer susceptibility genes in black South African breast cancer patients***". Oral presentation at the Molecular Biosciences Research Thrust Annual Research Day (University of the Witwatersrand, Johannesburg, 8th December 2016).

3. Reabetswe Pitere, Tasha Wainstein, Fiona Baine, Amanda Krause. "***Investigating known cancer susceptibility genes in black South African breast cancer patients***". Poster presentation at the Wits Faculty of Health Sciences Research Day (University of the Witwatersrand, Johannesburg, 1st September, 2016).

4. Reabetswe Pitere, Tasha Wainstein, Fiona Baine, Amanda Krause. "***Investigating known BRCA2 mutations in a cohort of black South African breast cancer families***". Poster presentation at the 16th Biennial Congress of the Southern African Society for Human Genetics (Pretoria, 16th-19th August, 2015).

ABSTRACT

Breast cancer is an increasingly common cause of morbidity and mortality in black South African women. Of all diagnosed cases of breast cancer in European populations, approximately 5-10% of these arise due to an inherited mutation in a cancer susceptibility gene. *BRCA1* and *BRCA2* gene mutations are the primary contributors to inherited breast cancer (IBC). However, mutations in other high, moderate and low susceptibility genes have also been identified. Previous studies indicate that approximately 10% of young black South Africans with breast cancer have a deleterious mutation in either *BRCA1* or *BRCA2* gene**.** It would, therefore, be pertinent to determine what is contributing to disease in the remainder of the young high-risk black South African breast cancer patients by investigating other genes which are known to confer cancer susceptibility.

In addition to a breast cancer syndrome, biallelic *BRCA2* mutations can also result in a Fanconi anaemia (FA) phenotype. Previous studies have identified overlapping *BRCA2* mutations, c.582G>A and c.5771_5774delTTCA, in patients with either breast cancer or FA. Approximately 80% of black South African patients with FA are found to be homozygous for the c.637_643delTACCGCC mutation in the *FANCG* gene. The relationship between this *FANCG* mutation and breast cancer in the black South African population has not been previously determined.

The main aim of the study was to increase current knowledge of the molecular basis of breast cancer in the black South African population. The study initially focussed on genotyping the two *BRCA2* mutations (c.582G>A and c.5771_5774delTTCA) in a sample of black South African women with breast cancer to determine the frequencies of these proposed common mutations in this population. This was followed by the construction of haplotypes using the *BRCA2* mutations to evaluate the presence of a founder effect for each mutation. The study also aimed to screen for the *FANCG* c.637_643delTACCGCC mutation in breast cancer patients to determine its role in breast cancer development.

Finally, screening of 26 cancer susceptibility genes in a targeted gene panel was undertaken in breast cancer patients who had previously tested negative for mutations in the *BRCA* genes.

A total of thirty-five patients were successfully genotyped for two common *BRCA2* mutations; one patient (2.86%) tested positive for the c.5771_5774delTTCA mutation and none of the patients (0%) tested positive for the c.582G>A mutation. Using haplotype analysis, the presence of a founder effect for these two mutations could not be determined due to inconclusive evidence.

One hundred and nineteen (119) patients were genotyped for the *FANCG* mutation and they all tested negative for the mutation. Eighteen patients were screened using the targeted gene panel and three deleterious mutations were identified in the *ATM*, *BARD1* and *RAD50* genes. A total of 248 benign polymorphisms as well as 308 variants of uncertain significance were also identified in all of the genes tested.

The results of this study highlight the importance of local research to determine population-specific variation which can subsequently be used to tailor population-specific testing strategies. Additionally, the findings from the screening intensify the need to move from single-gene testing to multiple-gene testing in the black South African population. These results will positively impact on the management of patients who have been found to have deleterious mutations and can also impact on our ability to prevent cancer diagnoses in at-risk relatives through increased disease surveillance. The research will contribute to an improved genetic counselling service with more information which can be provided to current and future patients. Hopefully with additional research, targets for therapeutics will be identified and long-term management and prevention strategies of breast cancer can be implemented.

ACKNOWLEDGEMENTS

A very heartfelt thank you to the following people and organisations:

1. The patients. Without your willingness to participate, this research would have never been possible. I hope that the results from this study will one day aid in alleviating the disease you and your families are burdened with.
2. My supervisors: Ms Tasha Wainstein, Dr Fiona Baine-Savanhu and Prof Amanda Krause for your unlimited support, guidance, patience and encouragement through the trials, errors and the ultimate success of the project.
3. University of the Witwatersrand for funding me and the project through the Postgraduate Merit Award and Faculty Research Committee grant.
4. National Research Foundation for the Scarce Skills Master's Scholarship which enabled me to complete my studies;

   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.
5. AstraZeneca Pharmaceuticals for a research grant which enabled me to perform next generation sequencing and immensely contributed towards the success of this project.
6. Whitehead Scientific for providing us with a discount on the next generation sequencing kit used in the study and for the travel grant which enabled me to present my findings at a local conference.
7. The Wits University Margot Lachmann travel grant which enabled me to attend a next generation sequencing course at the Wellcome Genome Campus in Cambridge, United Kingdom. The course equipped me with both the theoretical and practical skills I required for the project.
8. Ms Roshilla Ramdin at Whitehead Scientific for her availability and constant support throughout the next-generation sequencing phase of the study.
9. Dr Nadia Carstens for her bioinformatics training and support which greatly contributed to the analysis of the next generation sequencing data.

10. Drs Senzo Mtshali and Arshad Ismail at the NICD for kindly allowing us to make use of their infrastructure whenever it was required.
11. Molecular Diagnostics laboratory staff that were always available to assist with laboratory-related queries and provided access to patient samples.
12. Clinical genetics staff who kindly allowed me to make use of patient files.
13. Thandiswa, Carl, Jorge, Mahtaab and Liesl for always being willing to assist with all software, technical and theoretical queries I had.
14. Ms Mirriam Hlongwane who was always available to assist with administration-related queries.
15. Postgraduate students. Thank you kindly for all the support, motivation and endless laughter.
16. To everyone at the Wits/NHLS Human Genetics Division. Thank you.

TABLE OF CONTENTS

DECLARATION i

DEDICATION ii

PRESENTATIONS ARISING FROM THIS RESEARCH PROJECT iii

ABSTRACT iv

ACKNOWLEDGEMENTS vi

LIST OF FIGURES xiii

LIST OF TABLES xiv

ABBREVIATIONS xvi

GLOSSARY xix

1 INTRODUCTION 1

1.1 An overview of breast cancer 1

1.2 The genetic aetiology of inherited breast cancer 4

1.2.1 The *BRCA1* gene 6

1.2.2 The *BRCA2* gene 8

1.2.3 Other breast cancer susceptibility genes 9

1.3 Molecular genetic testing for individuals at high-risk for developing breast cancer 14

1.4 Breast cancer in African populations 15

1.5 Breast cancer in the South African context 18

1.5.1 Inherited Breast Cancer in the Afrikaner population 18

1.5.2 Inherited Breast Cancer in the Ashkenazi Jewish population 19

1.5.3 Inherited Breast Cancer in the Black South African and Mixed Ancestry Population 20

1.6 Rationale for study 22

1.6.1 The *BRCA2* c.5771_5774delTTCA and c.582G>A mutations 23

1.6.2 The *FANCG* c.637_643delTACCGCC mutation 24

1.6.3 Next generation sequencing 24

1.7 Aims and objectives 24

2 SUBJECTS AND METHODS 26

2.1 Study participants 26

2.2 Quantitative and qualitative analysis of DNA 29

2.3 Screening for the *BRCA2* founder mutations 29

2.3.1 *BRCA2* c.5771_5774delTTCA mutation 29

2.3.2 *BRCA2* c.582G>A mutation 32

2.4 Haplotype analysis for the *BRCA2* c.5771_5774delTTCA and c.582G>A mutations 34

2.4.2 SNP genotyping 38

2.4.3 Hardy-Weinberg equilibrium 41

2.4.4 Linkage disequilibrium 42

2.4.5 Haplotype analysis using PHASE 42

2.5 Screening for the *FANCG* c.637-643delTACCGCC mutation 43

2.6 Next generation sequencing 44

2.6.1 DNA purification of genomic DNA 47

2.6.2 Library preparation 47

2.6.3 Bridge amplification and sequencing-by-synthesis 51

2.6.4 Alignment and variant calling 52

3 RESULTS 55

3.1 Mutation screening of the *BRCA2* c.5771_5774 and c.582G>A mutations 55

3.2 Haplotype analysis 57

3.2.1 Hardy-Weinberg equilibrium 58

3.2.2 Linkage disequilibrium 59

3.2.3 Haplotype analysis 59

3.3 Mutation screen of the *FANCG* c.637_643delTACCGCC mutation 61

3.4 Next generation sequencing 62

3.4.1 Library preparation 63

3.5 MiSeq sequencing results for the Multiplicom BRCA Hereditary Cancer MASTR™ Plus kit 65

3.5.1 Previously described Class I and Class II Variants 68

3.5.2 Previously described Class IV and V Variants 68

3.5.3 Class III Variants 69

3.6 Identification of variants in African Genome database 79

3.7 Variant verification using Sanger sequencing 80

4 DISCUSSION 82

4.1 *BRCA2* mutation screening 82

4.2 Haplotyping analyses of the *BRCA2* region 84

4.3 *FANCG* c.637-643TACCGCC mutation screen 86

4.4 Next generation sequencing: multi-gene panel testing 87

4.4.1 Previously described Class I and II variants 88

4.4.2 Previously described Class IV and V variants 89

4.4.3 Class III variants 92

4.4.4 Interesting case identified from the gene panel screening 97

4.4.5 Appropriateness of the gene panel in the black population 98

4.5 Contribution of study to participants and their families 98

4.6 Limitations of study 99

4.7 Future prospects 100

4.7.1 Functional and segregation analysis 100

4.7.2 Next generation sequencing 101

4.7.3 Epigenetic studies 103

4.7.4 Genome-wide association studies 103

4.7.5 Environmental and genetic contributors to breast cancer 104

5 CONCLUSION 105

6 REFERENCES 107

6.1 Journal articles 107

6.2 Books 125

6.3 Electronic sources 125

7 APPENDICES 127

Appendix A: Ethics clearance certificate 127

Appendix B: Consent form 128

Appendix C: Recipes for laboratory solutions 129

Appendix D: Protocols 131

1. The *BRCA2* c.5771_5774delTTCA mutation PCR protocol and conditions 131

2. The *BRCA2* c.582G>A mutation PCR protocol and conditions 132

3. Microsatellite and *FANCG* 7 bp mutation genotyping protocol and conditions 133

4. SNP genotyping 134

Appendix E: File extraction and input of genotypes for PHASE haplotyping software 136

Appendix F: Multiplicom protocol 138

Appendix G: Assay specification sheet 148

Appendix H: Genotyping results for the SNP and microsatellite screening 149

Appendix I: Haplotypes generated by PHASE software 154

Appendix J: Quality metrics obtained from the Illumina MiSeq instrument 158

Appendix K: Benign and likely benign variants... 159

Appendix L: Novel variants 170

Appendix M: ExAC reported variants 174

Appendix N: Variants previously reported in 1000 Genomes and dbSNP 177

Appendix O: Variants previously reported in dbSNP, ExAC and 1000 Genomes 189

LIST OF FIGURES

Figure 1-1: The distribution of breast cancer incidence and mortality rates worldwide. 3

Figure 1-2: The genetic contribution to breast cancer cases. 5

Figure 1-3: A schematic representing the functional domains of the BRCA1 protein. 7

Figure 1-4: A schematic representing the functional domains of the BRCA2 protein. 9

Figure 1-5: The Fanconi anaemia/BRCA pathway. 13

Figure 2-1: An agarose gel electrophoresis image illustrating the restriction digest fragments for the *BRCA2* c.5771_5774delTTCA mutation. 31

Figure 2-2: An agarose gel electrophoresis image illustrating possible outcomes of the ARMS-PCR for the *BRCA2* c.582G>A mutation. 34

Figure 2-3: The position of five markers selected within and flanking the *BRCA2* gene in respect to positions of the *BRCA2* c.5771_5774delTTCA and c.582G>A mutations........... 35

Figure 2-4: An agarose gel image following digest using the *MluCI* enzyme to detect the alleles of the rs9590940 SNP. 41

Figure 2-5: An agarose gel image showing the purified amplicon libraries. 49

Figure 3-1: Capillary electropherograms illustrating the presence and absence of the *FANCG* c.637_643delTACCGCC mutation. 62

Figure 3-2: Image of a Bioanalyzer trace indicating the the internal control. 64

Figure 3-3: Example image of a Bioanalyzer trace for a successful library. 64

Figure 3-4: Example image of a Bioanalyzer trace for an unsuccessful library. 65

Figure 3-5: A sequencing chromatogram illustrating the forward sequence region of the *PTEN* gene containing the supposed c.802A>TT splice-site variant.................................. 80

Figure 3-6: A sequencing chromatogram illustrating the reverse sequence region of the *PTEN* gene containing the supposed c.802A>TT splice-site variant.................................. 81

LIST OF TABLES

Table 1-1: Pathogenic germline mutations identified in the black South African population group using next generation sequencing and MLPA techniques (Francies *et al*., 2015) 21

Table 2-1: The number and ethnicity of participants in each study objective 28

Table 2-2: Oligonucleotide sequences used to amplify the *BRCA2* region to detect the presence of the c.5771_5774deTTCA mutation 30

Table 2-3: Oligonucleotide sequences used to amplify the *BRCA2* region to detect the presence of the c.582G>A mutation 32

Table 2-4: Oligonucleotide sequences used to amplify the region containing the 17xAT microsatellite upstream of the *BRCA2* gene 36

Table 2-5: Oligonucleotide sequences used to amplify the region containing the 21xTG microsatellite downstream of the *BRCA2* gene 36

Table 2-6: Single nucleotide polymorphisms selected for the haplotype study of the *BRCA2* gene 39

Table 2-7: Oligonucleotide sequences used to amplify the region containing the single nucleotide polymorphism with ID rs9590940 (Chen, 2016) 40

Table 2-8: Oligonucleotide sequences used to amplify the region containing the *FANCG* c.637_643delTACCGCC mutation (Morgan *et al*., 2005). 43

Table 2-9: The 26 cancer susceptibility genes included in the Multiplicom panel for inherited breast cancer. 45

Table 2-10: The proposed classification system for sequence variants identified in genetic testing by the International Agency for Research on Cancer (Plon et al., 2008). 54

Table 3-1: A summary of the *BRCA2* c.5771_5774delTTCA and/or c.582G>A mutations identified in the breast cancer and Fanconi anaemia studies conducted in the Division of Human Genetics (Wits/NHLS) 56

Table 3-2: A summary of the genotypes observed for the three *BRCA2* single nucleotide polymorphisms in the control group (n=100) 58

Table 3-3: The frequencies of observed and expected heterozygotes, and Hardy-Weinberg equilibrium p-values computed using gPLINK for each SNP located in the *BRCA2* gene and selected for haplotyping. 58

Table 3-4: The haplotypes constructed by the PHASE software for the mutation carriers of the *BRCA2* c.582G>A mutation........59
Table 3-5: The haplotypes constructed by the PHASE software for patients with the BRCA2 c.5771_5774delTTCA mutation........60
Table 3-6: Shared haplotypes observed in either the *BRCA2* c.5771-5774delTTCA or c.582G>A mutation-positive individuals and control samples........61
Table 3-7: The total number of variants observed in each gene, for each of the 18 samples sequenced........67
Table 3-8: Three previously reported pathogenic mutations in known inherited cancer susceptibility genes which were identified in this study using a next generation sequencing approach in three different individuals........68
Table 3-9: The sub-classification of the various variants of uncertain significance identified in this study........70
Table 3-10: The sixteen variants of uncertain significance that were identified in this study in inherited cancer susceptibility genes and evidence in support of their pathogenicity in the form of minor allele frequencies and *in silico* modelling outputs........71
Table 3-11: The twenty-one variants which have been classified as having conflicting interpretations of pathogenicity were identified in this study........73
Table 3-12: Novel sequence variants predicted to be possibly disease causing by three or more in *silico* prediction tools........76
Table 3-13: A total of 49 variants which were only reported on ExAC were identified in this study........77
Table 3-14: The eight variants selected and searched in African Genome Variation Project (AGVP) databases to determine their frequencies in African population/s........79

ABBREVIATIONS

%	Percent
µl	Microliters
ARMS-PCR	Amplification-Refractory Mutation System PCR
ATM	Ataxia Telangiectasia Mutated gene
BARD1	BRCA1 Associated RING Domain 1 gene
BIC	Breast Cancer Information Core
BLAST	Basic Local Alignment Search Tool
Bp	Base pair
BRCA1	Breast Cancer 1 gene
BRCA2	Breast Cancer 2 gene
BRCT	BRCA1 C terminus
BRIP1	*BRCA1* Interacting Protein C-Terminal Helicase 1 gene
°C	Degrees Celsius
CDH1	Cadherin 1, Type 1 gene
cDNA	Complementary DNA
CE	Capillary electrophoresis
CHBAH	Chris Hani Baragwanath Academic Hospital
CHEK1	Checkpoint Kinase 1 gene
CHEK2	Checkpoint Kinase 2 gene
CMJAH	Charlotte Maxeke Johannesburg Academic Hospital
CNV	Copy number variation
ddH_2O	Double distilled water
DNA	Deoxyribonucleic Acid
EtBr	Ethidium Bromide
ExAC	Exome Aggregation Consortium
FA	Fanconi Anaemia
FANCD1	Fanconi Anaemia Complementation group D1 gene
FANCG	Fanconi Anaemia Complementation group G gene

FGFR2	Fibroblast growth factor receptor 2 gene
gDNA	Genomic DNA
Her2	Human epidermal growth factor receptor 2
HJH	Helen Joseph Hospital
HLA	Human Leukocyte Antigen
HREC	Human Research Ethics Committee
IBC	Inherited Breast Cancer
Kb	Kilobase
MAF	Minor Allele Frequency
MIDs	Molecular identifiers
MRE11A	Meiotic recombination 11 homolog A gene
mRNA	messenger Ribonucleic Acid
NCBI	National Center for Biotechnology Information
NCR	National Cancer Registry
Ng	Nanogram
NGS	Next Generation Sequencing
NHLS	National Health Laboratory Services
OMIM	Online Mendelian Inheritance in Man
PALB2	Partner and localizer of *BRCA2* gene
PCR	Polymerase Chain Reaction
PTEN	Phosphatase and Tensin Homolog gene
RAD50	RAD50 gene
RE	Restriction endonuclease
RFLP	Restriction fragment length polymorphism
RING	Really Interesting New Gene
SNP	Single Nucleotide Polymorphism
STK11	Serine/threonine kinase 11 gene
TBE	Tris-Borate-Ethylenediaminetetraacetic acid
T_m	Melting temperature
TNRC9	Trinucleotide-repeat-containing 9 gene

TP53	Tumour Protein 53 gene
TS	Tumour suppressor
UV	Ultraviolet
V	Volts
VUS	Variants of uncertain significance
WES	Whole exome sequencing
WGS	Whole genome sequencing
WHO	World Health Organization
WITS	University of the Witwatersrand

GLOSSARY

Positional cloning - A technique used to identify a disease-causing gene based on its genomic location on a chromosome. This is performed by looking at linkage markers whose location is known in the chromosome. Following identification of the candidate gene, mutation analysis in affected individuals is performed to identify the relevant mutation/s.

Pseudogenes – Genomic sequences that resemble known genes but do not produce a functional protein

Tumour suppressor gene – A gene whose function is to suppress uncontrolled cell growth and proliferation to prevent tumour development

Alu sequences – Interspersed, DNA repeat sequences

Biallelic mutations – Mutations that are present on both copies of a specific gene.

Founder mutations – Mutations that are present in a small group of individuals that has migrated away from a larger group such that the deleterious mutations become increasingly prevalent in the smaller group

Triple-negative breast cancer – breast cancer which is negative for the expression of oestrogen and progesterone receptors and human epidermal growth factor receptor 2 (Her2) expansions

Hamartoma tumours – Benign tumours which are composed of a mixture of cells and tissues which are found at the site of the growth

Polymerase chain reaction – A technique used to produce multiple copies of a particular region of DNA

Restriction fragment length polymorphism – A variation in a homologous DNA sequence that can be detected by using restriction endonucleases which recognise a specific pattern of DNA sequence and produces fragments of varying lengths. The fragments can be used

to define the genotype of a polymorphism which creates or abolishes a restriction enzyme site.

Amplification Refractory Mutation System-PCR - A technique used to detect single nucleotide polymorphisms in a DNA sequence; it can also be used for the detection of small deletions and single nucleotide changes. Four primers are used in the experiment; the two outer primers are used to amplify the target sequence while the two inner primers are specific to the SNP or small deletion of interest

Capillary electrophoresis – A high-throughput technique which makes use of an electric field to separate molecules based on their mobility and size when voltage is applied. The technique is used in the process of genotyping by separating fragments of different sizes which can subsequently be analysed.

Sequencing adaptors - short oligonucleotides which bind to the flow cell and contain binding sites so that the unique sample barcodes and sequencing primers can be bound and identified

Molecular identifiers (MIDs) - unique barcodes assigned to each DNA sample so that the samples can be differentiated from each other during data analysis

1 INTRODUCTION

This chapter will provide a discussion of breast cancer and briefly elaborate on the incidence rates of the disease with specific reference to the African context. This will be outlined in Section 1.1. Section 1.2 will provide a detailed explanation of inherited breast cancer (IBC) and how different genes are involved in increasing susceptibility to disease. The genes to be discussed in this chapter are either involved in repair of deoxyribonucleic acid (DNA) or the regulation of the cell cycle. Some of the genes to be discussed are involved in the Fanconi Anaemia/Breast cancer (FA-BRCA) pathway which is responsible for the repair of damaged DNA. When mutations occur in this pathway, one of two phenotypes can be observed; Fanconi anaemia or breast cancer. The distinction between the disorders and how the same genes can give rise to the different disorders will be discussed below. The section will contextualise IBC in the black South African population. The chapter will then conclude with a rationale for conducting this project and its importance.

1.1 An overview of breast cancer

Breast cancer is classified as a complex or multifactorial disease which arises as a result of associations between genetic and external contributors (Zhang *et al*., 2012). External factors which can interact with an individual's genetic material to cause cancer may include biological carcinogens (e.g. viruses), physical carcinogens (e.g. ultraviolet radiation) or chemical carcinogens (e.g. benzene) (Lawson & Heng, 2010; Schwarzman *et al*., 2015; World Health Organization [WHO], 2015). The interaction of carcinogens with genetic material can lead to the transformation of a normal cell into a tumorigenic cell. The tumorigenic cell grows and divides but never reaches apoptosis. Many of these cells develop and eventually form a mass of tissue that is known as a tumour. The tumour becomes malignant, begins to grow uncontrollably and can spread to other sites (Nussbaum *et al*., 2007; Strachan & Read, 2010).

Breast cancer is one of the leading causes of female morbidity and death worldwide with approximately 1.67 million women diagnosed in 2012 (Ferlay *et al*., 2015). Estimates show that approximately 522 000 deaths were reported globally in 2012 (Ferlay *et al*., 2015). African countries are reported to have the lowest incidence rates of breast cancer in comparison to other parts of the world as observed in Figure 1-1 (World Health Organization [WHO], 2016). The lower reported incidences could be due to a scarcity of epidemiological studies, under-diagnosis or protective environmental and genetic factors (Oluwagbemiga *et al*., 2012). Additionally, in African countries, cancers are not detected at earlier stages due to a lack of proper health care facilities and a scarcity of early detection programmes (Ferlay *et al*., 2015; Ginsburg *et al*., 2016). However, it has been shown that the incidence rates are increasing in Africa (WHO, 2016). This may be the result of an epidemiological transition; the decline of communicable diseases (due to improvements in treatment options) and an associated rise of non-communicable diseases, thereby exposing the increasing incidences (Dalal *et al*., 2011).

In contrast to the low incidence rates, the mortality rates appear to be higher in African countries compared to other parts of the world (Figure 1-1). The higher mortality rates can be ascribed to poorly resourced health facilities in limited resource countries. As such, cancers are detected at later stages when the disease has progressed extensively, leading to a greater chance of mortality (Ferlay *et al*., 2015: Ginsburg *et al*., 2016).

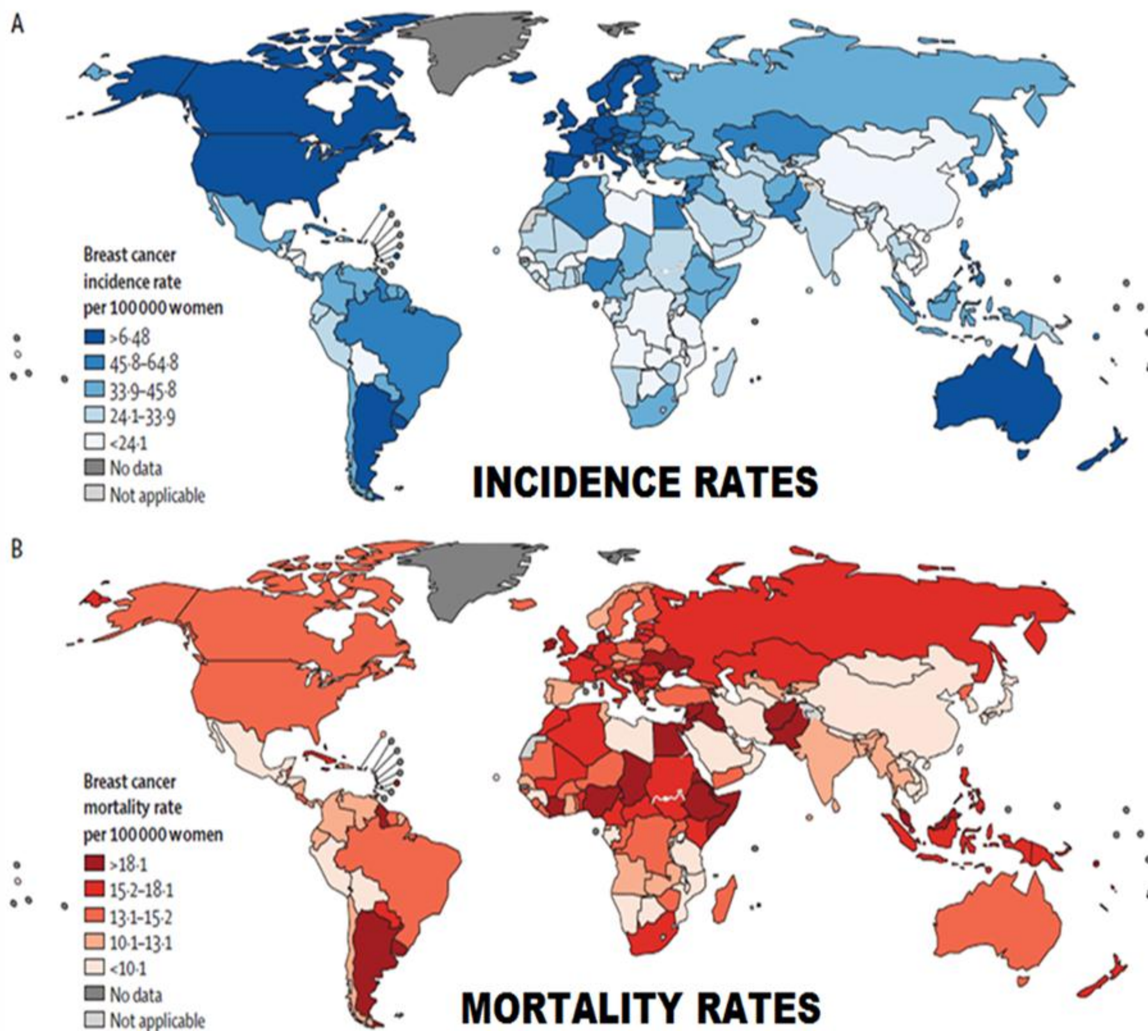


Figure 1-1: The distribution of breast cancer incidence and mortality rates worldwide.
A) Incidence rates of breast cancer across the world, showing higher rates in high-income countries and lower rates in middle and low income countries. B) Mortality rates of breast cancer worldwide. Higher mortality rates are observed in many African countries in comparison to North America and Europe. (Ferlay *et al*., 2015).

There are multiple risk factors known to predispose individuals to breast cancer development; these include reduced physical activity, poor dietary practices and delayed childbearing (Shetty, 2013; Brinton *et al*., 2014). Other risk factors are early-onset menarche, late stage menopause, the use of oral contraceptives or exposure to exogenous oestrogen through industrialisation and having a genetic predisposition (Brinton *et al*., 2014; Ginsburg *et al*., 2016).

About 5-10% of all breast cancer diagnoses are estimated to be inherited, caused by the transmission of mutated genes from parents to offspring (Peshkin *et al*., 2010). Inherited breast cancer shows high genetic heterogeneity and it can be caused by mutations in several different genes (Melchor & Benitez, 2013).

1.2 The genetic aetiology of inherited breast cancer

Inherited breast cancer is primarily caused by germline mutations in the *BRCA1* and *BRCA2* genes. Early estimates indicated that mutations in these two genes account for approximately 20-25% of IBC cases caused by known genes (Easton, 1999). However, other cancer susceptibility genes have also been implicated in IBC. These include Phosphatase and Tensin Homolog (*PTEN*), Cadherin 1, Type 1 (*CDH1*) and Tumour Protein 53 (*TP53*) which are considered to confer high risk for breast cancer when mutated and together, account for approximately 5% of IBC (Borresen *et al*., 1992; Li *et al*., 1997; Pharoah *et al*., 2001). Genes such as Checkpoint Kinase 2 (*CHEK2*), Ataxia Telangiectasia Mutated (*ATM*) and BRCA1 Associated RING Domain 1 (*BARD1*), highlighted in Figure 1-2, are considered to be moderate penetrance genes and explain an additional 5% of IBC (Borresen *et al*., 1990; Thai *et al*., 1998; Bell *et al*., 1999). Fibroblast Growth Factor Receptor 2 (*FGFR2*) and Trinucleotide-repeat-containing 9 (*TNRC9*) confer a small increase in risk for IBC (Easton *et al*., 2007). Another 41 low penetrance genes have been identified and together with previously identified genes, may account for approximately 14 % of IBC (Michailidou *et al*., 2013; Melchor & Benitez, 2013). On-going research is revealing more genes which are associated with breast cancer susceptibility and an additional 65 susceptibility loci were recently identified and reported (Michailidou *et al*., 2017).

In patients where a mutation in a known cancer susceptibility gene has not been identified, disease may result from multiple low-penetrance genes acting together (Smyth *et al*., 2015). Alternatively, a large proportion (approximately 51% of cases) can be

attributed to mutations in unidentified genes (Easton *et al*. 2007; Cox *et al*., 2007; Gracia-Aznarez *et al*., 2013).

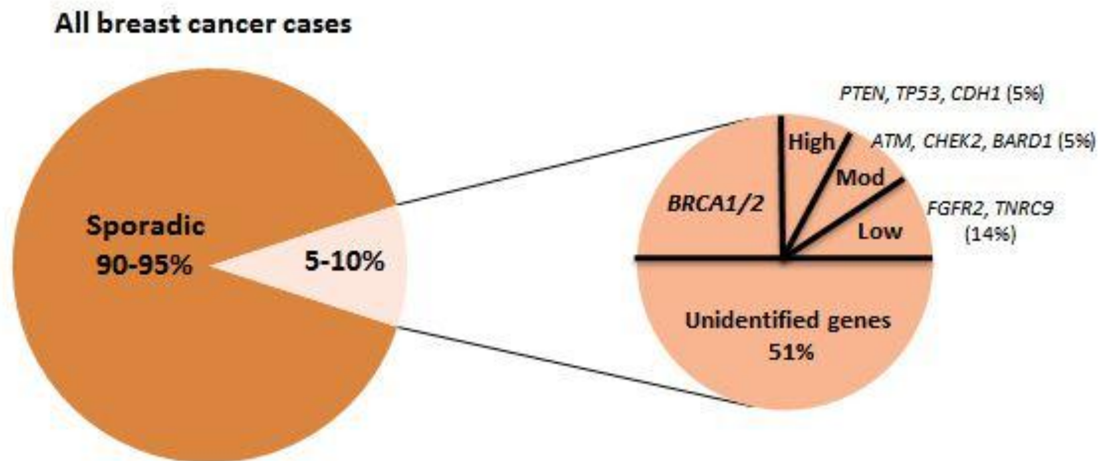


Figure 1-2: The genetic contribution to breast cancer cases.
Inherited breast cancer is rare and is observed in approximately 5-10% of all patients diagnosed with breast cancer. The majority of inherited cases are attributed to mutations in *BRCA1* and *BRCA2*. In addition to *BRCA1/2*, other high, moderate and low susceptibility genes account for a proportion of the cases. The majority of patients, however, do not carry mutations in known genes. It is therefore likely that breast cancer is caused by other genes that are yet to be determined.

Inherited breast cancer due to *BRCA1* or *BRCA2* mutations is inherited in an autosomal dominant manner and shows incomplete penetrance (Hall *et al*., 1990; Miki *et al*., 1994). Thus, each offspring born to an individual carrying a pathogenic mutation in *BRCA1* or *BRCA2* has a 50% chance of inheriting this mutation. As a result of the incomplete penetrance, not all individuals who have inherited the pathogenic variant will manifest disease. The penetrance associated with mutations in *BRCA1* or *BRCA2* has been shown to vary between and within families. Females with a *BRCA1* or *BRCA2* deleterious mutation have a 39 to 85% risk of developing breast cancer (Yurgelun *et al*., 2015). The *BRCA2* gene confers a slightly lower penetrance of breast cancer compared to the *BRCA1* gene. Additionally, *BRCA1* mutations are more common in high-risk western European and North American populations than those in *BRCA2* (Ford *et al*., 1998). In contrast, *BRCA2* mutations appear to be more prevalent than *BRCA1* mutations in individuals of African descent (Pal *et al*., 2013).

1.2.1 The *BRCA1* gene

The first breast cancer gene *BRCA1* (Breast Cancer Gene 1; OMIM 113705), was localised in 1990 (Hall *et al*., 1990). The gene was mapped to chromosome 17q21 by a linkage analysis study in a cohort of individuals with early-onset breast cancer. In 1994, positional cloning was utilised to identify the causal gene, *BRCA1* (Miki *et al*., 1994).

The *BRCA1* gene is a tumour suppressor and it is involved in maintaining genomic integrity. It comprises 24 exons, 22 of which are coding, and spans approximately 125 kilobases (kb) of genomic DNA (Miki *et al*., 1994; Cunningham *et al*., 2014). The *BRCA1* gene has a pseudogene (*BRCA1P1*/*ΨBRCA1*) located 45 kb upstream of the *BRCA1* 5' terminus (Brown *et al*., 1996; Tessereau *et al*., 2015). The pseudogene spans approximately 14 kb of the genome and it arose from a duplication event. It comprises the *BRCA1* promoter region, exon 1 and the majority of exon 2 (Brown *et al*., 1996; Tessereau *et al*., 2015). The wildtype allele of *BRCA1* is transcribed to a 7.8 kb messenger RNA (mRNA) which is then translated to a breast cancer type 1 susceptibility protein (BRCA1) comprised of 1863 amino acids (Miki *et al*., 1994). The protein is expressed in various tissues such as breast and ovarian tissues, as well as pancreatic and prostate tissues (Miki *et al*., 1994; Cavanagh *et al*., 2015).

The protein, as illustrated in Figure 1-3, is made of the RING (Really Interesting New Gene) domain and two BRCT (BRCA1 C terminus) domains which are located on the amino and carboxyl termini, respectively (Meza *et al*., 1999; Manke *et al*., 2003). The RING domain associates with BARD1 (BRCA1-associated RING domain protein 1) to mediate functions such as ubiquitination, DNA repair and transcription regulation. Additionally, association of BARD1 with the BRCA1 RING domain is necessary in E3 ubiquitin ligase activity. The interaction is required for the tumour suppressor function of BRCA1 so that the appropriate target protein substrate can be ubiquitinated and subsequently degraded (Meza *et al*., 1999).

The BRCT domain's function is to bind phosphorylated proteins (e.g. ATM) which are mostly involved in response to DNA damage (Manke *et al*., 2003). The damage to DNA can arise during DNA replication processes and from environmental exposure to ultraviolet and ionising radiation (Jackson & Bartek, 2009). The RING and BRCT domains of BRCA1 make it possible for the protein to bind and interact with various other proteins to retain genomic stability (Ford & Easton, 1995). The BRCA1 protein, for instance, interacts with PALB2 and BRCA2 to facilitate homologous recombination-mediated repair. Mutations in the protein disrupt the interaction with PALB2 and BRCA2, leading to failure in the DNA repair mechanism (Sy *et al*., 2009). Accumulation of mutations in the cell may subsequently result in tumorigenesis.

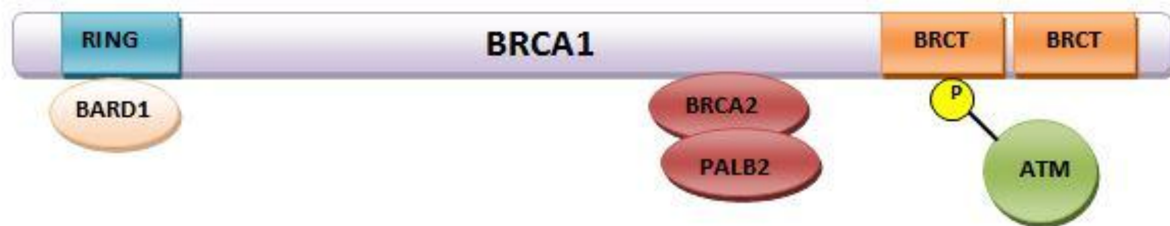


Figure 1-3: **A schematic representing the functional domains of the BRCA1 protein.**
The protein is composed of 1863 amino acids and has functional domains; the RING and two BRCT domains. The RING domain associates with BARD1 so that appropriate damaged substrates are ubiquinated and subsequently degraded. The BRCT domain recruits and binds phosphorylated proteins such as ATM in response to DNA damage. Proteins such as BRCA2 and PALB2 interact with BRCA1 to facilitate homologous recombination-mediated repair.

More than 1600 different pathogenic mutations have been identified in the *BRCA1* gene (Petrucelli *et al*., 2010). The mutations observed include single nucleotide variants (SNVs), small insertions and deletions, and splice-site mutations. A significant proportion of the mutations are located within exon 11, which is the largest exon (Dodova *et al*., 2015). Large duplications and deletions account for approximately one-third of mutations and can be attributed to the homologous recombination events between *BRCA1* and its pseudogene (Zhang *et al*., 2010; James *et al*., 2015).

1.2.2 The *BRCA2* gene

The *BRCA2* (Breast Cancer Gene 2; OMIM 600185) gene is a tumour suppressor and it was mapped in 1994 (Wooster *et al*., 1994). Wooster *et al*. (1995) identified the cDNA sequence of *BRCA2* by using a positional cloning technique. The gene is located on chromosome 13q12.3, contains 27 coding exons, and spans an approximate region of 85kb of the genome (Wooster *et al*., 1995; Cunningham *et al*., 2014). The *BRCA2* gene produces a 10.4kb transcript which encodes the breast cancer type 2 susceptibility protein (BRCA2), comprised of 3418 amino acids (Wooster *et al*., 1995). The BRCA2 protein shares similar functions with BRCA1 such as maintaining the integrity of the genome (Sy *et al*., 2009).

The BRCA2 protein is composed of a transcription activation domain on its amino or N-terminus, eight copies of a repetitive motif (known as BRC repeats) and a DNA-binding domain located centrally and two nuclear localisation signals (NLS) on the carboxyl C-terminus, as demonstrated in Figure 1-4 (Bork et al., 1996; Bignell *et al*., 1997; Milner *et al*., 1997; Yang *et al*., 2002). The N-terminal activation domain binds the p300/CBP-associated factor (P/CAF) which is a transcription co-activator. P/CAF has acetyltransferase activity which is important in regulating transcription (Fuks *et al*., 1998). The BRC repeats bind RAD51 to BRCA2 at sites of DNA damage to facilitate homologous-mediated DNA repair (Yuan *et al*., 1999). Downstream of the BRC repeats is a DNA-binding domain which assists BRCA2 to bind to single and double-stranded DNA (Yang *et al*., 2002). The NLS on the C-terminus is responsible for the transfer of proteins into the nucleus (Yano *et al*., 2000). This enables BRCA2 to be transported into the nucleus to function in the FA/BRCA pathway in response to DNA damage as discussed in Section 1.2.3. Deficiency of BRCA2 due to an accumulation of mutations and subsequent inactivation of the domains, leads to chromosomal instability (Yang *et al*., 2002).

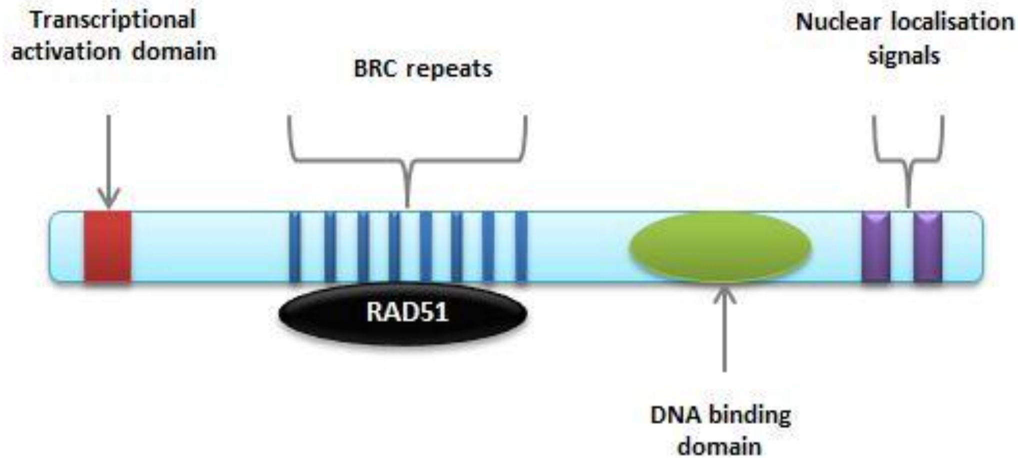


Figure 1-4: A schematic representing the functional domains of the BRCA2 protein.
The transcriptional activation domain lies on the N-terminus of the protein. The BRC repeats which bind RAD51, and the DNA binding domains are located in the centre of the protein. The nuclear localisation signals (NLS) resides on the C-terminus and facilitate transfer of the protein into the nucleus.

Over 1900 different pathogenic mutations have been identified in *BRCA2*, most of which result in truncated, non-functional proteins (Dodova *et al*., 2015). Unlike with*BRCA1*, large duplications and deletions do not occur frequently in *BRCA2* and this can be ascribed to fewer Alu sequences being observed in *BRCA2*. The majority of mutations in the gene are found in exons 10 and 11, which are the largest exons (Gayther *et al*., 1999; Thompson & Easton, 2004).

1.2.3 Other breast cancer susceptibility genes

Mutations in genes other than *BRCA1* and *BRCA2* have been associated with IBC. These genes encode proteins which interact and function in a common pathway to repair damage in DNA. This explains why overlapping phenotypes are observed when deleterious mutations are present in any of the genes. Rare deleterious mutations have been identified in genes considered to be high-, moderate-, and low-risk cancer genes. Some of the better known susceptibility genes are discussed below.

1.2.3.1 The *PTEN* gene

The Phosphatase and tensin homolog (*PTEN*) gene is a tumour suppressor gene thatis located on chromosome 10q23.31 (Steck *et al*., 1997). The gene is composed of nine exons and an additional exon ("5b") which is not included in the main transcript (Sharrard & Maitland, 2000). The PTEN protein is an enzyme involved in regulating cell growth by restraining the progression of the cell cycle at the G1 phase (Li & Sun, 1998). Germline mutations in *PTEN* cause a spectrum of disorders known as PTEN Hamartoma-Tumour Syndrome (Bland *et al*., 2017). These syndromes are inclusive of Cowden Syndrome, which is an adult-onset cancer predisposition disorder, and Bannayan-Riley-Ruvalcaba Syndrome which arises in childhood and is often characterised by a large head circumference. The disorders are inherited in an autosomal dominant manner and are characterised by benign hamartomatous tumours. Mutations in the *PTEN* gene can give rise to different cancers such as those of the breast, prostate and brain. Females with a *PTEN* mutation have an approximate 67% to 85% lifetime risk of developing breast cancer (Bland *et al*., 2017).

1.2.3.2 The *TP53* gene

The *Tumour protein 53 (TP53)* gene is located on chromosome 17p13.1 (Bourdon *et al*., 2005). The gene contains 11 exons, three transcription initiation start sites and two promoter regions (Bourdon *et al*., 2005). The wildtype form of the gene encodes a TP53 protein which is essential in regulating the cell cycle, repair of damaged DNA and mediating apoptotic responses of damaged telomeres. In addition, the protein also regulates senescence which is important in suppressing the growth of tumours (Artandi & Attardi, 2005). When the gene has a pathogenic mutation, it is known to cause Li-Fraumeni syndrome (Li & Fraumeni, 1969; Sorrell *et al*., 2013). Li-Fraumeni syndrome (LFS) is inherited in an autosomal dominant manner and is associated with a predisposition to many cancers such as breast cancer, brain cancer, sarcomas and adrenal cortical carcinomas (Sorrell *et al*., 2013). Breast cancer associated with Li Fraumeni syndrome (LFS) usually presents between the ages of 20 and 30 and the tumours are aggressive.

Additionally, the majority of *TP53* breast tumours have been shown to be positive for the over expression of the HER2 receptor. Breast cancer associated with *TP53* mutation is almost 100% penetrant in females (Sorrell *et al*., 2013).

1.2.3.3 The *CDH1* gene

Cadherin-1 (*CDH1*) is a gene located on chromosome 16q22.1 (Mansouri *et al*., 1988). It spans a region of approximately 100 kb of gDNA and contains 16 exons (Berx *et al*., 1995). The gene encodes a tumour suppressor protein, epithelial cadherin (E-cadherin), which functions in cell-to-cell adhesion to form tissues (Petridis *et al*., 2014). *CDH1* inactivating mutations make individuals susceptible to the development of diffuse gastric cancer. Inactivating heterozygous mutations can also lead to colorectal or lobular breast cancers. Individuals carrying the mutated gene have an approximately 39% lifetime risk of developing lobular breast cancer (Petridis *et al*., 2014).

1.2.3.4 Fanconi anaemia (FA) genes

Mutations involved in causing breast cancer have also been shown to give rise to a different phenotype, Fanconi anaemia (FA) when biallelic mutations are present. FA is a chromosomal breakage disorder and can be inherited in an autosomal or X-linked recessive manner (Savage *et al*., 2016). It is often characterised by congenital and developmental abnormalities, aplastic anaemia, solid tumours and bone marrow failure. Additionally, individuals with FA have an increased predisposition to developing cancer (Meyer *et al*., 2014; Sawyer *et al*., 2015). Fanconi anaemia is a heterogenous disorder which can arise due to mutations in at least 21 different genes. These genes encode proteins that function together to repair damage incorporated during DNA replication (Dong *et al*., 2015).

The *FANCD1* gene, which encodes the Fanconi Anaemia Complementation Group D1 protein, is known to cause a rare FA subtype (Dong *et al*., 2015; Savage *et al*., 2016). Molecular investigations were conducted to determine the relationship between the *BRCA1* and *BRCA2* genes, and Fanconi Anaemia (Howlett *et al*., 2002). Biallelic mutations in the *BRCA2* gene were observed in cells derived from a patient with FA-D1 subtype. The results of the study revealed that *FANCD1* and *BRCA2* are in fact the same gene (Howlett *et al*., 2002). Approximately 3% of all FA cases are a result of rare biallelic mutations in *FANCD1*/*BRCA2* (Alter *et al*., 2007). Individuals carrying biallelic *BRCA2* mutations have been shown to present with a more severe phenotype of FA and typically develop cancer within the first 10 years of diagnosis. This is contrary to classic FA, in which the symptoms are milder and cancer may develop in the second or third decade of life (Alter *et al*., 2003; Rosenberg *et al*., 2008).

Biallelic mutations in *BRCA1* (also known as *FANCS*) predispose patients to a distinct subtype of FA, FA-S (Sawyer *et al*., 2015). The patient described in the study by Sawyer *et al.,* 2015 presented with early-onset breast cancer and congenital abnormalities consistent with FA. Genetic testing showed that the patient harboured deleterious mutations in both of her *BRCA1* alleles (Sawyer *et al*., 2015).

Other FA genes have been reported as moderate-penetrance genes for breast cancer in their heterozygous form. These genes with overlapping FA and IBC phenotypes include *BRIP1* (*FANCJ*) and *PALB2* (*FANCN*). Biallelic mutations in these genes are associated with an FA phenotype, while heterozygous mutations are associated with IBC (Dong *et al*., 2015).

1.2.3.4.1 The Fanconi anaemia/BRCA pathway

The Fanconi anaemia/breast cancer (FA/BRCA) pathway as shown in Figure 1-5, is involved in the recognition of DNA damage and its subsequent repair (Dong *et al.*, 2015). The two pathogenic phenotypes observed due to overlapping mutations in the pathway, demonstrate the close proximity in which the FA and *BRCA* genes function together in the cell.

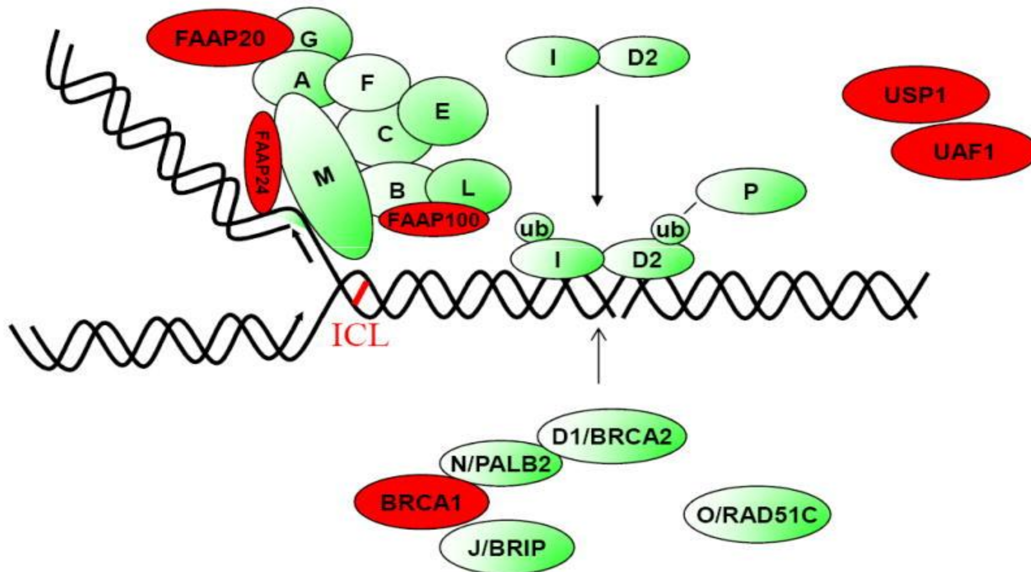


Figure 1-5: The Fanconi anaemia/BRCA pathway.
BRCA1 and BRCA2 interact with the Fanconi anaemia nuclear protein complex and other proteins to facilitate the repair of damaged DNA. Disruption of this pathway leads to clinical features observed in BRCA FA patients. Biallelic *BRCA2* mutations often result in a more severe FA, the FA-D1 subtype (D'Andrea, 2013).

Fanconi anaemia and breast cancer genes function in a common pathway but the disorders differ in their clinical symptoms and presentation, as well as the mode in which they are inherited. Individuals with FA who are found to have mutated genes which could be implicated in breast cancer susceptibility may have parents who are also at increased risk of developing the associated cancers. Screening and management for such individuals would be advisable in these instances. The existence of both phenotypes within a single pathway does indicate the importance of examining all Fanconi anaemia associated genes for susceptibility to breast cancer development in heterozygous form.

1.3 Molecular genetic testing for individuals at highrisk for developing breast cancer

Genetic testing is a method that is used to identify variants that may be involved in causing a particular genetic disease. The testing is useful in the diagnosis of symptomatic individuals, and in screening for mutations in disease genes in individuals at high risk of a disease (Kroese *et al*., 2004). Population-specific genetic testing or screening can be conducted to determine specific, high frequency variants or founder mutations in individuals in a population that has an increased risk for disease (e.g. the Ashkenazi Jewish population) (Sequeiros *et al*., 2012). For those patients ineligible for founder mutation testing, single gene sequencing for the most likely genes (*BRCA1* and *BRCA2*) and large genomic rearrangement (LGR) testing using a technique such as multiplex ligation probe amplification (MLPA) can be provided (Sohn *et al*., 2016; Winship & Southey, 2016).

In many laboratories across the world, next generation sequencing (NGS) is being performed to determine the presence of deleterious mutations in cancer susceptibility genes (Crawford *et al*., 2017). NGS techniques are high-throughput and enable hundreds or thousands of genes or genomic regions to be analysed simultaneously, making sequencing time and cost effective (Winship & Southey, 2016). The different NGS approaches include whole-genome sequencing (WGS), whole-exome sequencing (WES) and targeted gene panels (Rehm *et al*., 2013). The entire genome of an organism is screened using WGS whereas in WES, only the coding regions of the genome are sequenced (Rehm *et al*., 2013).The targeted gene approach differs from both WGS and WES in that only genes known to be associated with a particular disease or phenotype are investigated (Rehm et al., 2013). Examining a limited number of genes allows a greater depth of coverage. In addition, regions of heterozygosity can be confidently identified when the sequence coverage is of great depth. The data generated from gene panels are more manageable both in storage and interpretation (Rehm *et al*., 2013).

When gene panels yield no results, WES can be employed but this is in rare, exceptional cases (Noh *et al*., 2016). This is often performed in a research setting and where results are suspected to be pathogenic; a diagnostic test is set up to validate the mutation. Since many variants of uncertain significance (VUS) are produced from NGS techniques, subsequent functional studies are necessary to determine the pathogenicity of these variants and their association with disease (Noh *et al*., 2016).

Next generation sequencing was utilised in the African Genome Variation Project (AGVP) in an attempt to characterise the genetic diversity seen in Africa (Gurdasani *et al*., 2015). In the study, 1 481 individual samples were genotyped using the HumanOmni2.5 genotyping chip comprised of approximately 2.5 million markers and 320 individual samples were assessed using whole genome sequencing (WGS). The results from the project will offer comprehensive genomic data which is applicable to African individuals. In addition, the genome data can be used as a frame of reference for studies conducted in individuals of African ancestry. The individuals in the study were recruited from 18 populations in Sub-Saharan Africa (SSA) including Yoruba in Nigeria, Luhya in Kenya and Zulu and Sotho from South Africa (Gurdasani *et al*., 2015). The advent of NGS techniques holds promise to identify genetic causes associated with breast cancer and other non-communicable diseases in African populations where knowledge is limited.

1.4 Breast cancer in African populations

Breast cancer is understudied in African populations. The majority of studies have been conducted in European populations and thus, under-represent the genomic diversity observed in African populations. There is currently limited information on ways to mitigate challenges associated with poor disease prognosis, testing and management. However, in recent years, studies have been performed to gain more insight into breast cancer in the African context.

Individuals of African ancestry have been documented to present with a severe and aggressive phenotype; their breast cancer diagnoses are often at earlier ages and they generally have a poorer disease prognosis compared to individuals of European descent. In addition, patients of African descent present more frequently with triple-negative breast cancer (TNBC) (DeSantis *et al*., 2016). This trend has been documented in African-American individuals and individuals in various parts of the African continent.

The incidence rates of breast cancer in individuals of African descent are increasing. It was noted that in African-Americans, the incidence rates of breast cancer were estimated to be 119.4 per 100 000 individuals between the years 1998 and 2002 (Smigal *et al*., 2006) and this figure increased to 124.3 per 100 000 in the period between 2008 and 2012 (DeSantis *et al*., 2016). In Africa, approximately 100 000 new cases of breast cancer were reported and 49 000 deaths documented in the year 2012 (Ferlay *et al*., 2015). Incidence rates differ slightly between the different parts of the African continent; per 100 000 women, the incidence rates were 26.8 in Middle Africa, 30.4 in Eastern Africa, 38.6 in Western Africa, 38.9 in Southern Africa and 43.2 in Northern Africa in 2012 (Ferlay *et al*., 2015). The incidence rates in all African regions might be underestimated due to an under-representation of data from rural regions. Patients in urban areas have easier access to healthcare services than those in rural communities and thus patients in rural communities may be undiagnosed (Ferlay *et al*., 2015).

An earlier study performed in the United States of America was carried out in order to evaluate the distribution of *BRCA1* and *BRCA2* mutations in African-American and white women diagnosed with early-onset breast cancer (Haffty *et al.,* 2006). The study revealed that African-American women had more *BRCA2* mutations than *BRCA1* mutations and a greater number of variants of uncertain significance, compared to white women who were found to carry *BRCA1* mutations predominantly. These results are consistent with what is currently being observed in breast cancer patients of African descent (Schoeman *et al*., 2013).

In North Africa, a study was conducted in Algeria to identify mutations in the *BRCA1* gene in patients with triple-negative breast cancer (TNBC) (Gaceb *et al*., 2017). Investigations were conducted only for the *BRCA1* gene, as the TNBC phenotype is associated more often with *BRCA1* germline mutations than with *BRCA2* mutations. DNA samples from 103 patients were examined and eight (7.8%) found to carry a pathogenic mutation in *BRCA1* (Gaceb *et al*., 2017).

A recent pilot study in Tunisia investigated the use of MLPA to determine the prevalence of large genomic rearrangements in *BRCA1* and *BRCA2,* and the *CHEK2* 1100delC mutation in that population (Riahi *et al*., 2017). The 36 individuals in the study had previously tested negative for *BRCA1* and *BRCA2* point mutations. From this study, 5.5% of the patients were found to have a rearrangement in *BRCA1* and no rearrangements in *BRCA2* were identified. In addition, none of the patients tested positive for the *CHEK2* mutation (Riahi *et al*., 2017).

Genome-wide association studies (GWAS) and fine mapping have been used to identify more than 100 breast cancer susceptibility loci in women of European descent with breast cancer (Feng *et al*., 2017). Due to the reduced linkage disequilibrium detected in African populations, the markers previously identified in European populations may not accurately associate the genetic variant to disease risk. The determination of loci associated with breast cancer in individuals of African descent was carried out by performing fine-mapping in 6 522 breast cancer patients and 7 643 controls. This revealed 74 breast cancer risk variants. Of the 74 variants, 68 were found to be more common in the women included in the study, with minor allele frequencies (MAF) more than 0.05. When odds ratios were examined, it was revealed that 54 of the variants had ratios similar to what has been previously described in European ancestry. Twelve of the 54 variants showed a statistical significance with a p-value less than 0.05. Although the study shows similarities in risk loci across ethnicities, additional research is required to characterise fully their association with breast cancer development particularly in African women.

The studies above highlight some of the research that has been done to try and understand the genetic aetiology of breast cancer in individuals of African ancestry. Although the studies have provided some significant results, more research is still warranted.

1.5 Breast cancer in the South African context

According to the National Cancer Registry (NCR) of South Africa, breast cancer is the most common cancer diagnosed in females, with approximately 8 203 cases diagnosed in 2012 (NCR, 2012). The lifetime risk for disease is approximately 1 in 26 across all ethnic groups. It is estimated that 1 in 12 white South African females will develop breast cancer in their lifetime, while the lifetime risk for a black female is estimated to be 1 in 45 (NCR, 2012). Previous studies suggest that breast cancer has increased in incidence in the black South African population (NCR, 2012). The ratio of diagnosed cases of black to white women rose from 1:6 in 1993 to 1:4 in 2012, indicating an increase in the number of diagnosed cases in the black population (Loubser *et al*., 2008; NCR, 2012). A number of studies have been performed to determine the genetic basis of breast cancer in South Africa by looking at *BRCA1*, *BRCA2* and other susceptibility genes.

1.5.1 Inherited Breast Cancer in the Afrikaner population

Various studies have been performed to understand the genetic aetiology of breast cancer specifically in the Afrikaner population due to the high incidence observed in this group. In a 2004 study, the *BRCA1* gene was screened in 90 South African families of European origin with breast and/or ovarian cancer to detect predisposing mutations (Reeves *et al*., 2004). The families were recruited from Tshwane (formerly Pretoria) and Bloemfontein. Of these families, 18 (20%) were identified to have a deleterious mutation in the *BRCA1* coding region and four of the mutations had not been previously reported. One mutation, c.2641G>T (p.Glu881Ter), was present in five Afrikaner families and reported as a novel founder mutation (Reeves *et al*., 2004).

In order to further identify predisposing variants in *BRCA1* and *BRCA2*, 147 Afrikaner breast and/or ovarian cancer families were screened. From the screen, three recurrent mutations were observed; c.1493delC and c.2641G>T in *BRCA1*, and the c.7934delG variant in *BRCA2*. These mutations were observed in 97/103 families who carried a *BRCA* mutation, accounting for 94% of *BRCA*-positive mutations. Genealogical and genotype analyses revealed that the three variants are founder mutations for the Afrikaner population (van der Merwe & van Rensburg, 2009).

In a more recent study, a retrospective file review was performed on individuals of Afrikaner ancestry who reside in Johannesburg (Seymour *et al*., 2016). The files under review were those of 86 unrelated individuals who were provided with genetic counselling services between the years 2001 and 2014. Of the 86 individuals whose files were reviewed, 54 (62.8%) underwent genetic testing for *BRCA1* and *BRCA2* genes. Eighteen (33.3%) of the individuals tested positive for a *BRCA* mutation and 14 of the 18 (77.8%) carried one of the reported founder mutations (Seymour *et al*., 2016).

1.5.2 Inherited Breast Cancer in the Ashkenazi Jewish population

As with the Afrikaner population, mutations that are more common in individuals of Ashkenazi Jewish ancestry have been identified. The *BRCA1* c.185delAG frameshift mutation was first observed in approximately 1% of the 858 Ashkenazi Jewish individuals seeking genetic testing for conditions unrelated to cancer. The finding suggested an especially high risk of developing breast and/or ovarian cancer for one in every hundred Ashkenazi Jewish women (Struewing *et al*., 1995). This mutation was then shown to lie in a founder Ashkenazi Jewish haplotype (Neuhausen *et al*., 1996). In a South African study, four families of Ashkenazi Jewish ancestry with a positive family history of breast cancer were all found to carry the same c.185delAG mutation (Reeves *et al*., 2004).

The *BRCA1* c.1493insC and *BRCA2* c.8162delG mutations have also been identified as common mutations for the Ashkenazi Jewish population (Schoeman *et al*., 2013). These three mutations are offered as a first-line test for individuals affected by breast cancer and with Ashkenazi Jewish ancestry (Schoeman *et al*., 2013).

The studies above on Afrikaner and Ashkenazi Jewish populations in South Africa demonstrate the importance of identifying founder mutations as they prove to be a time-saving and cost effective means of testing. The cost of Afrikaner founder mutation testing for example, ranges between R1,600.00 and R3,500.00 while costs for comprehensive screening for *BRCA1* and *BRCA2* range between R9,000.00 and R25,000.00 (prices will vary between laboratories and depends on the testing approach) (Seymour *et al*., 2016). The detection rate and pricing for founder mutation testing in this population demonstrates the utility of identifying founder mutations in different ethnic groups and importantly, for the black South African population which has been previously understudied.

1.5.3 Inherited Breast Cancer in the Black South African and Coloured Populations

In an effort to better understand the genetic aetiology of breast cancer, a study was performed in the Western Cape Xhosa and coloured (mixed ancestry) populations (van der Merwe *et al*., 2012). The study screened for mutations in *BRCA1* and *BRCA2* genes in 16 black and 105 coloured women. Four pathogenic mutations were identified; c.1504_1508delTTAAA in *BRCA1* and c.2826_2829del, c.6447_6448dupTA and c.5771_5774delTTCA in *BRCA2*. The latter mutation was observed in 25% (4/16) of Xhosa women and 3.8% (4/105) of coloured females included in the study. Haplotype analysis was performed to determine the presence of a founder effect. The mutation has since been documented as a founder mutation for the Xhosa and coloured populations (van der Merwe *et al*., 2012).

The above-mentioned founder mutation (*BRCA2* c. 5771_5774delTTCA) was identified in a separate study conducted in breast cancer patients based in Gauteng Province. The study was performed in the Division of Human Genetics at Wits University and investigated the presence of identify germline mutations in *BRCA1*, *BRCA2* and *PALB2* in 108 individuals from four South African ethnic groups (black, white, coloured and Indian) using a next-generation sequencing approach and MLPA (Francies *et al*., 2015). Of the 108 individuals in the study, 85 (78.7%) were black. The deleterious mutations identified in the black South African population in this study are summarised in Table 1-1 below:

Table 1-1: Pathogenic germline mutations identified in the black South African population group using next generation sequencing and MLPA techniques (Francies *et al*., 2015)

Gene	**Exon**	**HGVS nomenclature**	**Frequency of mutation**	**Effect of mutation on protein**
BRCA1	4	c.212G > A	1/85	Missense
BRCA1	10	c.1155G > A	1/85	Nonsense
BRCA1	10	c.1953_1954insA	1/85	Frameshift
BRCA2	7	c.582G > A	1/85	Nonsense
BRCA2	11	c.5771_5774delTTCA	1/85	Frameshift
BRCA2	23	c.9097_9098insA	1/85	Frameshift

HGVS - Human Genome Variation Society

In addition to the presence of the *BRCA2* c.5771_5774delTTCA Xhosa/mixed ancestry founder mutation which was identified in one of the 85 patients (1.2%), the *BRCA2* c.582G>A mutation (see Table 1-1 above) was also of particular interest. It had been previously reported once in the black South African population in the same Division (Chen, 2015). It had also been reported as a causative mutation in a FA patient found to have biallelic *BRCA2* mutations as the cause of disease (Feben *et al.,* 2017). Upon scrutinising patient files, these two individuals, both with breast cancer, were found to be unrelated

to each other or to the child affected with FA. This may suggest a potential founder mutation for the black South African population.

Studies conducted in the past have shown that there is poor knowledge of breast cancer in the majority of black South African women, especially in rural communities (Maree *et al*., 2013; Ramathuba *et al*., 2015). A study undertaken by Maree *et al*., 2013 in a poor community in Tshwane, Gauteng Province, revealed that 80.6% of the individuals included in the study were unaware of the risk factors, warning signs and screening associated with breast cancer (Maree *et al*., 2013). In a study conducted in Thohoyandou, Limpopo Province, 69% of the females in the study had never heard of breast cancer (Ramathuba *et al*., 2015). Educating the public (particularly in rural communities) about breast cancer may therefore aid in identifying at-risk individuals. This could also aid in characterising *BRCA1* and *BRCA2* gene mutations, as well as mutations in other cancer susceptibility genes that contribute to the breast cancer burden in the black South African population. Identification of such mutations has the potential to improve genetic testing approaches and counselling in the future. Furthermore, identification of mutations presents an opportunity to develop better management and treatment options, as well as personalised or population specific healthcare approaches.

1.6 Rationale for study

Various research projects have been conducted at the Division of Human Genetics (University of the Witwatersrand and the National Health Laboratory Service) to identify *BRCA1 and BRCA2* mutations in black South African individuals at increased risk for inherited breast cancer (Chen, 2015; Francies *et al*., 2015). From these studies, a small number of rare pathogenic mutations have been identified. Two *BRCA2* mutations of particular interest, c.5771_5774delTTCA (p.Ile1924ArgfsTer38) and c.582G>A (p.Trp194Ter) were observed in unrelated patients as described in Section 1.5.2 above. The c.5771_5774delTTCA mutation had been previously reported as a founder mutation in

the Western Cape Xhosa and coloured population groups (van der Merwe *et al*., 2012). The *BRCA2* c.582G>A mutation has been reported on the National Center for Biotechnology Information (NCBI) ClinVar database (accessible at https://www.ncbi.nlm.nih.gov/clinvar/) as a pathogenic variant for familial breast cancer.

Although these efforts have identified some mutations of in *BRCA1* and *BRCA2* in the black South African population, additional data is still required. This data includes understanding the contribution of specific mutations to the incidence of IBC in this population which would hopefully aid in streamlining testing approaches, which in turn, could potentially result in cost-effective genetic counselling and testing services in the local black population. The use of targeted next generation sequencing technologies to explore the contributions of genes other than *BRCA1* and *BRCA2* in a setting with limited resources needs to be evaluated and considered for long-term application for inherited breast cancer.

1.6.1 The *BRCA2* c.5771_5774delTTCA and c.582G>A mutations

In several studies conducted in the Division, the *BRCA2* c.582G>A and c.5771_5774delTTCA mutations have been observed in a number of black patients with breast cancer and Fanconi anaemia (Chen, 2016; Francies *et al*., 2015; Feben *et al*., 2017). However, a family history of cancer was not observed in either of the families of the probands with breast cancer. In one of the studies, a proband with an FA phenotype was found to be a *BRCA2* compound heterozygote for the above-mentioned mutations (Feben *et al*., 2017). The presence of the c.582G>A and c.5771_5774delTTCA mutations in patients diagnosed with breast cancer and FA indicate that these mutations could potentially be founder mutations in the black South African population.

1.6.2 The *FANCG* c.637_643delTACCGCC mutation

A common seven base-pair mutation, c.637_643delTACCGCC, has been identified in the *FANCG* gene in the black Southern African population (Morgan *et al*., 2005). The mutation was identified in 33/40 (82.5%) of patients of African ancestry with FA (Morgan *et al*., 2005). The gene product (FANCG) has been previously shown to directly interact with *BRCA2* in response to DNA damage through homologous recombination repair (Hussain *et al*., 2003). Due to this strong interaction, it is important to determine whether *FANCG* mutations contribute to a breast cancer phenotype given that *BRCA2* mutations contribute to a FA phenotype.

1.6.3 Next generation sequencing

Previous studies in the Division indicate that approximately 10% of high risk black patients carry pathogenic germline mutations in either *BRCA1* or *BRCA2 (*Chen, 2015; Francies *et al.,* 2015*)*. It is therefore important to identify genes and mutations other than *BRCA1* and *BRCA2*, given that many high-risk patients do not carry a *BRCA* mutation. The next step would be to examine other known genes that predispose individuals to inherited breast cancer when mutated.

1.7 Aims and objectives

The overall aim of the study is to advance current knowledge on the contribution of known cancer susceptibility genes to inherited breast cancer in the black South African population.

The following objectives were established in order to achieve this aim:

A. To screen DNA samples from black South African women and women of mixed ancestry, diagnosed with breast cancer, for the two pathogenic *BRCA2* mutations:

c.5771_5774delTTCA and c.582G>A to determine if they represent common mutations.

B. To perform haplotype analysis across the *BRCA2* gene in order to determine whether a founder effect exists for each of the two mutations.

C. To screen DNA samples from black South African women and those of mixed ancestry, diagnosed with breast cancer, for the *FANCG* c.637_643delTACCGCC mutation to examine its contribution to the inherited breast cancer phenotype.

D. To utilise a targeted gene sequencing panel to screen multiple susceptibility genes in high-risk patients for whom no mutation has been identified in the *BRCA* genes.

2 SUBJECTS AND METHODS

This chapter will describe the participants included in the study and the methodologies that were utilised to achieve the aims and objectives outlined in the previous section. Section 2.1 describes the participants and the selection criteria used to determine which participants were eligible for the study. Section 2.2 describes how the quality and quantity of DNA was analysed. Section 2.3 outlines the *BRCA2* mutation screening methodologies. Section 2.4 describes the haplotyping methodologies employed. Section 2.5 describes how the *FANCG* mutation was screened and Section 2.6 will discuss the next-generation sequencing portion of the study including the methodologies and analysis tools used for data annotation.

The study was approved by the Human Research Ethics Committee (HREC) Medical, University of the Witwatersrand (WITS) under clearance certificate number: M16111159 (see Appendix A). All solutions used in this chapter are detailed in Appendix C.

2.1 Study participants

Participants in the study were seen and assessed through the Clinical and Counselling Section of the Division of Human Genetics, University of the Witwatersrand and the National Health Laboratory Service (Wits/NHLS). The patients were assessed at various tertiary hospitals in and around Johannesburg, namely: Charlotte Maxeke Johannesburg Academic Hospital (CMJAH), Chris Hani Baragwanath Academic Hospital (CHBAH), Helen Joseph Hospital (HJH) and Rahima Moosa Mother and Child Hospital (RMMCH).

The individuals included in the study were all South African women of black or mixed ancestry who were thought to be at high-risk for inherited breast cancer based on their age of diagnosis (less than 50 years). At the time of genetic counselling, blood was drawn

and DNA was extracted under routine diagnostic procedures by the Molecular Diagnostic Laboratory at the Division. All participants signed an informed consent form (Appendix B) to participate in research studies related to their primary disease (breast cancer). The inclusion criteria for this study included:

1) A diagnosis of breast cancer made before the age of 50 years

2) Patients with breast cancer diagnosed older than 50 years with a positive family history of cancer/s

For the next generation sequencing portion of the study, an exception to the third inclusion criterion was made for a portion of patients (27.8%) who had previously tested negative for mutations in the *BRCA* genes. This was because the aim of this portion of the study was to determine the presence of other inherited cancer susceptibility gene mutations in this population. It was, therefore, necessary to select high-risk candidates to ensure the greatest possible yield in a small sample of patients. Table 2-1 summarises the number of participants for each objective of the study.

A control group of 100 black South African female individuals were also included in the study. These DNA samples were sourced from the Birth-To-Twenty cohort study (Richter *et al*., 2005), comprised of approximately 3000 individuals born in the Johannesburg-Soweto metropolitan area in the year 1990. The individuals were 25 years of age at the time of DNA analysis and breast cancer status was unknown.

Table 2-1: The number and ethnicity of participants in each study objective

Study objectives	**Participants (N)**		
	Black	Mixed Ancestry	Controls
A: Screening for the c.5771_5774delTTCA *BRCA2* mutation	27	8	0
B: Screening for the c.582G>A *BRCA2* mutation	27	8	0
C: Haplotyping analysis	5*	0*	100
D: Screening for the *FANCG* c.637_634delTACCGCC mutation	119	0	0
E: Next generation sequencing using a targeted breast cancer gene panel (24-sample library preparation kit)	24	0	0

*This number was derived from individuals who were found to have one of the two *BRCA2* mutations during the initial screening process (n=1 black individual and n=0 mixed ancestry individuals) as well as an additional four patients who had previously been found to have these mutations in other research projects.

Objectives A and B of the study involved screening 35 women (of African and coloured) for the *BRCA2* c. 582G>A and c.5771_5774delTTCA mutations; these women had not been previously tested for *BRCA1* and *BRCA2* mutations.

Objective C was to perform haplotype analysis for the five individuals who had previously tested positive for the *BRCA2* c.582G>A or c.5771_5774delTTCA mutations.

Objective D was to screen for the *FANCG* c.637_634delTACCGCC mutation in all 119 women who had previously tested negative for *BRCA1* and *BRCA2* mutations across all breast cancer studies performed previously in the Division, and including those from this study.

Objective E was to perform targeted sequencing on 24 women who had either not been previously screened for *BRCA1* and *BRCA2* mutations (N=18); OR had tested negative (N=7). The number was limited by the funding available for a single library preparation kit (24 reactions).

2.2 Quantitative and qualitative analysis of DNA

The DNA samples were retrieved from storage and analysed for quality. The DNA was quantified on a NanoDrop ND-1000 spectrophotometer (*NanoDrop technologies, Inc.*) to determine concentration. The integrity of the DNA was then analysed by agarose gel electrophoresis to confirm whether it could be used in downstream applications. This was performed by resolving the DNA on a 0.8% agarose gel stained with Ethidium Bromide (EtBr) (*Sigma-Aldrich*). Subsequent to staining, the DNA was combined with Ficoll based loading dye. The DNA was then loaded onto the gel and subjected to a voltage of 150 volts (V) in 1 x Tris-Boric acid EDTA (TBE) buffer alongside a $1kb^+$ DNA ladder (*Invitrogen™*). The gel was visualised using the Vacutec gel documentation system (*Vacutec*) which emits UV light to excite the EtBr fluorescence. The gel image was captured by GeneSnap v6.08 software (*Syngene*). Following quality control, non-degraded DNA samples were diluted to a concentration of 100ng/µl and used in the screening reactions as described in Section 2.3 below.

2.3 Screening for the *BRCA2* founder mutations

In order to screen for the two *BRCA2* mutations which were thought to be of particular relevance to the black and mixed ancestry South African populations, it was necessary to design a genotyping assay appropriate for each mutation.

2.3.1 *BRCA2* c.5771_5774delTTCA mutation

The c.5771_5774delTTCA *BRCA2* mutation was screened for by restriction fragment length polymorphism (RFLP) analysis. The method is cost effective and can easily distinguish affected individuals from those who are unaffected by cleaving DNA when the restriction endonuclease site is present.

2.3.1.1 Primer design

Primers were manually designed to amplify the region of interest; sequences and details are given in Table 2-2 below. The melting temperature (T_m) between the primers is less than 5℃ which improves specificity of primer binding to the target sequence.

Table 2-2: Oligonucleotide sequences used to amplify the *BRCA2* region to detect the presence of the c.5771_5774deTTCA mutation

Primer Name	Direction	T_m (℃)	Length	Sequence
c5771_Cont_F	Forward	53.6	28	5'- TGA AAG ACA TAT TTA CAG ACA GTT TCA G 3'
c5771_Cont_R	Reverse	56.7	28	5' AGA CTG ACT TAT GAA GCT TCC CTA TAC T -3'

2.3.1.2 Polymerase chain reaction

Polymerase chain reaction (PCR) was used to amplify the region of interest containing the *BRCA2* c.5771_5774delTTCA mutation prior to RFLP analysis. To determine the optimal temperature for amplification, a gradient PCR was performed with annealing temperatures ranging between 50°C and 60°C. The optimal annealing temperature (T_a) for this reaction was determined to be 58°C. The PCR mixture was optimised to make up a total PCR reaction of 20µl as outlined in Appendix D1. Amplification was performed in the T100™ Thermal Cycler (*Bio-Rad Laboratories, Inc.*) using the thermal cycling conditions detailed in Appendix D1.

2.3.1.3 Agarose gel electrophoresis

The presence of the amplicon was confirmed by the electrophoresis of the PCR products electrophoresed through a 2% agarose gel stained with EtBr (*Sigma-Aldrich*). The PCR products were electrophoresed in 1 x TBE alongside a $1kb^+$ DNA ladder (*Invitrogen™*) to enable product size estimation. The gel was visualised as previously described in Section 2.2 and the expected PCR product size was 324 bp.

2.3.1.4 Restriction fragment length polymorphism analysis

Following amplification, RFLP analysis was used to determine the presence of the mutation. The restriction endonuclease (RE), Hpy*188I,* was selected using NEBcutter V2.0 (*New England BioLabs, Inc*). The digest conditions were as follows: the samples were incubated on a heating block at 37℃ overnight, followed by the inactivation of the enzyme at 65℃ for 20 minutes. The incubation period was prolonged as the standard digestion time of two hours was insufficient.

The digest products were electrophoresed through a 3% agarose gel stained with EtBr at 100V for one hour and 30 minutes. The gel images were visualised and captured as previously described in Section 2.2. Following fragmentation of the amplicon by the endonuclease, fragments of the following sizes were expected in the non-mutation carriers; 43 bp, 74 bp, 91 bp and 116 bp. When the mutation was present, an additional fragment size of 161 bp was observed as illustrated in Figure 2-1.

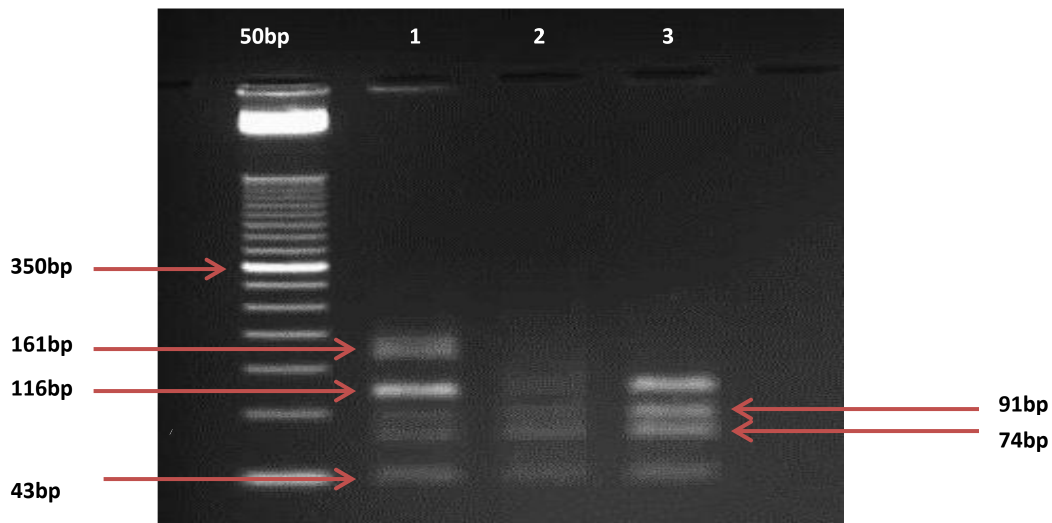


Figure 2-1: An agarose gel electrophoresis image illustrating the restriction digest fragments for the *BRCA2* c.5771_5774delTTCA mutation.
The digest products were electrophoresed alongside an Invitrogen 50 bp ladder. Lane 1 of the gel depicts a heterozygous individual or positive control (PC). Lanes 2 and 3 are the negative control (NC) samples of individuals who do not carry the mutation. For a homozygous individual (not depicted), the observed amplicon sizes would be expected at 161 bp, 116 bp and 43 bp.

2.3.2 *BRCA2* c.582G>A mutation

The presence of the BRCA2 c.582G>A mutation was investigated by using a tetra-primer amplification refractory mutation system-PCR (ARMS-PCR) for genotyping.

2.3.2.1 Primer design

The primers were designed using the Geneious version 7.1.9 software (*Geneious*). The primer sequences, primer directions and melting temperatures are indicated in Table 2-3. The difference in melting temperature between the primers was kept at less than 5℃ to ensure that the primer pairs anneal to the target region efficiently.

Table 2-3: Oligonucleotide sequences used to amplify the *BRCA2* region to detect the presence of the c.582G>A mutation

Primer name	**Sequence**	**Melting temperature (T_m)**
c.582_inner_F	5'- CTG AGG TGG ATC CTG ATA TGT CTA GG -3'	57.6℃
c.582_inner_R	5'- TGG GTG GTG TAG CTA AAG AAC TTC AT -3'	58℃
c.582_Cont_F	5'- CAA ATC TGT ACC TAG CAT TCT GCC TC -3'	57.3℃
C.582_Cont_R	5'- CAG AAG TAT TAG AGA TGA CAA TTA TCA ACC TC -3'	55.3℃

The sequence below indicates the region being targeted and the position of the primers in the sequence.

CAAATCTGTACCTAGCATTCTGCCTCatacaggcaattcagtaaacgttaagtgaaataaagagtgaatgaaaaaat
aatatccttaatgatcagggcatttctataaaaaataaactattttctttcctcccagggtcgtcagacaccaaaacatatttctga
aagtctaggagctgaggtggatcctgatatgtcttggtcaagttctttagctacaccacccacccttagttctactgtgctcataggt
aataatagcaaatgtgtatttacaagaaagagcagatGAGGTTGATAATTGTCATCTCTAATACTTCTG

The sequence regions colour-coded in green indicate the binding position of the outer primers. These primers are essential in amplifying the region of interest containing the single nucleotide change. The region in red and underlined is the position of the nucleotide change. The regions colour-coded in yellow indicate the binding position of the inner primers which are used to amplify the respective alleles of the nucleotide change to determine the presence of the mutation.

2.3.2.2 ARMS-PCR for the c.582G>A mutation

A temperature gradient experiment was performed to determine the optimal annealing temperature for all four primers in the same system. The temperatures tested ranged between 50℃ and 60℃; and 58.8℃ was selected as the most optimal annealing temperature. The PCR was set-up using the PCR components and volumes outlined in Appendix D2 which were added to make up a total reaction volume of 20µl. A non-template control (NTC) was also included in the set-up to monitor contamination in the system. PCR was performed in the T100™ Thermal Cycler (*Bio-Rad Laboratories*) and the cycling conditions are outlined in Appendix D2.

2.3.2.3 Agarose gel electrophoresis

Following amplification, the PCR products were loaded on a 2% agarose gel stained with EtBr and electrophoresed in 1 x TBE for 60 minutes at 150V. Alongside the PCR products was a 1kb^{+} GeneRuler size standard (*Thermo Fisher Scientific™*) which was used to estimate the size of the PCR products. The electrophoresis image was visualised as described in Section 2.2. The expected fragment sizes for a homozygous individual not carrying the muataion were 164 bp for the wild-type allele and 324 bp for the control amplicon. For a heterozygous mutation carrier, fragments with sizes 164 bp, 230 bp and 324 bp would be observed. The 230 bp amplicon indicated the presence of the mutation as indicated in Figure 2-2.

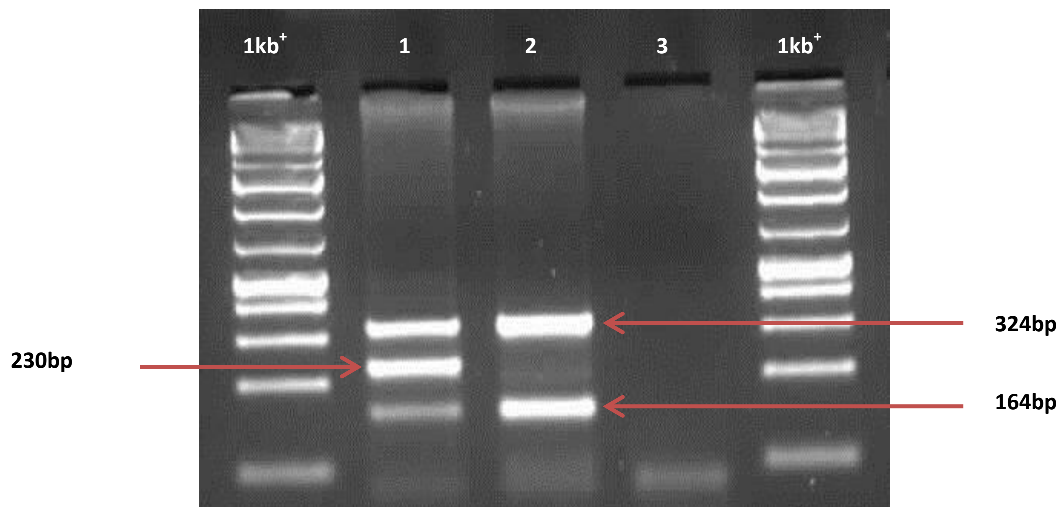


Figure 2-2: An agarose gel electrophoresis image illustrating possible outcomes of the ARMS-PCR for the *BRCA2* c.582G>A mutation.

The PCR products were electrophoresed alongside a 1kb^{+} GeneRuler ladder. In lane 1 is a heterozygous mutation carrier; the 164 bp fragment is the wildtype, the 230 bp fragment is the mutant and the 324 bp is the control fragment. Lane 3 is the non-template control which is used to determine whether contamination occurred. There are no amplicons in this lane, indicating that there was no contamination in the system. Where an individual is homozygous for the mutant allele (not depicted), amplicons only at 230 bp and 324 bp would be expected.

2.4 Haplotype analysis for the *BRCA2* c.5771_5774delTTCA and c.582G>A mutations

Haplotype analysis was performed for the two *BRCA2* mutations (c.5771_5774delTTCA and c.582G>A) to determine the presence of a founder effect for both of the mutations. The haplotypes were constructed by using SNP and microsatellite markers. Three SNPs across the *BRCA2* gene and two microsatellites flanking the gene region were used as shown in Figure 2-3. The microsatellites selected in the study were 17xAT upstream of the *BRCA2* gene and 21xTG located downstream of the gene. The 17xAT microsatellite lies upstream of the gene, approximately 47kb and the 21xTG microsatellite is approximately 409kb. The gene region containing the microsatellites was examined using the University of California, Santa Cruz (UCSC) Genome Browser https://genome.ucsc.edu/. The 17xAT and 21xTG microsatellites were selected as they were located closest to the *BRCA2* gene, in comparison to other microsatellites outside the gene region.

The SNP markers that were selected have the following dbSNP identification codes: rs4986860, rs9590940 and rs1801426 and were genotyped in mutation-positive individuals and control individuals. The genotypes of the control group were necessary to determine the frequency and distribution of alleles in the general population.

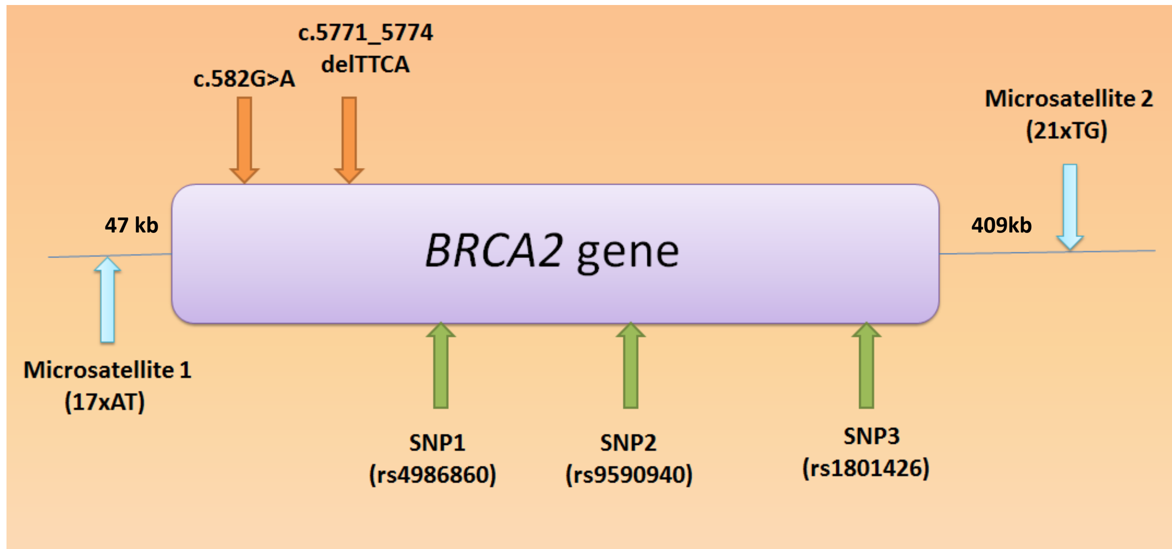


Figure 2-3: The position of five markers selected within and flanking the *BRCA2* gene in respect to positions of the *BRCA2* c.5771_5774delTTCA and c.582G>A mutations.
The five markers selected were two microsatellites (indicated by the blue arrows), and three SNPs indicated by the green arrows. The selected markers were used to determine whether a founder effect exists for the two *BRCA2* mutations that were genotyped.

2.4.1.1 Primer design for the microsatellite markers

Primers were designed to amplify the region containing each microsatellite to be genotyped as follows:

- Primers were designed manually to flank the region with the marker.
- Where possible, the 5' and 3'-terminus had either a guanine (G) or cytosine (C) nucleotide to enhance primer binding.
- The difference in T_m between the forward and reverse primer was kept at less than 5°C so that the primers annealed efficiently to the target.

- OligoAnalyzer Tool (*Integrated DNA Technologies,*) was used to determine primer properties such as primer length and T_m and whether there was self- or hetero-dimerisation and hairpin formation.
- Basic Local Alignment Search Tool (BLAST), an online database which finds regions of similarity between sequences was used to confirm that the primers bind to chromosome 13, upstream and downstream of the *BRCA2* gene.

The primer sequences used to amplify the microsatellite repeats are given in Tables 2-4 and 2-5.

Table 2-4: Oligonucleotide sequences used to amplify the region containing the 17xAT microsatellite upstream of the *BRCA2* gene

Primer name	Sequence
17xAT_microsat_F	*5'- CCA TAT TAT AGT TAA AGA GAA AGC TAG -3'
17xAT_microsat_R	5'- GGC ATT TTA AAA AAA GGA ATA CTA ATC -3'

*The forward primer was labelled with a 5' HEX fluorescent dye

A long stretch of adenine nucleotides was found upstream of the 17xAT microsatellite. The primer lengths were therefore increased to 27 bp to raise the melting temperatures and enhance primer binding and specificity.

Table 2-5: Oligonucleotide sequences used to amplify the region containing the 21xTG microsatellite downstream of the *BRCA2* gene

Primer name	Sequence
21xTG_microsat_F	*5'- ATT TCT TAT AAA CCC TTT CAG TC -3'
21xTG_microsat_R	5'- GTT CGT AAT ACG AAC TAT AAA C -3'

*The forward primer was labelled with a 5' HEX fluorescent dye

2.4.1.2 Polymerase chain reaction for microsatellite markers

The PCR set-up was performed by adding the PCR components to make up a reaction volume of 25µl as described in Appendix D3. The primer concentrations used for the

genotyping of the 17xAT microsatellite were 10µM. When the same primer concentrations were used to genotype the 21xTG microsatellite, no results were observed. The concentration was increased to 50µM, which still yielded no results. A concentration of 100µM for both the reverse and forward primers was tested and results were observed. This concentration was subsequently used to screen all mutation-positive samples and the control samples. The PCR was performed in a T100™ Thermal Cycler (*Bio-Rad*) for both microsatellites as shown in Appendix D3.

2.4.1.3 Capillary electrophoresis for the microsatellite markers

Capillary electrophoresis (CE) was used to separate the various fragments by size, in order to determine the alleles present at each microsatellite locus. The forward primers were modified with a fluorescent tag to enable detection of the fragments detection by laser excitation. The reaction was prepared by adding 1µl of undiluted PCR product, 0.3µl 500 ROX size standard (*Applied Biosystems*) and 8µl Hi-Di™ formamide to each well on the MicroAmp 96-well plate (*Applied Biosystems*). The plate was then placed on a 95℃ heating block for 2 minutes to denature the DNA and immediately cooled at a temperature of -20℃. The plate was loaded and analysed on the ABI 3130xl Genetic Analyzer (*Applied Biosystems*). GeneMapper version 4.0 and GeneMapper ID version 3.2 (*Applied Biosystems*) software were used to estimate the size of each amplicon generated during PCR. Allele sizes were estimated using the height of the highest peak observed. The formulae below were used to calculate the number of repeats each individual carried for the two microsatellites.

Formula for 17xAT = [(observed amplicon size-79)/2]

The 79 was derived as follows: 27 bases (forward primer) + 27 bases (reverse primer) + 25 bases (between the microsatellite repeat sequence and the primers)

Formula for 21xTG = [(observed amplicon size-52 bp)/2]

> The 52 was derived from: 23 bases (forward primer) + 22 bases (the reverse primer) + 7 bases (between the microsatellite repeat sequence and the primers)

Following the microsatellite genotyping, the SNP markers were genotyped to be used for the construction of haplotypes.

2.4.2 SNP genotyping

The *BRCA2* gene was explored to identify appropriate SNPs distributed across the gene. The SNPs were selected from the dbSNP database in NCBI (http://www.ncbi.nlm.nih.gov/projects/SNP/) and Ensembl Genome Browsers (http://www.ensembl.org/index.html). Criteria for selection were that the SNPs occurred within the coding region and had a minor allele frequency (MAF) greater than 0.05 (or 5%). SNPs within the coding region were selected as they may directly affect the way in which a gene product functions, as well as its interaction with other proteins. The MAF is a score used to indicate how diverse and frequent a SNP is in the general population. The greater the MAF, the more frequent the SNP is in the population. It is also important to note that the MAF can differ between populations and where a SNP is associated with disease, the MAF will differ between the case and control groups. For this study, SNPs selected were required to have a MAF greater than 5% in individuals with African ancestry and not individuals in the general population. Table 2-6 shows the SNP identification codes and genomic position according to the GRCh37/h19 human genome build as well as the technique that was chosen to genotype each SNP.

It was, unfortunately, noted by the candidate very late in the research process that the criterion related to MAF was incorrectly applied and thus, selected SNPs had a MAF less than 5%, indicative of rare SNPs instead of those frequently occurring in the population. The methodology and results discussed here represent SNP results generated and analysed despite the knowledge of the candidate's error and the resultant. A discussion of

the appropriate SNP selection and future plans to rectify this are elaborated on in Section 4.2.

Table 2-6: Single nucleotide polymorphisms selected for the haplotype study of the *BRCA2* gene

SNP ID	Reference SNP alleles	Genomic position	Minor allele frequency (MAF) 1000 Genomes	Genotyping technique
rs4986860	A/G	13:32929309	G = 0.01	Real-time PCR
rs9590940	A/C	13: 32944667	C = 0.01	RFLP
rs1801426	A/G	13: 32972884	G = 0.04	Real-time PCR

2.4.2.1 TaqMan SNP genotyping

Of the three SNPs selected for the haplotyping analysis, two (rs4986860 and rs1801426) were genotyped using predesigned TaqMan® SNP Assays (*Applied Biosystems*). The TaqMan SNP probe-based assays rely on sequence-specific detection so that only the region of interest is amplified and specific single base changes are detected (McGuigan & Ralston, 2002). The commercially available assays comprise of two primers (forward and reverse) which flank the SNP region of interest and two TaqMan probes. One probe is labelled with VIC® reporter dye which is used to detect Allele 1 of the SNP and the other is labelled with a FAM™ reporter dye to detect Allele 2. The probes are labelled with a fluorescently labelled reporter dye on the 5' end and make use of the 5' nuclease activity of the Taq polymerase enzyme. At the 3' end of the probes is a quencher dye which is not fluorescently labelled. The quencher is responsible for suppressing fluorescence when the probe has not bound to its complementary sequence on the target DNA. The assay allows for the target of interest to be monitored throughout the entire amplification run, in real-time (McGuigan & Ralston, 2002).

Prior to amplification, the DNA and assay components were prepared on a MicroAmp 384-well plate (*Applied Biosystems*). The assay was optimised using various parameters and the most optimal reaction was determined to be one in which the 40X (stock) probe assay was used with 2.25µl of a 10ng/µl concentration of gDNA. The PCR set-up is outlined in Appendix D. The plate was loaded onto the 7900HT Fast Real-Time PCR System (*Applied*

Biosystems). The amplification thermal cycling conditions are detailed in Appendix D4. Sequence Detection System version 2.3 (*Applied Biosystems*) was used for capturing of data and subsequent analysis.

2.4.2.2 RFLP SNP genotyping

The third SNP, rs9590940, was genotyped by RFLP. The primer sequences are found in Table 2-7 below.

Table 2-7: Oligonucleotide sequences used to amplify the region containing the single nucleotide polymorphism with ID rs9590940 (Chen, 2016)

Primer Name	Sequence
BC2 19 fwd US2 01	5'- TAAGGCAGTTCTAGAAGAATGAA -3'
BC2 19 rvs US1 01	5'- GAAACTCATCTCAAACAAAC -3'

PCR amplification was optimised prior to genotyping to determine the ideal temperature at which primers bind to the target DNA. The range of temperatures tested was between 55℃ and 65℃. The optimal annealing temperature was determined to be 60℃. The PCR was set-up to make up a 20µl reaction as shown in Appendix D4. Amplification was performed in the Bio-Rad T100™ thermal cycler (*Bio-Rad Laboratories, Inc*) by using the conditions detailed in Appendix D4. Following amplification, PCR products were subjected to restriction by digesting the products with Mlu*CI* endonuclease which was selected using NEBcutter (*New England BioLabs, Inc*).

For the fragmentation reaction, 10µl of the PCR product was added to 2µl of the Mlu*CI* enzyme, 2µl of the CutSmarter buffer and 6µl of ddH_2O in a final volume of 20µl.The reactions were incubated at 37℃ overnight and the enzyme subsequently inactivated at 80℃ for 20 minutes. The digested products were resolved on a 3% agarose gel stained with EtBr alongside a 25 bp *Invitrogen™* ladder in 1 x TBE at 100V for approximately 90 minutes for genotype determination. The expected fragment sizes were 46 bp, 74 bp, 140

bp in the presence of the adenine (A) allele. The cytosine (C) allele abolishes a RE site and results in as 46 bp fragment and a larger fragment of 214 bp shown in Figure 2-4.

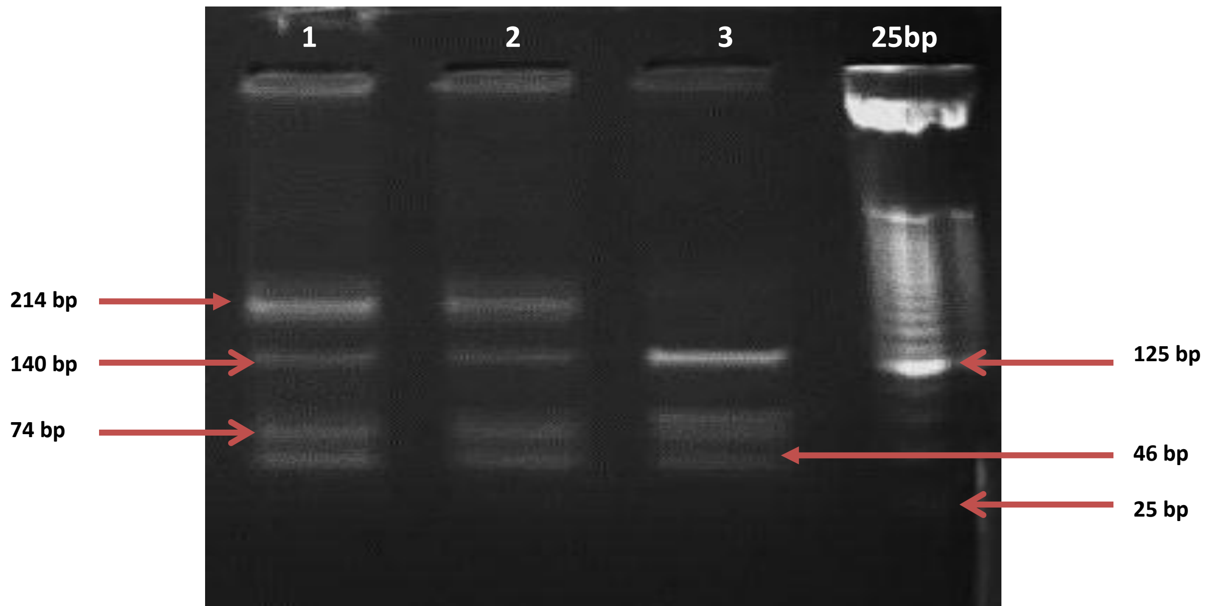


Figure 2-4: An agarose gel image following digest using the *MluCI* enzyme to detect the alleles of the rs9590940 SNP.
Lanes 1 and 2: heterozygous (A/C) genotype determined by the presence of all three fragment sizes (46 bp, 74 bp and 140 bp). Lane 3: homozygous (A/A) genotype determined by the presence of only two fragment sizes (46 bp and 214 bp). The digest products were electrophoresed alongside a 25 bp ladder to enable size estimation.

2.4.3 Hardy-Weinberg equilibrium

Subsequent to SNP genotyping, the results were used to determine whether the SNPs are in Hardy-Weinberg equilibrium. Hardy-Weinberg equilibrium (HWE) suggests that allele and genotype frequencies in a particular population will remain constant from one generation to the next, given that no evolutionary influences act on the population (Wigginton *et al*., 2005). Evolutionary influences may include the presence of mutations, genetic drifts, non-random mating, migration or natural selection (Wigginton *et al*., 2005). If the population is in HWE, the sum of the allele and genotype must equal 1 (Rodriguez *et al*., 2009). That is:

Allele frequencies: $p + q = 1$ and Genotype frequencies: $(p + q)^2 = p^2 + 2pq + q^2 = 1$

In this study, the DNA control samples were genotyped and the selected SNPs were tested to determine whether HWE exists for the SNPs. This was important to eliminate the likelihood of associating the markers with disease association based on the frequencies at which they occur in the control sample population. The online tool, gPLINK (freely accessible from http://zzz.bwh.harvard.edu/plink/gplink.shtml) was used to calculate the HWE p-value. A p-value of less than 0.05 is indicative of a SNP that is not in HWE.

2.4.4 Linkage disequilibrium

Linkage disequilibrium (LD) is defined as the non-random association of alleles between different loci more often than would be expected by chance (Slatkin et al., 2008). Two measures are often used to estimate LD; they are linkage disequilibrium correlation coefficient (D') and correlation coefficient of allele frequencies (r^2) (Slatkin *et al*., 2008). D' is the difference between the observed and the expected frequency of a given haplotype and r^2 is the correlation between a pair of loci. If D' and r^2 are both equal to one, this indicates complete and perfect LD, respectively; that is, two loci have not been separated by recombination events and will be co-inherited together (Slatkin *et al*., 2008). The Ensembl genome browser (https://m.ensembl.org) was used to estimate LD between the two microsatellite markers (17xAT and 21xTG).

2.4.5 Haplotype analysis using PHASE

In this study, PHASE software (version 2.1.1) was used to assign possible haplotypes belonging to each individual. Microsatellite and SNP genotypes from mutation-positive individuals and unrelated controls were used to determine potential allelic combinations inherited on each chromosome. PHASE is freely available online from http://stephenslab.uchicago.edu/software.html. The program makes use of a Bayesian statistical method to remodel haplotypes by predicting those that are expected to occur naturally in the population (Stephens *et al*., 2001). The software can utilise SNP, microsatellite, tri-allelic SNP and Human Leukocyte Antigen (HLA) allele data (Stephens *et*

al., 2001; Stephens & Donnelly, 2003). The software was run on a Linux operating system. The file extraction of the files and the commands used to input data into the software are outlined in Appendix E. Phasing of control and patient samples was performed to investigate the presence of disease-associated haplotypes. Haplotypes found to be highly associated with the respective *BRCA2* mutations and rare in the general population or the control sample would support the existence of a founder effect.

2.5 Screening for the *FANCG* c.637-643delTACCGCC mutation

DNA samples belonging to 119 individuals with breast cancer who tested negative for *BRCA1* and *BRCA2* mutations were screened to determine whether the *FANCG* c.637-643TACCGCC mutation confers a breast cancer risk in its heterozygous state. The primer sequences used for screening can be found in Table 2-8 below.

Table 2-8: Oligonucleotide sequences used to amplify the region containing the *FANCG* c.637_643delTACCGCC mutation (Morgan *et al*., 2005).

Orientation	**Primer sequence**
Forward	*5'- CCC AGG GAT TGA AGG ATG TC -3'
Reverse	5'- GCA TGA GAC TGG AGG ACC AC -3'

*The forward primer was fluorescently labelled with a 5' 6-FAM dye.

The DNA was amplified and genotyped using CE by using the same conditions outlined in Appendix D3. The PCR products were diluted to 1:50 concentration using 1xTE buffer and these were subsequently used for CE analysis.

2.6 Next generation sequencing

In this study, a commercially available targeted gene panel was selected for screening in order to assess its utility in a South African patient cohort. The BRCA Hereditary Cancer MASTR™ (Multiplex Amplification of Specific Targets for Resequencing) Plus kit was supplied by *Multiplicom NV*. The panel is target-amplification based and 561 amplicons were generated following library building. Primers are designed to bind to the specific genomic regions of interest such that only those regions are amplified to produce libraries and to be sequenced later .The assay was used to identify various germline mutations and copy number variations (CNVs) in 26 genes (Table 2-9) known to predispose individuals to inherited breast, ovarian, pancreatic and other related cancers.

Table 2-9: The 26 cancer susceptibility genes included in the Multiplicom panel for inherited breast cancer.
The table shows the chromosomal location of the genes, their function and the associated cancer phenotypes and other diseases which can arise when mutations are identified (Genetics Home Reference, 2018).

Gene	OMIM	Chromosomal location	Normal Function	Associated cancer/s	Other associated disease/s	Disease characteristics
ATM	607585	11q22.3	Regulation of cell growth and division	Breast, ovarian, stomach, bladder cancers	Ataxia-telangiectasia	Telangiectasia, muscle and neurological degeneration
BARD1	601593	2q35	DNA repair and transcriptional regulation	Breast and ovarian cancers, and neuroblastoma	N/A	N/A
BLM	210900	15q26.1	Unwinding of DNA helix	Breast, colorectal, skin cancers and melanoma	Bloom syndrome	Short stature, skin rash upon sun exposure
BRCA1	113705	17q21	DNA repair and regulates embryonic development	Breast, ovarian, prostate, colon and pancreatic cancers	N/A	N/A
BRCA2	600185	13q12.3	DNA repair and transcriptional regulation	Breast, ovarian, prostate and colon cancers	Fanconi anaemia	Aplastic anaemia, bone marrow failure
BRIP1	605882	17q23.2	Double-stranded break repair	Breast cancer	Fanconi anaemia	Aplastic anaemia, bone marrow failure
CDH1	192090	16q22.1	Cell-to-cell adhesion	Breast, ovarian, prostate and gastric cancers	Blepharocheilodontic (BCD) syndrome	Congenital disorder. Lower eyelids and inability of eyelids to close completely in some instances
CHEK2	604373	22q12.1	Regulation of cell growth and division	Breast, ovarian and prostate cancers	N/A	N/A
EPCAM	185535	2p21	Cell-to-cell adhesion	Breast, ovarian, stomach and skin cancers	Lynch syndrome	Large intestine and rectum malignancy
FAM175A	611143	4q21.23	DNA repair and regulation of the cell cycle	Breast cancer	N/A	N/A
MEN1	131100	11q13.1	Regulation of cell growth and division	Breast, pancreatic and lung cancers	Multiple endocrine neoplasia	Endocrine gland malignancies
MLH1	120436	3p22.2	DNA repair	Breast and ovarian cancers	Lynch syndrome	Large intestine and rectum malignancy
MRE11A	600814	11q21	Homologous recombination and maintenance of telomere length	Breast and ovarian cancers	Ataxia-telangiectasia-like disorder 1	Abnormal eye movements, cerebellar ataxia

Gene	OMIM	Chromosomal location	Normal Function	Associated cancer/s	Other associated disease/s	Disease characteristics
MSH2	609309	2p21-p16.3	DNA repair	Ovarian cancer	Lynch syndrome	Large intestine and rectum malignancy
MSH6	600678	2p16.3	DNA repair	Ovarian cancer	Lynch syndrome	Large intestine and rectum malignancy
MUTYH	604933	1p34.1	DNA repair	Breast, thyroid and duodenal cancers	Familial adenomatous polyposis	Adenomatous polyps
NBN	602667	8q21.3	DNA repair	Breast, ovarian and prostate cancers	Nijmegen breakage syndrome	Short statue, recurrent respiratory tract infections, intellectual disability
PALB2	610355	16p12.2	Homologous recombination repair	Breast and ovarian cancers	Fanconi anaemia	Aplastic anaemia, bone marrow failure
PMS2	600259	7p22.1	DNA repair	Ovarian cancer	Lynch syndrome	Large intestine and rectum malignancy
PTEN	601728	10q23.31	Regulation of cell growth and division an promotes apoptosis of damaged cells	Breast, lung and prostate cancers	Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome	Growth of multiple hamartomas in the skin, thyroid glands or breast tissue
RAD50	604040	5q31.1	Double-standed break repair	Breast and ovarian cancers	Nijmegen breakage syndrome	Short statue, recurrent respiratory tract infections, intellectual disability
RAD51C	602774	17q22	Homologous recombination repair	Breast and ovarian cancers	Fanconi anaemia	Aplastic anaemia, bone marrow failure
RAD51D	602954	17q12	Homologous recombination and repair	Breast and ovarian cancers	N/A	N/A
STK11	602216	19p13.3	Regulation of cell growth and division, maintenance of energy in a cell	Breast, ovarian, pancreatic and skin cancers	Peutz-Jeghers syndrome	Growth of multiple hamartomas in the gastrointestinal tract
TP53	191170	17p13.1	Regulation of cell growth and division	Breast, bladder, ovarian cancers and head and neck squamous cell carcinoma	Li-Fraumeni syndrome	Predisposition to the development of cancerous tumours
XRCC2	600375	7q36.1	Homologous recombination repair	Breast cancer	Fanconi anaemia	Aplastic anaemia, bone marrow failure

2.6.1 DNA purification of genomic DNA

Prior to library preparation and sequencing, the DNA samples were purified using the QiAmp DNA Micro Kit (*Qiagen*) which utilises silica-based membranes. The DNA volume utilised for purification was 10µl. The purification step ensures that buffers or RNA and protein contaminants which may interfere with or inhibit the sequencing reaction are removed. The DNA was eluted in a final volume of 30µl in nuclease-free water.

Following elution, the samples were quantified using the NanoDrop ND-1000 spectrophotometer (*NanoDrop technologies, Inc.*). The selected samples had OD260/280 and OD260/230 ratios of 1.8 or greater as recommended by the manufacturer. The selected samples were then diluted in nuclease-free water to a final concentration of 10ng/µl. A qubit quantitation assay on the Qubit® 3.0 fluorometer (*Invitrogen™*) was used to measure the final DNA concentration for each sample.

2.6.2 Library preparation

A total of five PCR mixes or plexes (plex 1 to plex 5) were supplied in the kit. Each PCR plex consists of oligonucleotides specific to different regions targeting the 26 genes as well as deoxynucleotide triphosphates (dNTPs). The amplification is expected to generate a total of 561 amplicons. The manufacturer's instructions were followed for a 24 sample assay; 240µl of each PCR plex was added to 3.6µl of the Taq DNA polymerase to make up the respective master mixes (1-5). For each plex PCR, 10µl of the mastermix is combined with 5µl of the DNA sample. Therefore for 24 samples, the set up involves 120 separate reactions. A negative control (no DNA template) was included to ensure no contamination occurred during the PCR set-up. Amplification was performed on the GeneAmp® PCR System 9600 (*Applied Biosystems*) as outlined in Appendix F. For detailed library preparation steps, please refer to the Multiplicom protocol in Appendix F.

2.6.2.1 Purification of amplicon libraries to remove unbound primers

The PCR products from the five plex reactions were combined using the mixing scheme outlined in the assay specification sheet (Appendix G). The 24 amplicon libraries were purified using magnetic Agencourt® Ampure® XP beads (*Beckman Coulter, Inc*) to remove buffer, excess unbound primers and dNTPs. Twenty microlitres of the amplicon library was combined with 24µl of the beads. Amplicons bind to the beads when introduced to a magnetic field and produce a pellet. The unwanted products aggregated in the supernatant and could then be subsequently discarded. The bead-amplicon pellet was then washed twice with 200µl of freshly prepared 70% ethanol and the ethanol discarded. The pellet was allowed to dry for five minutes to evaporate the ethanol completely to avoid its interference with the sequencing reaction. The pellet was re-suspended in 20µl of nuclease-free water and this combination was placed on a magnetic stand and allowed to separate. The eluent contained the purified amplicons and was transferred to a clean tube.

2.6.2.2 Quality control to confirm the annealing of primers to target genes

A quality control step was undertaken to confirm that the primers targeting the 26 genes of interest had successfully annealed to the target regions. This was performed by combining 5µl of the amplicon library with 3µl of Ficoll dye, which was then electrophoresed through a 2% agarose gel stained with GelRed™ (*Biotium, Inc*), alongside a 1kb+ GeneRuler ladder (*ThermoFisher Scientific™*). The gel images were captured by Omega Fluor™ gel documentation system (*Gel Company*) as illustrated in Figure 2-5. The expected fragment size was approximately 450 bp.

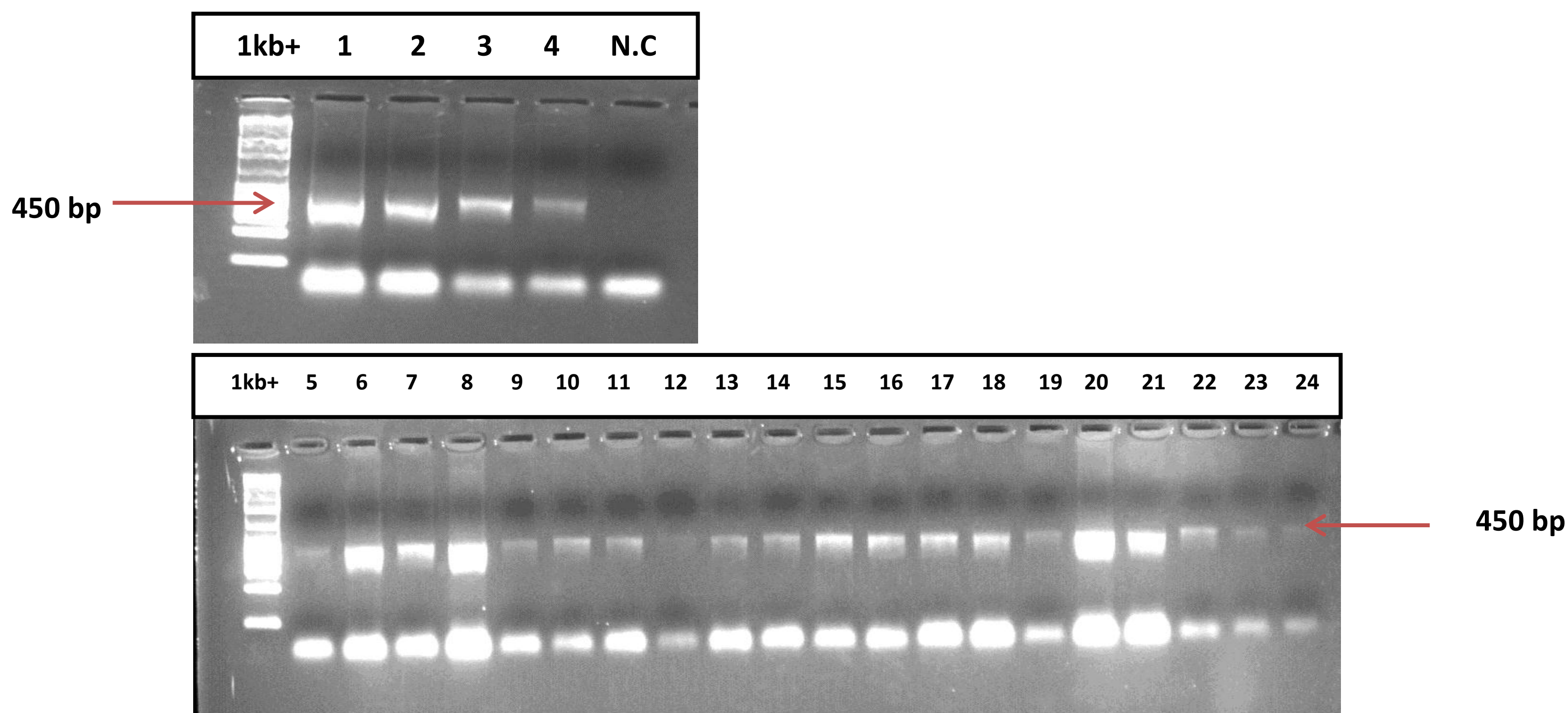


Figure 2-5: An agarose gel image showing the purified amplicon libraries.
All 24 samples were amplified with primers targeting the genomic regions containing the 26 cancer susceptibility genes under study. The samples were electrophoresed alongside a 1kb+ marker. Successful amplification is shown by the presence expected of the expected 450 bp fragment in lanes 1 to 24. Excess primers are seen at the bottom of the gel image, smaller than 100 bp (the smallest band on the ladder). The negative control (N.C) has no amplicon present, indicating that no contamination occurred during the amplification set up.

2.6.2.3 PCR for the annealing of adaptors and molecular identifiers

Following purification, each library was tagged with p5 and p7 adaptors as well as molecular identifiers (MIDs). Dual indexing was used so that each sample has a unique combination of two barcodes attached at the 5' and 3' ends. PCR was performed to bind adaptors and MIDs to the amplicon libraries. The purified amplicon libraries were diluted in nuclease-free water in a two-step serial dilution to 1:1000, 1:500 or 1:250 depending on the intensity of the PCR product amplicon shown in Figure 2-5. Two microliters of the diluted amplicon libraries were used for the annealing amplification of MIDs and adaptors as detailed in Appendix F.

2.6.2.4 Quality control to confirm annealing of adaptors and molecular identifiers

Following PCR amplification, a quality control (QC) step was undertaken using a 2100 Bioanalyzer (*Agilent Technologies*). The Bioanalyzer can be used to quantify or size DNA, RNA and proteins. It was used here to confirm successful amplification by detecting a digital signal between 350 and 550 bp.

2.6.2.4.1 Purification of the tagged amplicons

The tagged amplicon libraries were purified again using the Ampure beads. This was to ensure that excess unbound primers (as seen in Figure 2-5), adaptors and molecular identifiers are removed as they could interfere with sequencing. Smaller fragments tend to be preferentially sequenced and this could lead to a loss of sequence data from the actual sample libraries that are meant to be sequenced.

A total of thirty microliters of the amplicon library was combined with 25.5µl of the Ampure beads (as specified in Appendix G), mixed by pipetting several times and allowed to incubate at room temperature for five minutes to ensure that the amplicons had efficiently bound to the beads. The mixture was then subjected to a magnetic field and the purification and elution of DNA performed as previously described in Section 2.6.1.

2.6.2.4.2 Pooling of the purified tagged amplicon libraries

The purified amplicon tagged libraries were quantified using a spectrophotometer. The concentrations were then calculated in nanoMolar (nM) as follows:

$$\text{Sample molarity (nM)} = \frac{\text{sample concentration [ng/µl]} \times 10^{6}}{656.6 \times \text{average amplicon size}}$$

Subsequent to calculating the molarity, the purified libraries were diluted to 4 nM in Tris-Cl buffer and 5µl of each library were pooled in a 1.5ml microcentrifuge tube and mixed thoroughly. Five microliters of the pooled libraries was combined with 5µl of Sodium Hydroxide (NaOH) at normality of 0.2N to denature the amplicons. This was followed by adding 990µl of pre-chilled HT1 hybridisation buffer to the pooled libraries to dilute them to a concentration of 20pM. The pooled libraries were quantified on the Bioanalyzer to confirm the final concentration.

The PhiX control v3 was also diluted to 4 nM by combining 2µl of 10nM PhiX library and 3µl of 10mM Tris-Cl. The PhiX library is from the PhiX genome and is used as an internal control for the sequencing and alignment to estimate phasing and error rate. The diluted control was also denatured by adding 5µl of 0.2 N NaOH to 5µl of the 4nM PhiX library and diluted to 20pM with 990µl of pre-chilled HT1 buffer. The PhiX control 5% spike-in was performed by combining 30µl of the denatured and diluted PhiX to 570 µl of the pooled library. This was followed by loading the library onto the reagent cartridge from the Illumina Miseq reagents kit v3 (Illumina). For detailed instructions on denaturing and diluting libraries, the Illumina custom protocol which can be found by accessing the following URL:

https://support.illumina.com/content/dam/illumina-support/documents/documentation/system_documentation/miseq/miseq-denature-dilute-libraries-guide-15039740-03.pdf

2.6.3 Bridge amplification and sequencing-by-synthesis

Prior to sequencing, a manifest file and sample sheet were uploaded to the Miseq instrument. The manifest file comprises amplicon and primers sequences which are being targeted in the assay. The sample sheet contains a list of the samples and the unique barcode sequences attached to each sample. Following preparation, the samples were

amplified in the MiSeq instrument by bridge amplification and sequenced by synthesis by attaching complementary, reversible dNTPs to the strand.

2.6.4 Alignment and variant calling

Following the sequencing run, reads with similar nucleotide orders were clustered together and the forward and reverse reads were paired. The reads were aligned to form contigs (overlapping fragments of DNA which form part of a specific region of DNA). The contigs were aligned to a reference genome for comparison and to identify variants. The MASTR Reporter data analysis software (*Multiplicom NV*) was used for contiguous sequence alignment to the reference genome and for variant calling. The software provides information such as read depth, zygosity, location of variant in the genome as well as the variant change observed in the sequence.

A software tool called wANNOVAR, which is freely available at http://wannovar.wglab.org/index.php, was used for the functional annotation of the variants identified by the MASTR Reporter software. The tool extracts data from public databases such as 1000 Genomes, Exome Aggregation Consortium (ExAC), ClinVar and dbSNP.

The 1000 Genomes database characterises whole-genome human genetic variation (in the form of SNPs, insertions/deletions and structural variants) from various populations across the different continents (Auton *et al*., 2015). The ExAC database contains exome sequencing data from different disease-specific and population genetic studies (Lek *et al*., 2016). The ClinVar database combines human genetic variation with clinical phenotypes (Landrum *et al*., 2014). The dbSNP database is a collection of genome-wide SNPs and other small-scale genetic variations such as insertions/deletions and microsatellites (Sherry *et al*., 2001). wANNOVAR makes use of these public databases to provide as much

comprehensive information about a variant as possible. For variants which had not been previously classified and with no information available on wANNOVAR and the public databases, five *in silico* prediction tools were used to score pathogenicity. The prediction tools are listed below:

Variant effect predictor – The predictor tool is part of the Ensembl database, and can be utilised to annotate, analyse and provide an impact score of variants in both coding and non-coding regions of the genome (McLaren *et al*., 2016). The software scores variants as having either low, modifier, moderate and high impact. Low and modifier impact scores suggest that the variants lie in the location of a transcript where they do not produce a damaging consequence on the protein product. The modifier impact, in addition, suggests that the variant change may not alter protein function but it is possible that it is regulating the expression of a transcript that lies downstream (McLaren *et al*., 2016).

Combined Annotation Dependent Depletion (CADD) – A prediction tool that is used to determine the degree of pathogenicity of a variant (Kircher *et al*., 2014). The tool scores SNVs as well as insertions and deletions by using multiple annotations (such as allelic diversity, pathogenicity and regulatory function) and comparing variants that have survived natural selection to variants that have been simulated by computer programs. The tool can predict variants in coding and non-coding regions by providing a numerical score. A score of 15 or more is suggestive of a potentially deleterious variant (Kircher *et al*., 2014).

Sifting Intolerant from Tolerant (SIFT) - A tool that predicts the pathogenicity of SNVs (missense and silent) which are located in the coding regions and are thought to affect protein function. The prediction tool scores variants as either tolerated or deleterious (Sim *et al*., 2012).

Polymorphism Phenotyping v2 – A tool that predicts the possible effect that an SNV has on the function of a protein and its folding structure. The tool scores the variants as either benign, possibly deleterious and probably deleterious (Adzhubei *et al*., 2013).

MutationTaster – A tool that is used to for variant annotation and to determine the pathogenicity of a sequence variant. The tool evaluates variants that are protein coding and those that are non-coding. The software predicts variants as either polymorphic or disease-causing (Schwarz *et al*., 2014).

Based on the outcomes of these *in silico* analyses, variants that were identified and previously unclassified were assigned with possible classification using the International Agency for Research on Cancer (IARC) Unclassified Genetic Variants Working Group Five-class System (Plon *et al.,* 2008). Table 2-10 below describes the five class system of variants proposed for use in genetic testing.

Table 2-10: The proposed classification system for sequence variants identified in genetic testing by the International Agency for Research on Cancer (Plon et al., 2008).

Class	Description
5	Definitely Pathogenic
4	Likely Pathogenic
3	Uncertain
2	Likely Not Pathogenic or of Little Clinical Significance
1	Not pathogenic or of No Clinical Significance

3 RESULTS

In Chapter 3, the results and outcomes of the different objectives will be discussed. The chapter begins with Section 3.1 which provides a description of the genotyping data from the two *BRCA2* mutation genotyping experiments. The results obtained from the haplotyping analysis follows in Section 3.2. Section 3.3 presents the results from the *FANCG* mutation screen. The chapter concludes with a presentation of the results from the targeted gene panel component of the study in Section 3.4.

3.1 Mutation screening of the *BRCA2* c.5771_5774 and c.582G>A mutations

A group of black South African women diagnosed with breast cancer and thought to be at increased risk of having inherited breast cancer were screened for two *BRCA2* mutations of interest based on previous research in the Division of Human Genetics (Wits/NHLS). Genotyping was performed to determine the frequency of these mutations in the sample group and whether the mutations can in future be employed for targeted diagnostic testing.

A single patient of the total 35 tested (equating to 2.86% of the sample) was found to carry the c.5771_5774delTTCA mutation. The positive result obtained for the c.5771_5774delTTCA mutation was verified by the Molecular Diagnostic Laboratory at the Division and communicated to the patient by a Genetic Counsellor. None of the 35 individuals screened for the c.582G>A mutation were identified to carry it.

Having observed these two mutations across multiple studies in two different sample groups of unrelated individuals (i.e. breast cancer patients and FA patients), the overall importance of the mutations in inherited cancer syndromes was interrogated. To contextualise and understand their importance, literature from all studies which involved

BRCA2 screening were perused to determine and compare the frequencies at which these mutations were present in the different studies. Table 3-1 summarises the occurrence of the two *BRCA2* mutations in all breast cancer and FA studies conducted in the Division to date including this study.

Table 3-1: A summary of the *BRCA2* c.5771_5774delTTCA and/or c.582G>A mutations identified in the breast cancer and Fanconi anaemia studies conducted in the Division of Human Genetics (Wits/NHLS)

Study	**Sample size**	**c.5771_5774delTTCA**		**c.582G>A**	
		N	Frequency (%)	N	Frequency (%)
Chen, 2015	33	0	0.00	1	3.03
Francies *et al*., 2015	85	1	1.18	1	1.18
Feben *et al*., 2017	4*	2	50.00	1	25.00
Pitere, 2018 (Current)	35	1	2.86	0	0.00
Total	157	4	2.55	3	1.91

*Individuals in this study were from two families in which the probands were found to be biallelic carriers of *BRCA2* mutations. One proband was a compound heterozygote for the two *BRCA2* mutations investigated in this study and the other proband was a compound heterozygote of the c.5771_5774delTCCA mutation and another *BRCA2* mutation not under investigation in this study. The study was focussed on Fanconi anaemia, while all the other studies in the table were on breast cancer.

A total of 157 individuals have been investigated among the various studies of these, the c.5771_5774delTTCA mutation has been identified four times while the c.582G>A mutation has been identified three times. In the study performed by Feben *et al*., 2017, two unrelated families were investigated where the probands were diagnosed with FA. Upon genetic testing, one of the probands was found to be heterozygous for the c.5771_5774delTTCA and another *BRCA2* mutation, while the other proband was found to be a compound heterozygote of both the *BRCA2* mutations investigated in this study. In both families, no history of an IBC syndrome was reported. The remaining studies focused on samples of patients who were diagnosed with breast cancer at younger ages. Collectively, these two mutations account for only 4.46% across the various studies. The

genotyped mutations were further investigated to determine whether they could potentially indicate the presence of a founder effect.

3.2 Haplotype analysis

As discussed in section 2.4.2, the methodology employed to carry out this section of the study was inherently flawed as a result of incorrectly selecting rare SNPs instead of frequently occurring SNPs. The data is nonetheless presented here to demonstrate the work of the candidate.

A total of five mutation carriers were included for haplotyping as DNA samples were only available from those samples; one of the samples was positively identified for the mutation in this study and the remaining four were positively screened in previous studies. Of the five mutation carrier samples investigated in this study, two carried the *BRCA2*c.582G>A mutation and three had the c.5771_5774delTTCA mutation. In addition to the mutation carriers, 100 controls were included for the haplotype analysis to investigate the presence of a founder effect in the black South African population for the two above-mentioned mutations. Five markers were selected for the haplotype analysis as previously mentioned in Section 2.4 and all the genotypes obtained for all mutation carrier samples and the controls are listed in Appendix H.

The microsatellites were genotyped using CE. Of the two microsatellites genotyped, the 17xAT microsatellites showed diverse variation of allele sizes compared to the 21xTG where the variation observed was minimal. As a MAF of less than 0.05 was selected in error, it is not surprising that the SNP markers showed only minor variation in the control samples and only a few heterozygous individuals were identified. In contrast, the mutation-positive individuals were found to be monomorphic for the major allele “A” at all three SNPs indicating that no variation was present in these individuals. Table 3-2

presents the genotypes identified for each of the three SNPs and the number of heterozygotes for each marker. None of the genotyped individuals were observed to be homozygous for the minor alleles ("C" and "G") for all SNP markers.

Table 3-2: A summary of the genotypes observed for the three *BRCA2* single nucleotide polymorphisms in the control group (n=100)

	rs4986860 (A/G)	**rs9590940 (A/C)**	**rs1801426 (A/G)**
Homozygous (major allele)	AA = 95	AA = 95	AA = 77
Heterozygous	AG = 5	AC = 5	AG = 23
Homozygous (minor allele)	GG = 0	CC = 0	GG = 0

3.2.1 Hardy-Weinberg equilibrium

Subsequent to SNP genotyping, the SNP results were used to calculate HWE to determine whether there are evolutionary forces (i.e. natural selection or inbreeding) acting on the population. Table 3-3 shows the different HWE p-values computed using the gPLINK software. A p-value of more than 0.05 indicated a SNP that was in HWE. The data in Table 3-3 show that all SNP p-values computed were greater than 0.05 indicating that the SNPs were indeed in HWE.

Table 3-3: The frequencies of observed and expected heterozygotes, and Hardy-Weinberg equilibrium p-values computed using gPLINK for each SNP located in the *BRCA2* gene and selected for haplotyping.

SNP ID	**Genotype**	**Observed heterozygote frequencies**	**Expected heterozygote frequencies**	**P-value**
rs4986860	AG	0.05	0.04875	1
rs9590940	AC	0.05	0.04875	1
rs1801426	AG	0.23	0.2036	0.3534

3.2.2 Linkage disequilibrium

Linkage disequilibrium was measured for the 17xAT and 21xTG microsatellite markers to determine whether they are inherited together. The results obtained from the Ensembl genome browser measured the average linkage disequilibrium correlation coefficient (D') as 0.92 and the correlation coefficient of allele frequencies (r^2) as 0.23. The simulated r^2 indicates that the alleles at the loci are rare. These results, therefore, indicate that both loci will always be co-inherited (due to high D') when the rare alleles are present.

3.2.3 Haplotype analysis

Haplotypes were constructed using SNP and microsatellite marker genotyping data. The analysis was performed to determine how frequently the haplotypes observed co-segregate with disease occur in the general population and whether a founder effect exists for either of the two mutations in the black South African population. For the mutation-positive samples, seven possible haplotypes were constructed using the PHASE software and 93 haplotypes were estimated for the control samples (Appendix I). The paired constructed haplotypes for individuals with the c.582G>A and c.5771_5774delTTCA are presented in Tables 3-4 and 3-5, respectively. The haplotypes were analysed separately by comparing mutation-carriers of each mutation. This was performed to determine whether similar haplotypes segregate with the same mutation and to provide evidence for or against the presence of a founder effect.

Table 3-4: The haplotypes constructed by the PHASE software for the mutation carriers of the *BRCA2* c.582G>A mutation.

Mutation	Patient	Possible haplotypes
c.582G>A	1	11 A A A 21 11 A A A 21
	2	12 A A A 19 22 A A A 19

The haplotypes in are presented by the order of markers as such: 17xAT SNP1-SNP2-SNP3 21xTG (as illustrated in Figure 2-3). The results obtained and presented in Table 3-4 indicate that individuals with the *BRCA2* c.582G>A mutation do not share a full haplotype. However, the small number of patients analysed and the minimal variation observed at the SNP regions do not conclusively provide evidence for or against the presence of a founder effect. The same comparison was performed for the c.5771_5774delTTCA and the results are depicted in Table 3-5 below.

Table 3-5: The haplotypes constructed by the PHASE software for patients with the BRCA2 c.5771_5774delTTCA mutation

Mutation	Patient	Possible haplotypes
c.5771_5774delTTCA	3	10 A A A 21 10 A A A 21
	4	12 A A A 18 12 A A A 21
	5	11 A A A 17 10 A A A 21

Patients 3 and 5 share a common 10 AAA 21 potential haplotype. Patient 4, however, shows a different haplotype. The monomorphic SNPs in the haplotypes do not provide enough evidence to segregate potential disease-susceptibility haplotypes from normal haplotypes.

Subsequent to comparing haplotypes to individuals carrying the same mutation, the mutation carrier samples and control samples were compared to each other to determine haplotypes appearing in both groups and how frequently they occur. Five common haplotypes between the mutation-positive and control samples were observed and are presented in Table 3-6, along with the number of alleles in which they appeared in the control samples, the c.5771-5774delTTCA and c.582G>A mutation-positive samples.

Table 3-6: Shared haplotypes observed in either the *BRCA2* c.5771-5774delTTCA or c.582G>A mutation-positive individuals and control samples.

Haplotype	c.5771-5774delTTCA mutation-positive	c.582G>A mutation-positive	Number of alleles (controls)
10 AAA 21	3	0	1
11 AAA 17	1	0	17
12 AAA 18	1	0	1
12 AAA 19	0	1	1
12 AAA 21	1	0	8

The haplotype 10 AAA 21 appeared more frequently in the c.5771_5774delTTCA mutation individuals than in the controls, potentially segregating with disease but larger studies would be required to be more definitive. The haplotypes 11 AAA 17 and 12 AAA 21 appeared frequently in the control alleles and may be indicative of haplotypes common in the normal population and not segregating with disease. The haplotypes 12 AAA 18 and 12 AAA 19, however, are less informative about disease-segregation as they each appeared once in the mutation-carrier group and the control group.

Biases were expected in the interpretation of haplotypes and allelic frequencies as the control group was larger in size than the mutation-carrier group. The haplotype analysis failed to reveal a founder haplotype for either of the *BRCA2* mutations. The accuracy of the results was further skewed by the incorrect SNPs under study and the small sample size of the mutation-positive individuals.

3.3 Mutation screen of the *FANCG* c.637_643delTACCGCC mutation

Following the haplotype analyses, the *FANCG* seven base-pair c. 637_643delTACCGCC mutation was investigated because FA gene mutations in heterozygous form have been shown to contribute to the IBC phenotype. The *FANCG* mutation is a founder mutation in the black South African population which, when in its homozygous or compound

heterozygous state with another *FANCG* mutation, gives rise to FA (Morgan *et al*., 2005; Wainstein *et al*., 2013). As such, the possibility of its contribution to the IBC was investigated when it is in its heterozygous state. A sample of 119 black and coloured patients suspected to be at increased risk for IBC were screened for the mutation, using PCR amplification followed by capillary electrophoresis. The wild-type allele obtained from the genotyping was 102 bp but when the mutation was present, the mutant allele was 95 bp as illustrated in Figure 3-1.

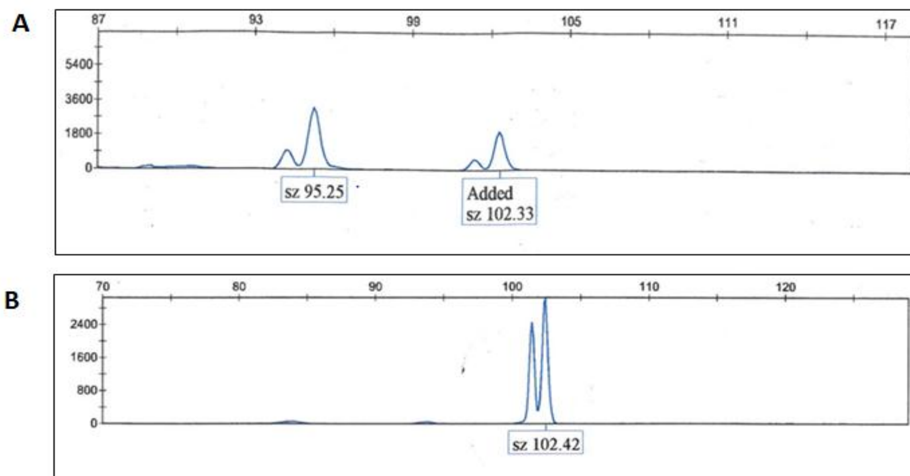


Figure 3-1: Capillary electropherograms illustrating the presence and absence of the *FANCG* c.637_643delTACCGCC mutation.
A) A heterozygous individual with the wild-type allele, (102 bp) and the mutant allele (95 bp) resulting from the seven base pair deletion. B) A homozygous individual for the wild-type allele. For a person suspected of having an inherited breast cancer syndrome, it would be expected of them to have the mutation in its heterozygous state as in A should the disease be associated with the presence of the *FANCG* mutation.

None of the 119 individuals screened for the *FANCG* gene mutation tested positive. It is unlikely that the mutation has a significant role in breast cancer development or that a larger sample size needs to be tested to provide statistical evidence.

3.4 Next generation sequencing

For the final part of the study, a next generation sequencing approach was used to examine patients thought to be at increased risk of an IBC syndrome for pathogenic

mutations in 26 inherited cancer susceptibility genes. All patients (n=24) selected for targeted sequencing were black South African females, diagnosed with breast cancer at the age of 50 years or younger. A number of these patients (27.8%) have previously been tested and found not to have mutations in *BRCA1* or *BRCA2* but were included here to enrich results for other cancer susceptibility genes. Twenty-six cancer susceptibility genes were screened for as shown in Section 2.6.2. This section will be divided into the quality control results determined during library preparation and the results determined following sequencing on the Illumina MiSeq instrument.

3.4.1 Library preparation

Library preparation was performed by rounds of amplification to target the regions of interests, attach adaptors and MIDs to the various samples.

3.4.1.1 Quality control to confirm the annealing of primers to target genes

The first step in the preparation of libraries was to perform a multiplex PCR that allowed all primers targeting the 26 genes of interest to anneal to the DNA samples. This was followed by purification and then gel electrophoresis to determine whether the primers bound and if amplification was successful as shown in Figure 2-5. All 24 samples had amplicons of approximately 450 bp in size, indicating successful amplification and purification.

3.4.1.2 PCR for the annealing of adaptors and molecular identifiers

PCR was performed to anneal adaptors and MIDs to the DNA samples. The MIDs were annealed to ensure that samples can be distinguished from each other after sequencing. A quality control step was performed using the Agilent Bioanalyzer to determine the fragment size of each amplicon library. Ten of the 24 libraries were randomly selected for

quality control and the results are shown in Figures 3-2, 3-3 and 3-4. The Bioanalyzer traces were expected to show peaks between 350 bp and 550 bp. An internal control was included to determine whether the instrument was detecting signal.

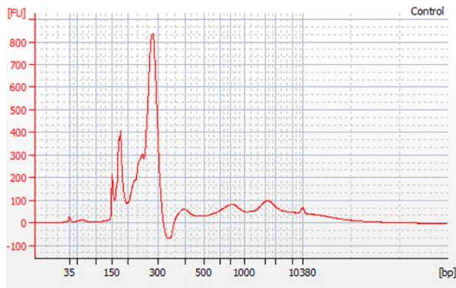


Figure 3-2: Image of a Bioanalyzer trace indicating the the internal control.
An internal control was included to ensure the accuracy and function of the instrument. The internal control fragment was sized at approximately 250 bp as expected.

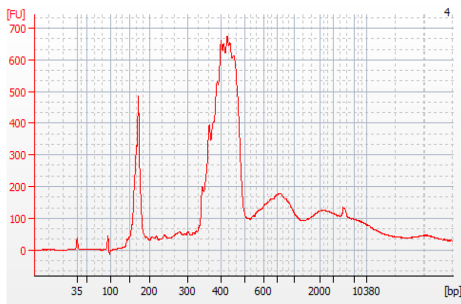


Figure 3-3: Example image of a Bioanalyzer trace for a successful library.
Six of the ten selected libraries produced the expected signal between 350 bp and 550 bp indicating successful annealing of adaptors and molecular identifiers. The signal observed at approximately 160 bp is indicative of adaptor dimers present in the sample.

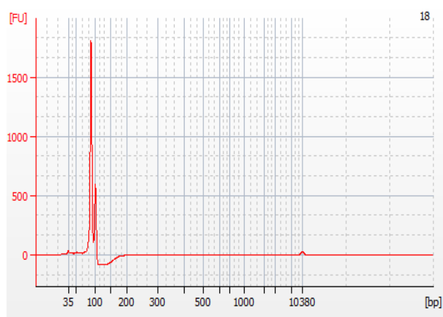


Figure 3-4: Example image of a Bioanalyzer trace for an unsuccessful library.
Four of the ten selected libraries were unsuccessful with no signal between 350 bp and 550 bp. The signal observed at 100 bp is indicative of primer dimers present in the sample.

3.4.1.3 Pooling of the purified tagged amplicons

Following Bioanalyzer analysis and molarity calculations (Section 2.6.4.2, page_), six libraries found to have a molarity under 7nM (four of these had no amplification as indicated in Figure 3-4), were excluded from further analyses. This meant that only 18 of the 24 libraries prepared could be taken forward for sequencing.

Each of the libraries was diluted to a concentration of 4nM. The libraries were pooled, denatured and diluted further to a concentration of 20pM as detailed in Section 2.6.4.2.2. The concentration of the pooled libraries was confirmed on the Bioanalyzer. The pooled libraries were loaded on the reaction cartridge and onto the Illumina MiSeq instrument for sequencing. The results obtained from the instrument are fully presented in Section 3.5.

3.5 MiSeq sequencing results for the Multiplicom BRCA Hereditary Cancer MASTR™ Plus kit

During sequencing, the quality metric was obtained as outlined in Appendix J. The quality metric shows that a phred quality score (Q30) of 88.4% was achieved by the platform. A cluster density of 1232K/mm^2 was obtained and 85.7% of the clusters are shown to have passed filters. These metrics indicate that the sequencing run was of good overall quality.

Subsequent to sequencing, the FASTQ files generated from the MiSeq instrument were aligned to human reference genome (GR37/19 build) using the MASTR Reporter data analysis software and variant calling files were obtained and variants were subsequently annotated.

A total of 657 variants were identified across the 26 genes investigated in the 18 patient samples. A total of 411 variants were located in the intronic regions, making up approximately 62.5% of all variants. The variants in the coding regions made up approximately 27.5% of all identified variants and 96.7% of them were single nucleotide variants (SNV). A summary of the total number of variants identified per sample library is outlined in Table 3-7. Subsequent to functional annotation of the variants, they were categorised as either benign, likely benign, variants of uncertain significance, likely pathogenic or pathogenic (see Table 2-10) based on evidence outlined in available literature and the prediction scores obtained from the *in silico* prediction tools.

Table 3-7: The total number of variants observed in each gene, for each of the 18 samples sequenced

	ATM	*BARD1*	*BLM*	*BRCA1*	*BRCA2*	*BRIP1*	*CDH1*	*CHEK2*	*EPCAM*	*FAM175A*	*MEN1*	*MLH1*	*MRE11A*	*MSH2*	*MSH6*	*MUTYH*	*NBN*	*PALB2*	*PSM2*	*PTEN*	*RAD50*	*RAD51C*	*RAD51D*	*STK11*	*TP53*	*XRCC2*	*Other*	TOTAL
Seq001	33	5	5	18	11	10	5	0	1	7	3	6	4	3	15	3	6	3	5	1	10	1	0	1	3	0	10	**169**
Seq002	31	7	6	18	8	10	4	0	1	6	2	5	1	6	10	2	13	9	22	3	5	1	1	6	2	1	3	**183**
Seq003	39	6	7	15	9	6	4	0	1	7	1	5	6	6	10	4	7	9	5	5	6	0	0	2	3	1	4	**168**
Seq004	41	9	3	8	11	7	3	0	4	5	2	3	8	7	11	3	11	4	28	4	7	1	1	6	4	1	6	**198**
Seq005	33	10	4	15	11	10	4	0	4	6	2	3	11	12	14	4	3	9	6	3	8	1	0	1	2	1	8	**185**
Seq006	34	8	12	12	10	6	2	1	3	8	4	1	7	6	11	3	14	9	6	3	5	0	1	7	0	0	9	**182**
Seq007	24	8	6	10	15	8	4	0	2	8	3	9	7	7	12	4	4	4	4	2	9	2	1	6	6	1	6	**172**
Seq008	28	10	8	18	11	8	3	1	2	6	2	6	10	5	11	3	12	7	4	2	9	1	0	7	3	1	7	**185**
Seq009	26	6	9	10	12	15	5	0	2	6	2	3	6	7	12	4	10	8	6	4	9	0	0	5	6	1	3	**177**
Seq010	29	4	5	12	12	10	7	0	3	8	2	2	7	11	10	2	7	9	6	3	6	2	0	5	2	0	4	**168**
Seq011	31	6	10	10	15	10	7	0	3	5	2	7	5	5	15	4	10	3	22	2	7	0	1	6	3	0	7	**196**
Seq012	32	11	6	10	12	7	3	0	2	3	2	5	1	7	10	4	7	5	9	4	10	1	0	6	3	0	8	**168**
Seq013	32	10	11	14	9	9	4	0	2	5	2	8	7	10	14	4	4	9	3	2	6	2	1	6	5	2	7	**188**
Seq014	28	6	7	11	12	9	8	0	2	5	2	4	7	5	8	2	2	8	4	4	7	2	1	4	3	1	5	**157**
Seq015	28	10	5	13	12	13	7	0	2	8	4	2	7	6	12	3	11	6	7	3	6	1	1	2	4	1	6	**180**
Seq016	30	10	6	21	10	7	7	1	2	4	2	6	6	3	11	3	14	8	22	3	10	1	0	6	3	3	6	**205**
Seq017	29	2	2	19	10	9	3	0	2	7	3	2	5	7	10	3	9	6	25	5	9	2	1	5	6	2	12	**195**
Seq018	25	10	4	8	9	10	4	1	2	4	2	5	8	9	10	2	13	9	8	1	6	1	1	6	2	1	10	**171**

*The column "Other" refers to variants that were located in intergenic and non-coding RNA regions

3.5.1 Previously described Class I and Class II Variants

A total of 248 previously reported as benign or likely benign variants were identified and are listed in Appendix K. The variants have been previously reported on ClinVar as either benign (class I) or likely benign (class II). The *TP53* variant, c.215C>T, has not been shown to confer risk to the development of breast cancer but it has been reported to affect the efficacy of drugs. Two intronic variants, c.214-44G>A in *RAD50* and c.8755-66T>C in *BRCA2* are both reported as benign and *in silico* prediction tools indicate that they do not alter the splice-site and have no damaging effect on the protein. Three of the benign variants have merged rs numbers in dbSNP. The variants with merged ID's are: c.3476-12_3476-10delTTC in *RAD50* was previously rs147883712 and is now rs730881922, c.165-13_165-10delGTTT in *PTEN* was rs531173559 and is now rs786204877, and c.706-5_706-4delTT in *PMS2* was rs531184698 and is now rs776641246. The benign variants have no consequence on phenotype and thus, do not predispose individuals to IBC. MAF for the variants were extracted from 1000 Genomes and ExAC databases.

3.5.2 Previously described Class IV and V Variants

Three previously reported pathogenic variants which led to a diagnosis of IBC were identified in three unique patients (16.67%) in this study and are listed in Table 3-8. The variants had not been reported in 1000 Genomes nor the ExAC database. Two of the three mutations were identified as truncating, nonsense mutations while the one was determined to be a missense mutation.

Table 3-8: Three previously reported pathogenic mutations in known inherited cancer susceptibility genes which were identified in this study using a next generation sequencing approach in three different individuals

Gene	HGVS DNA nomenclature	HGVS Protein nomenclature	Mutation type	dbSNP	References
ATM	c.1369C>T	p.Arg457Ter	Nonsense	rs749036865	van Os *et al.*, 2017
BARD1	c.3G>A	p.Met1Ile	Missense	rs587780031	N/A
RAD50	c.412C>T	p.Arg138Ter	Nonsense	rs786203485	Aloraifi *et al.*, 2015

The *ATM* mutation, c.1369C>T (p.Arg457Ter), is listed as pathogenic in the ClinVar database and has been previously identified in patients with Ataxia Telangiectasia (Hoche *et al*., 2014; van Os *et al*., 2017). The mutation results in a premature transcription stop codon, leading to a truncated mRNA transcript which may be degraded by nonsense-mediated mRNA decay or the transcript may produce a dysfunctional protein.

The *BARD1* mutation, c.3G>A (p.Met1Ile), is reported on ClinVar as pathogenic but no supporting citations are available. This mutation interrupts the transcription initiation codon by changing it from a Methionine to an Isoleucine amino acid. It is expected that a protein will not be produced as the translation signal (Methionine) is eliminated. This c.3G>A change has been seen to be pathogenic in genes such as *RAD50* and *BRCA2* in breast cancer patients (Konecny *et al*., 2011; Foley *et al*., 2015).

The *RAD50* mutation, c.412C>T (p.Arg138Ter), was identified as a novel mutation in a breast cancer patient who had previously tested negative for *BRCA1* and BRCA2 (Aloraifi *et al*., 2015). The mutation results in a truncated mRNA transcript and thus, predicted a dysfunctional protein (Aloraifi *et al*., 2015).

3.5.3 Class III Variants

Variants of uncertain significance (VUS) are variants whose protein effects, function and association to disease are unknown. To determine how frequent the VUS appear in the population, minor allele frequencies (MAFs) from 1000 Genomes and ExAC were used. The frequencies extracted were those observed in all populations and those solely in individuals of African descent (AFR). *In silico* prediction tools mentioned in Section 2.6.4 were used to predict the possible pathogenicity of the variants identified. To differentiate the various VUSs identified in the study, the following sub-classifications for the variants were implemented and are outlined in Table 3-9.

Table 3-9: The sub-classification of the various variants of uncertain significance identified in this study

Section	VUS Sub-Classification	Description	Number of Variants
3.5.3.1	True VUS	All reports in the literature suggest these are true class III variants.	16
3.5.3.2	Variants with conflicting interpretations	Variants with different classifications by different groups of individuals	21
3.5.3.3	Novel variants	Variants appearing for the first time in this study	95
3.5.3.4	ExAC reported variants	Variants have been previously reported on the ExAC database but have not been classified	49
3.5.3.5	dbSNP and 1000 Genomes reported variants	Variants have been previously reported in dbSNP and 1000 Genome databases but no there is no literature suggesting the classification of the variants	127
3.5.3.6	ExAC, dbSNP and 1000 Genomes reported variants	Variants have been previously reported in ExAC, dbSNP and 1000 Genome databases but no classification of the variants is available	98
Total			**406**

3.5.3.1 True variants of uncertain significance

A total of 16 variants which have been previously reported as true class III variants were identified in this study and are outlined in Table 3-10 below.

Table 3-10: The sixteen variants of uncertain significance that were identified in this study in inherited cancer susceptibility genes and evidence in support of their pathogenicity in the form of minor allele frequencies and *in silico* modelling outputs

Gene	DNA nomenclature	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Frequency in study	SIFT	Polyphen	CADD	MutationTaster	VEP
ATM	c.334G>A	p.Ala112Thr	Missense	rs146382972	0.0004	0.0015	0.0002	0.0025	1/18	Deleterious	Possibly damaging	27.6	Disease-causing	Moderate
	c.320G>A	p.Cys107Tyr	Missense	rs142358238	0.0004	0.0015	0.0001	0.0016	1/18	Deleterious	Benign	11.6	Disease-causing	Moderate
	c.4792C>A	p.Leu1598Ile	Missense	rs375190373	-	-	8.242E-06	0.00009625	1/18	Tolerated	Probably damaging	24	Disease-causing	Moderate
	c.2251-19T>C	-	Intronic	rs370713089	0.0002	0.0008	0.00007559	0.0009	1/18	N/A	N/A	1.3	Polymorphism	Modifier
BRCA1	c.1724A>G	p.Glu575Gly	Missense	rs111539978	-	-	8.256E-06	0.00009679	4/18	Tolerated	Benign	2.4	Polymorphism	Low
BRCA2	c.5414A>G	p.Asn1805Ser	Missense	rs80358765	-	-	0.00004129	0.0003	1/18	Tolerated	Possibly damaging	4.7	Polymorphism	Low
	c.8092G>A	p.Ala2698Thr	Missense	rs80359052	-	-	0.00003296	0.0002	1/18	Tolerated	Benign	0.02	Polymorphism	Low
	c.6842-73T>A	-	Intronic	rs11571673	0.0016	0.0061	-	-	1/18	Tolerated	Benign	3.2	Polymorphism	Low
FAM175A	c.1042A>G	p.Ala348Thr	Missense	rs12642536	0.42	0.12	0.4308	0.1523	2/18	Deleterious	Possibly damaging	17	Polymorphism	Moderate
MSH2	c.942+20_942+29delAAAAAAAAAA	-	Intronic	rs587779196	-	-	-	-	1/18	N/A	N/A	0.9	Polymorphism	Modifier
MSH6	c.457+52_457+53dupTG	-	Intronic	rs374618432			0.0003	.	2/18	N/A	N/A	2.8	Polymorphism	Modifier
	c.3557-4delT	-	Intronic	rs267608102	-	-	0.1386	0.2299	16/18	N/A	N/A	0.2	Polymorphism	Modifier
PMS2	c.2350G>A	p.Asp784Asn	Missense	rs143340522	0.0044	0.0091	0.0014	0.008	1/18	Tolerated	Probably damaging	31	Disease-causing	Moderate

Gene	DNA nomenclature	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Frequency in study	SIFT	Polyphen	CADD	MutationTaster	VEP
	c.320G>A	p.Arg107Gln	Missense	rs63751284	-	-	0.0000616	0	5/18	Deleterious	Probably damaging	37	Disease-causing	Moderate
	c.1379G>A	p.Gly460Asp	Missense	rs150201462	-	-	0.0000659	0.0000961	1/18	Tolerated	Benign	5.6	Polymorphism	Low
PTEN	c.882T>G	p.Ser294Arg	Missense	rs143335584	-	-	0.0000617	0.0006	1/18	Deleterious	Benign	15.25	Disease-causing	Moderate

Of the VUS detected in the targeted gene panel testing, six (highlighted in red) were predicted to be pathogenic by at least three of the five prediction tools used. Although not conclusive on its own, this does indicate that these variants may require further investigation to confirm their pathogenicity. The *MSH6* c.3557-4delT variant was previously observed in a Pakistani population at a MAF of 0.025. The variant was classified as a VUS as there was insufficient evidence to classify it as either benign or pathogenic (Hedge *et al*., 2005). The variants in Table 3-10 could possibly be common in the population and not conferring any predisposition to IBC as they appeared at high frequencies in this study.

3.5.3.2 Variants with multiple interpretations

These variants are a class of VUS that have been reported on NCBI's ClinVar database as having "conflicting interpretations of pathogenicity". These variants were studied by a consortium but the groups within the consortium have different interpretations on the effect of the variant on the protein. The variants, therefore, have multiple unconfirmed classifications. Twenty such variants were identified in the study and are identified in Figure 3-11.

Table 3-11: The twenty-one variants which have been classified as having conflicting interpretations of pathogenicity were identified in this study. ExAC and 1000 Genomes databases were used to determine the frequencies at which these variants appear in all populations and only in African populations. Five prediction tools SIFT,Polyphen, CADD, MutationTaster and Variant Effect Predictor (VEP) were used to estimate pathogenicity of each variant

Gene	DNA nomenclature	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Frequency in study	SIFT	Polyphen	CADD	MutationTaster	VEP
ATM	c.2096A>G	p.Glu699Gly	Missense	rs147934285	0.0018	0.0068	0.0002	0.0021	1/18	Deleterious	Possibly damaging	21.5	Disease-causing	Moderate
ATM	c.3118A>G	p.Met1040Val	Missense	rs3092857	0.014	0.053	0.0037	0.0414	2/18	Tolerated	Benign	0.19	Polymorphism	Low
BARD1	c.1518_1519delTGinsCT	p.Val507Leu	Non-frameshift substitution	rs386654966	-	-	-	-	9/18	N/A	N/A	8.4	Disease-causing	Modifier
BLM	c.1601A>G	p.Asn534Ser	Missense	rs35224686	0.0018	0.0068	0.0004	0.004	1/18	Tolerated	Benign	9.6	Polymorphism	Moderate
BLM	c.254G>C	p.Arg85Thr	Missense	rs141503266	0.0028	0.011	0.0006	0.0059	1/18	Tolerated	Benign	8.922	Polymorphism	Low
BLM	c.4077-10C>T	-	Intronic	rs145310008	0.0012	0.0045	0.0007	0.0073	1/18	N/A	N/A	0.19	Polymorphism	Modifier
BRCA2	c.1011C>T	p.Asn337Asn	Silent	rs41293473	0.0002	0.0008	3.33E-05	0.0002	1/18	Tolerated	Benign	2.9	Disease-causing	Low
BRCA2	c.6412G>T	p.Val2138Phe	Missense	rs11571659	0.0016	0.0061	0.0008	0.0098	1/18	Tolerated	Benign	2.6	Polymorphism	Low
BRCA2	c.593+3G>A	-	Intronic	rs80358013	0.0002	0.0008	3.30E-05	0.0004	1/18	N/A	N/A	0.04	Polymorphism	Modifier
BRCA2	g.5177C>G	c.517-4C>G	Intronic	rs81002804	-	-	0.00002481	0.0003	1/18	N/A	N/A	1.9	Polymorphism	Modifier
BRCA2	c.9648+54G>A	-	Intronic	rs11571823	0.0018	0.0068	-	-	1/18	N/A	N/A	2.1	Polymorphism	Modifier

Gene	DNA nomenclature	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Frequency in study	SIFT	Polyphen	CADD	MutationTaster	VEP
BRIP1	c.93+15G>A	-	Intronic	rs113052745	-	-	4.34E-05	0.0005	1/18	N/A	N/A	0.5	Polymorphism	Modifier
CDH1	c.1849G>A	p.Ala617Thr	Missense	rs33935154	0.014	0.051	0.0042	0.0462	1/18	Tolerated	Benign	10.5	Disease-causing	Moderate
MEN1	c.1621G>A	p.Ala541Thr	Missense	rs2959656	0.83	0.67	0.9378	0.7079	16/18	Tolerated	Benign	0.01	Polymorphism	Moderate
MEN1	c.655-6C>T	-	Intronic	rs77461664	0.006	0.021	0.0016	0.0161	3/18	N/A	N/A	0.07	Polymorphism	Modifier
MLH1	c.1732-19T>A	-	Intronic	rs77120160	0.0096	0.034	0.0028	0.0309	1/18	N/A	N/A	0.02	Polymorphism	Modifier
MSH6	c.4002-11_4002-10delTT	-	Intronic	rs587779306	-	-	0.1323	0.1439	9/18	N/A	N/A	0.04	Polymorphism	Modifier
MUTYH	c.158-12C>T	-	Intronic	rs367908623	-	-	0.00007356	0.0009	1/18	N/A	N/A	0.1	Polymorphism	Modifier
PALB2	c.629C>T	p.Pro210Leu	Missense	rs57605939	0.025	0.091	0.0061	0.0676	2/18	Tolerated	Benign	10.57	Polymorphism	Moderate
PMS2	c.1280G>A	p.Arg427His	Missense	rs112902065	0.0008	0.003	0.00004981	0.0006	2/18	Tolerated	Benign	10.57	Polymorphism	Moderate
RAD51D	c.983C>T	p.Thr328Ile	Missense	rs138969595	0.0008	0.003	0.00009885	0.0012	2/18	Deleterious	Benign	8.356	Polymorphism	Low

One of the variants in this category was classified by at least three predictions tools as pathogenic. This variant would therefore require additional evidence to confirm its pathogenicity. The silent variant, c.1011C>T in *BRCA2*, may not cause significant changes to the protein as the amino acid is retained at that position and the prediction tools do not deem it to be pathogenic. However, it is possible that the point mutation may alter the way in which anti-codons recognise this specific codon, which could lead to aberrations in mRNA translation. The role of this variant in translation can be analysed further by functional analysis to confirm the variant as not pathogenic.

The *MEN1* variant, c.1621G>A and the MSH6 c.4002-11_4002-10delTT variant were seen at high frequencies in this study and may possibly common variants rather than disease-causing. The *ATM* variant, c. 2096C>T which is highlighted in red, is the only variant that was predicted by at least three prediction tool as possibly deleterious. The variant changes glutamic acid to glycine which leads to changes in the composition of the side chain of the amino acid, which may ultimately affect protein folding and structure and possibly function. Further studies can be conducted in the future to determine the effect of variant on protein function.

The *BARD1* c.1518_1519delTGinsCT was observed at a high frequencyin this study. It has been reported on ClinVar with conflicting interpretations of pathogenicity.The variant replaces two adjacent nucleotides which lead to a missense variant in both positions. The missense variants, c.1518T>C and c.1519G>T, were observed to have frequencies of 0.8119 and 0.000424, respectively in an East Asian subpopulation. The c.1519G>T variant frequency is twice the estimated maximal expected allele frequency for a pathogenic *BARD1* variant (0.0002188). Both frequencies observed in the East Asian population may suggest that the c.1518_1519delTGinsCT variant is a benign polymorphism but further investigations are warranted to correctly classify the variant.

3.5.3.3 Novel variants

Another class of variants identified in this study are those categorised as novel. These variants have not been previously reported in the literature. A total of 95 novel variants were identified in this study. Thirteen of the 95 variants were predicted to be disease-causing by at least three of the *in silico* prediction tools used and they are highlighted in Table 3-12 below. The remaining 82 variants are listed in Appendix L and may possibly be benign but more confirmations will be required.

Table 3-12: Novel sequence variants predicted to be possibly disease causing by three or more in *silico* prediction tools.
A total of 95 novel variants were identified in this study. The 13 variants highlighted were predicted as possibly disease-causing variants.

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	Frequency in study	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.281T>A	108100000	*ATM*	p.Met94Lys	Missense	5/18	Tolerated	Probably damaging	15.42	Disease-causing	Moderate
c.286G>A	108100005	*ATM*	p.Glu96Lys	Missense	1/18	Tolerated	Probably damaging	37	Disease-causing	Moderate
c.290T>A	108100009	*ATM*	p.Ile97Asn	Missense	6/18	Deleterious	Probably damaging	23.5	Disease-causing	Moderate
c.302T>A	108100021	*ATM*	p.Val101Asp	Missense	11/18	Deleterious	Probably damaging	27.6	Disease-causing	Moderate
c.307T>A	108100026	*ATM*	p.Tyr103Asn	Missense	12/18	Tolerated	Possibly damaging	16.71	Disease-causing	Moderate
c.314T>A	108100033	*ATM*	p.Ile105Asn	Missense	17/18	Deleterious	Probably damaging	25.8	Disease-causing	Moderate
c.323C>A	108100042	*ATM*	p.Ala108Glu	Missense	15/18	Deleterious	Probably damaging	34	Disease-causing	Moderate
c.327C>A	108100046	*ATM*	p.Asn109Lys	Missense	16/18	Deleterious	Probably damaging	17.69	Disease-causing	Moderate
c.1887delC	94170382	*MRE11A*	-	Frameshift	1/18	N/A	N/A	35	Disease-causing	High
c.2478G>C	6013141	*PMS2*	p.Glu826Asp	Missense	3/18	Tolerated	Probably damaging	24.6	Disease-causing	Moderate
c.802-2delAinsTT	89720650	*PTEN*	-	Splicing	1/18	N/A	N/A	24	Disease-causing	High
c.802-1G>T	89720650	*PTEN*	-	Splicing	1/18	N/A	N/A	24.8	Disease-causing	High
c.802G>T	89720651	*PTEN*	p.Asp268Tyr	Missense	1/18	Deleterious	Probably damaging	23.1	Disease-causing	Moderate

The variants may require further study to determine their pathogenicity and role in breast cancer development. The *ATM* variants, c.281T>C, c. 290T>A, c.302T>A, c.307T>A, c.314T>A, c.323T>A and c.327T>A appear at high frequencies in this study. It is possible that the variants are benign variants and common in the population and thus, do not predispose individuals to inherited cancer syndromes.

3.5.3.4 ExAC only reported variants

A total of 49 variants identified in the have only been previously reported on the ExAC database. Four of the 49 variants were predicted by at least three prediction tools to be disease-causing and are outlined in Table 3-13. The remaining variants which may be likely benign are in Appendix M.

Table 3-13: A total of 49 variants which were only reported on ExAC were identified in this study.
The variants were simulated on five *in silico* prediction tools to determine their pathogenicity. Four of the 49 variants were predicted to be pathogenic by at least three prediction tools

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	Frequency in study	ExAC MAF	ExAC MAF (African)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.1536delA	91304139	*BLM*	-	Frameshift	18/18	4.0×10^{-3}	5.0×10^{-4}	N/A	N/A	17.9	Disease-causing	High
c.269C>G	6043405	*PMS2*	p.Ser90Cys	Missense	5/18	8.7×10^{-6}	0.0	Deleterious	Probably damaging	27.6	Disease-causing	Moderate
c.2156delT	131931452	*RAD50*	-	Frameshift	11/18	1.0×10^{-3}	1.2×10^{-3}	N/A	N/A	31.0	Disease-causing	High
c.3292C>T	131953889	*RAD50*	p.Arg1098Trp	Missense	1/18	1.7×10^{-5}	0.0	Deleterious	Probably damaging	24.7	Polymorphism	Moderate

The *BLM* c.1536delA variant was present in heterozygous form in the 18 individuals who were included in this study. The variant may not be pathogenic in its heterozygous state but its effects when homozygous may need to be investigated. The frameshift variant c.2156delT in the *RAD50* gene is located in a noncoding exon. The variant is, therefore, not expected to produce a deleterious protein but it will need to be further investigated as it may be a modifier of downstream genes.

3.5.3.5 dbSNP and 1000 Genomes reported variants

Another class of variants observed in this study were those which have been previously reported in dbSNP and 1000 Genomes databases. A total of 127 of these variants were identified in the study and their pathogenicity tested against five prediction tools. One hundred and seven variants of the 127 were determined to be intronic and the remainder were in untranslated regions, splice donor and acceptor sites and in between genes. Only one variant was in the exonic region and it was a synonymous variant. A variant in the *ATM* gene, c.6573-42delT, has merged rs numbers from rs398017511 and is now rs11366542. The variants identified are in Appendix N.

3.5.3.6 dbSNP, ExAC and 1000 Genomes reported variants

A total of 98 variants identified in this study have been previously described in the dbSNP database and the the 1000 Genomes and ExAC databases. The pathogenicity associated with each of the identified variants has not been reported. Thus, the five *in silico* tools utilised in this study were used to predict how deleteriousness of the variants and the results are in Appendix O. None of the variants were predicted to be disease-causing by at least three of the tools used, which could potentially mean that the variants do not confer any risk to disease-development. However, this will need to be validated in the future.

The c.20+141G>A variant in the *MRE11A* gene was identified in this study. The variant was found to be located in the splice donor site and it was homozygous. Further analysis will be required to determine whether the variant does not affect transcription or cause exon-skipping which could lead to aberrations in the protein function.

3.6 Identification of variants in African Genome database

Some of the variants identified in Section 3.5 were searched in the African Genome Variation Project (AGVP) database to determine how frequently the variants occur in African populations. Eight variants were selected, including the three pathogenic mutations in Section 3.5.2. The *PMS2* c.269C>G variant was previously only observed in the ExAC database as described in Section 3.5.3.4 (page _). The remaining four variants are novel class III variants which were predicted to be disease-causing by the *in silico* tools used in the study. The variants that were run through the AGVP database are in Table 3-14.

Table 3-14: The eight variants selected and searched in African Genome Variation Project (AGVP) databases to determine their frequencies in African population/s.
Three of the mutations (highlighted in red) have been previously reported as pathogenic. The remaining five variants were novel and predicted to be deleterious.

HGVS nomenclature	Gene	Protein nomenclature
c.1369C>T	*ATM*	p.Arg457Ter
c.286G>A	*ATM*	p.Glu96Lys
c.3G>A	*BARD1*	p.Met1Ile
c.1887delC	*MRE11A*	p.Glu632Hisfs*15
c.2478G>C	*PMS2*	p.Glu826Asp
c.269C>G	*PMS2*	p.Ser90Cys
c.802-2delAinsTT	*PTEN*	N/A
c.412C>T	*RAD50*	p.Arg138Ter

Of the eight variants selected and searched in AGVP, only the *RAD50* c.412C>T nonsense mutation was present in the dataset at a low frequency of 0.001 in 320 samples. The variants in the table are possibly very rare or need to be interrogated in larger datasets.

3.7 Variant verification using Sanger sequencing

A novel splice-site mutation in *PTEN*, c.802-2delAinsTT, was identified in a patient who was also found to carry the *ATM* c.1369C>T nonsense mutation. A similar *PTEN* mutation, c.802-2A>T, has been previously reported as a pathogenic mutation and has the dbSNP ID: rs587782455. The mutation has been associated with a hereditary cancer-predisposing syndrome and it was therefore pertinent to confirm whether this mutation was present. To confirm the presence of the *PTEN* mutation and a second pathogenic mutation in the same individual, Sanger sequencing (studying the forward and reverse sequences) was performed to determine whether the splice-site mutation was present or whether slippage occurred in the poly-T stretch surrounding the mutation. The results from the sequencing are shown in Figures 3-5 and 3-6.

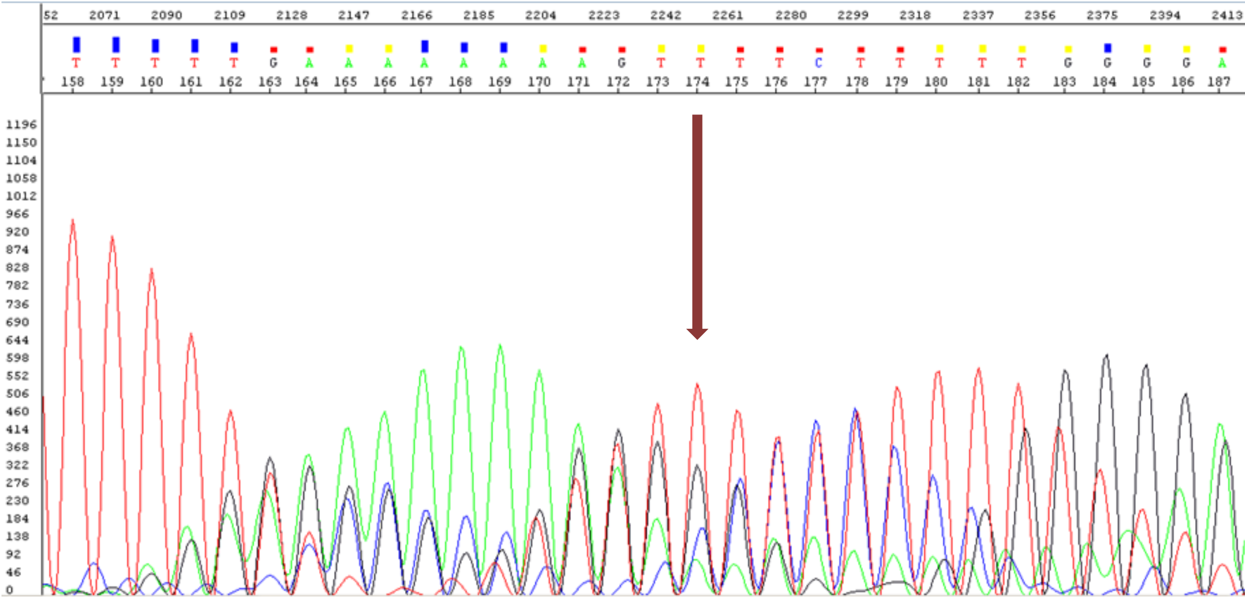


Figure 3-5: A sequencing chromatogram illustrating the forward sequence region of the *PTEN* gene containing the supposed c.802A>TT splice-site variant.
The image indicates a poor sequence region where multiple peaks are present at the site of the mutation (illustrated by arrow) which are possibly due to polymerase slippage

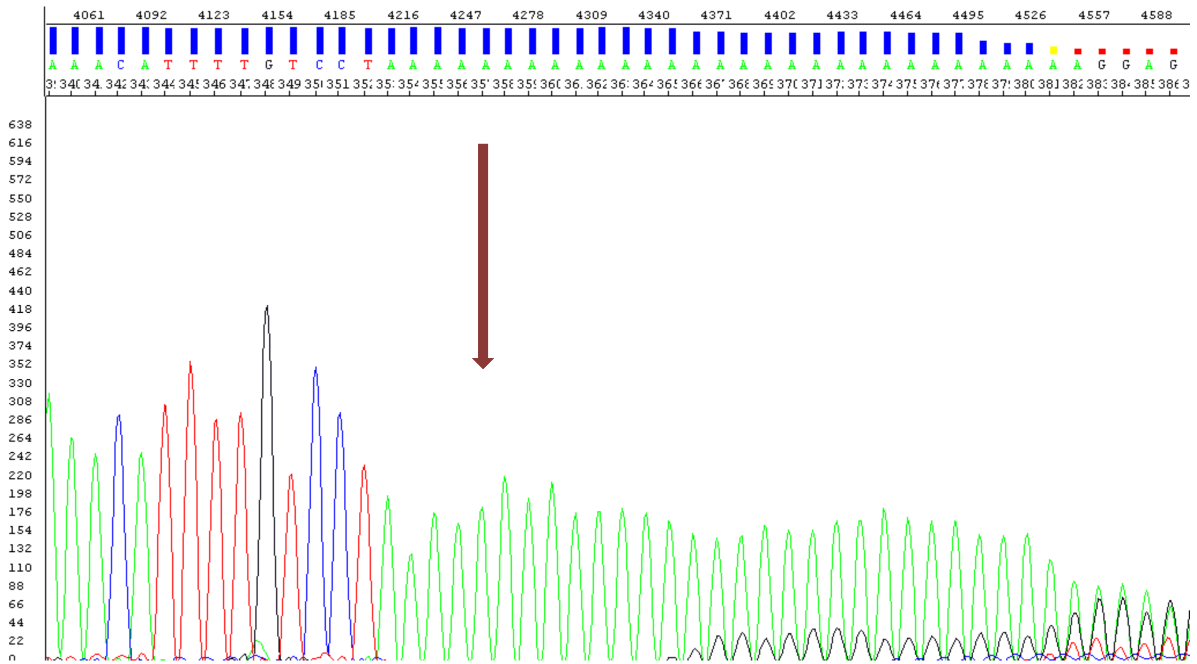


Figure 3-6: A sequencing chromatogram illustrating the reverse sequence region of the *PTEN* gene containing the supposed c.802A>TT splice-site variant.
At the position of the variant change (shown by arrow), a single peak is observed which indicates homozygosity of the "T" allele and the absence of a mutation.

The Figures in 3-5 and 3-6 indicate the forward and reverse sequences which were targeting the region of the c.802-2delAinsTT mutation. The reverse sequence in Figure 3-6 indicates a single peak which can be interpreted as a homozygous, normal genotype. This, therefore, excludes the presence of a mutation at that site in contrast to the forward sequence where the sequence is unclear. Figure 3-5 confirms that slippage of the polymerase occurred during sequencing of the forward sequence which ultimately led to a false-positive result being observed. These images highlight the importance of confirming NGS variants by using Sanger sequencing to ensure accurate results.

4 DISCUSSION

This chapter will discuss the findings from the study and the implication/s thereof. A discussion of the results of all study components will be covered in first part of the chapter. The latter portion of the chapter will discuss the implications of the study on the participant, their families as well as the genetic testing approach for black South African women with breast cancer. The chapter will conclude with a discussion of the limitations of the study and elaborate on ideas for future research prospects.

4.1 *BRCA2* mutation screening

Two mutations (c.582G>A and c.5771_5774delTTCA) in the *BRCA2* gene were screened for in breast cancer patients. The mutations were selected as they had been previously observed in a number of unrelated individuals with breast cancer who had been genetically screened at the Division of Human Genetics, Wits/NHLS. As such, the presence of a founder effect was suspected and the mutations were included in the study to elucidate their frequencies in black South African individuals with breast cancer.

The results from this study revealed only a single additional patient (in a group of 27 black and eight coloured women with breast cancer) who tested positive for the c.5771_5774delTTCA mutation while none tested positive for the c.582G>A mutation. Both mutations are thus relatively rare and this is supported by previous studies conducted in the Division. Studies performed to date, including the current one, in the Division show a cumulative frequency of 4.46% (7 out of 157) individuals who were identified as carriers of the c.5771_5774delTTCA mutation while 3 of 157 (1.91%) were identified as c.582G>A mutation carriers (Chen, 2015; Francies *et al*., 2015; Feben *et al*., 2017). All the individuals identified to carry the mutations in the Division were black and no coloured individuals were identified to carry either of the mutations. The c.5771_5774delTTCA mutation was previously reported in the Western Cape in black

Xhosa and coloured populations (van der Merwe *et al*., 2012). The results from the Western Cape study showed a 6.6% pick up rate for the *BRCA2* c.5771_5774delTTCA mutation; where 4/16 (25%) were black South African Xhosa individuals and 4/105 (3.8%) were coloured individuals. The differences in mutation frequency could be attributed to the studies being undertaken in different geographical regions with this study being performed in Gauteng and the other study in the Western Cape. In the Western Cape, the majority of the black South African habitants are of the Xhosa ethno-linguistic grouping and they account for approximately 25% of the provincial population, while the coloured population accounts for about 49% (Census, 2011). However, the demographics are different in Gauteng, with Xhosa speakers accounting for about 7.3% of the population (Census, 2011). It is possible that this mutation is enriched in the Xhosa group and not necessarily as prevalent in the Gauteng region which is known to be ethno-linguistically and culturally diverse. In addition to this mutation being reported in South African populations, mutation carriers have been previously observed in individuals residing in the Netherlands and the mutation was reported in the Breast Cancer Information Core [BIC] in the period between 1997 and 1998 (Breast Cancer Information Core [BIC], 2017). The coloured population in the Western Cape arose from admixture between Dutch settlers and indigenous South African populations, including those of the Xhosa tribe (Petersen *et al*., 2013). This explains why the mutation is also observed in the coloured population in the Western Cape region.

The *BRCA2* c.582G>A mutation is reported as a pathogenic mutation and was previously added to the BIC database in 1997 (BIC, 2017). In recent years, the mutation has been identified in four individuals of African descent in studies conducted in South Africa (Chen 2015; Francies *et al*., 2015; Feben *et al*., 2017). The mutation has been recently identified in an ovarian cancer patient in the Netherlands (Weren *et al*., 2017), suggesting that the mutation may predispose to hereditary breast and ovarian cancer syndromes as it has been previously seen in breast cancer patients and a patient with ovarian cancer (Chen 2015; Francies *et al*., 2015; Weren *et al*., 2017).

4.2 Haplotyping analyses of the *BRCA2* region

SNP and microsatellite-based haplotypes were constructed to determine whether a founder effect could be determined for the two proposed *BRCA2* mutations. The two microsatellite markers (17xAT upstream of *BRCA2* and 21xTG downstream of *BRCA2*) in this study showed a wider variation in the control group, compared to the mutation-carrier samples. This can be ascribed to the smaller sample size used for the mutation-carriers. The *BRCA2* SNP markers (rs4986860[1] (A/G), rs9590940 (A/C) and rs1801426 (A/G)) genotyped in this study were found to be polymorphic in the control samples but not in the mutation carrier group. The rs1801426 SNP shows a MAF frequency of 0.115 (23/200) in the black South African control population in this study but no homozygotes were observed for the minor allele. It is possible that the homozygote minor allele genotype is rare in this population.

The mutation carriers showed no variation and were monomorphic for all SNPs. This is possibly due to the error in selecting SNPs with MAF less than 0.05 (5%) which are quite rare in the general population. Additionally, the sample size of the mutation-positive samples was significantly low and this also, could have contributed to the lack of variation observed in this group of samples. This, overall, made it difficult to discriminate haplotypes showing disease-association from those found in the normal population. Individuals with the same *BRCA2* mutations were expected to have similar haplotypes but the results were inconclusive as only a small sample size was interrogated and the SNPs showed no variation. Thus, no significant founder haplotypes for IBC could be inferred from the results that were obtained from the study. However, if this section of the study were to be replicated should funds and additional mutation-positive samples become available in the future, the following SNPs could be used for the construction of haplotypes and it is anticipated that the desired results may be obtained; rs1799955 with MAF of 0.28 (28%), rs9634672 with MAF of 0.23 (23%) and rs7334543 with MAF of 0.22

[1]SNP rs4986860 has recently been merged with rs52813470 making it a triallelic SNP with A/C/G variation. The SNP was biallelic at the time when genotyping and haplotyping were performed in the study. Results may be bias and not representative of the C allele.

(22%). These SNPs together with an increased sample size may produce the founder haplotypes required to conclusively confirm the two *BRCA2* mutations as founder mutations for the black population in South Africa.

Founder mutation testing has proven useful in many populations as a starting point for the evaluation of HBOCS. An example is the Afrikaner population in whom three breast cancer founder mutations have been described as: *BRCA1* c.1374delC, *BRCA1* c.2641G>T and *BRCA2* c.7934delG (Reeves et al., 2004; van der Merwe *et al*., 2009). In a recent study, the prevalence of the Afrikaner *BRCA* mutations was determined in a group of patients found to be carrying a deleterious *BRCA* mutation. Of the 18 individuals who tested positive for a *BRCA* mutation, 14 (77.8%) carried one of the founder mutations, with the *BRCA2* c.7934delG being the most prevalent and occurring in 66.7% of the sample size (Seymour *et al*., 2016). Founder mutation screening has shown to be cost-effective for this population. The costs for Afrikaner founder mutation testing are estimated to range between R1600 and R3500 while costs for comprehensive screening for *BRCA1* and *BRCA2* range between R9000 and R25000 (Seymour *et al*., 2016).

Founder mutations are often limited to a population and only present for certain disorders. When such mutations are identified, it allows for cost effective testing to be performed as seen with the Afrikaner population for example (Seymour *et al*., 2016). It would be beneficial to identify founder mutations, if present, for the black South African population as this could be both time saving and cost effective.

The results obtained from the *BRCA2* mutation screen in this study illustrate that the mutations account for a minority of black South African HBOCS cases and thus, first line founder mutation screening may not be beneficial in a diagnostic setting as the positive-mutation rate would be low. The reduced cost of this type of testing does not appear to be justified due to the low yield of mutations and thus, the necessity to continue with full

sequencing and large rearrangement analysis in the *BRCA* genes as well as the potential need to screen other inherited cancer susceptibility genes is still required. Despite this, the concept of founder mutation testing is still a valid approach especially in a resource-limited setting and so additional work to evaluate the presence of other founder mutations in other genes should be considered.

4.3 *FANCG* c.637-643TACCGCC mutation screen

The c.637-643TACCGCC mutation in the *FANCG* gene was shown to account for 80% of mutant alleles in the *FANCG* gene for affected FA individuals in the black Southern African population (Morgan *et al*., 2005). The mutation was included for evaluation in this study to determine its contribution and connection to IBC. Other FA genes have been shown to be associated with an increased susceptibility to breast cancer as the two conditions are caused by genes which function in a common pathway. A total of 119 individuals suspected to have an inherited breast cancer syndrome were screened for the *FANCG* mutation and none of the individuals tested positive for the mutation. Considering that the carrier frequency of the founder *FANCG* mutation in the general population is approximately 1% (Morgan *et al*., 2005), it seems unlikely to contribute significantly to the emergence of IBC in this population.

The findings from this study are consistent with those found in previous studies. In one study, FA genes including *FANCG* were screened in 88 families with an IBC syndrome who had previously tested negative for *BRCA1* and *BRCA2* mutations. No obvious deleterious mutations were detected in *FANCG* in this study (Seal *et al*., 2003). In a study conducted in a Finnish population, FA gene mutations were screened to determine their contribution to inherited breast and prostate cancer syndromes. The *FANCG* c.832insG mutation was identified in one of 1174 breast cancer cases who was diagnosed at the age of 43 years and had no family history of disease. The mutation was not observed in the healthy control samples. It was postulated that although the heterozygous frame-shift mutation

was detected, it did not confer high or moderate risk for breast cancer (Mantere *et al*., 2015). It is possible that *FANCG* mutations do not make a significant contribution to IBC or they confer a very low risk for disease. Nevertheless, only the single founder mutation in the *FANCG* gene was evaluated in this study and so additional research may be warranted to exclude *FANCG* as a significant contributor to IBC in general.

4.4 Next generation sequencing: multi-gene panel testing

Next generation sequencing (NGS) technologies were first introduced in 2005 to overcome the limitations of the Sanger sequencing technology (Margulies *et al*., 2005). Unlike Sanger sequencing, NGS techniques are high-throughput and their applications include sequencing of the entire genome, exome (coding regions) or targeted regions of the genome. NGS technologies make use of massively parallel sequencing (MPS) to decipher the genetic code (Margulies *et al*., 2005). MPS enables multiple genes to be sequenced simultaneously on a large-scale, drastically reducing time and costs associated with sequencing in comparison to Sanger sequencing. The first human genome sequence cost approximately $2.7 billion in 2001. Over the years, the cost has reduced drastically to an estimated $1000 in 2016 (Weymann *et al*., 2017). This, therefore, enables scientists to perform large-scale sequencing at affordable prices.

The advent of NGS technologies has made it possible to study complex traits with less difficulty. Due to the nature of these traits being genetically heterogeneous, single gene testing for the associated genes is costly. NGS can therefore identify all predisposing variants and susceptibility loci associated with a particular complex trait in one test (Winship & Southey, 2016). Furthermore, NGS technologies can be effectively used to identify indels (insertions/deletions), single nucleotide variants (SNVs) and copy number variants (CNVs). Although there are a number of advantages to utilising NGS technologies, there are complications to these technologies; these include the identification of multiple

variants of uncertain significance and incidental findings. These remain a significant challenge for clinicians to report back to patients (Winship & Southey, 2016).

The NGS application utilised in this study was a targeted gene panel. Targeted gene sequencing is used to evaluate only genes that are of interest for a particular trait (Gray *et al*., 2015). This, therefore, increases the speed at which results are obtained and also significantly reduces costs and data associated with sequencing compared to whole-exome or whole-genome sequencing. Additionally, gene panels produce high read depth, approximately 500X, increasing the accuracy of the assay. Novel and rare variants are thus, more confidently called (Gray *et al*., 2015). Given the large quantity of variants as well as the high read depths achieved in this study, as well as the well-defined gene contributors to IBC, the choice of a targeted gene panel for this study was justified. For the purposes of this study, amplicon targeted sequencing was chosen to sequence the 26 cancer susceptibility genes of interest. Targeted sequencing of the 26 cancer susceptibility genes was performed to detect deleterious mutations in additional cancer genes (beyond *BRCA1* and *BRCA2*) that had not been previously investigated in the black South African population. Furthermore, this was performed to potentially increase the mutation detection rate in black South African patients with breast cancer suspected to have a hereditary component. Following sequencing and analysis, the variants were categorised as either benign (class I and class II), pathogenic (class IV and class V) or variants of uncertain significance (class III). The majority of the variants identified were found in the *ATM* gene and this can possibly be attributed to the large genomic size of the gene (approximately 150kb).

4.4.1 Previously described Class I and II variants

A total of 248 previously reported benign or likely benign variants were identified in this study as outlined in Appendix K. These variants do not confer a risk to the development of breast cancer and were reported in ClinVar as either benign or likely benign. Two intronic

variants, c.214-44G>A in *RAD50* and c.8755-66T>C in *BRCA2* are both reported as benign in ClinVar and *in silico* prediction tools show that they do not alter the splice-site and have no damaging effect on the protein. The allele frequencies in African populations in 1000 Genomes and ExAC datasets indicate that the derived allele for the majority of the variants are rarer than those that are ancestral. The benign and likely benign variants identified have been reported to have no consequence on phenotype and thus, do not predispose individuals to IBC. The large number of variants identified as benign or likely benign in this study serves to highlight the great genomic diversity which is observed in African genomes.

4.4.2 Previously described Class IV and V variants

Pathogenic variants are those which show a clear significance for a clinical phenotype and have a 99% minimum likelihood to be pathogenic (Plon *et al*., 2008). Some pathogenic variants are truncating and lead to the production of an incomplete mRNA transcript (Miller & Pearce, 2014). These variants introduce premature stop codons through the incorporation of point mutations and small insertions or deletions. The mRNA can be subjected to the nonsense–mediated mRNA decay pathway where it will be degraded, significantly reducing the levels of protein in the cell and ultimately leading to disease (Miller & Pearce, 2014). Two truncating variants, *ATM* c.1369C>T (p.Arg457Ter) and *RAD50* c.412C>T (p.Arg138Ter) which both result in a premature stop codon were identified in this study in two unrelated patients.

Ataxia Telangiectasia (AT) is a disorder inherited in an autosomal recessive manner and arises due to mutations in the *ATM* gene (van Os *et al*., 2017). The *ATM* c.1369C>T mutation observed in this study was previously identified in a homozygous and compound heterozygous state with another *ATM* gene mutation in patients with AT (Hoche *et al*., 2014; van Os *et al*., 2017). Women found to be heterozygous carriers of truncating variants in *ATM* have been shown to be at high risk for developing breast cancer as the

heterozygous variants tend to act as dominant negatives (Goldgar *et al*., 2011). In addition, the penetrance associated with truncating mutations in *ATM* was estimated to be similar to the penetrance associated with germline mutations in the *BRCA2* gene (Goldgar *et al*., 2011).

In a recent study, the *ATM* gene was investigated to determine its role in breast cancer development in a Moroccan population (Marouf *et al*., 2017). Breast cancer patients (N=163) and controls (N=150) were screened for the *ATM* c.7271T>G and c.1066–6T>G variants which had been previously associated with an increased risk for breast cancer development. Forty of the patients in the Moroccan had previously tested negative for mutations in the *BRCA1* and *BRCA2* genes and thus, further investigation was required. Subsequent to genotyping, the results from the study indicated that none of the patients screened for the *ATM* gene tested positive for the variants (Marouf *et al*., 2017). Whilst the Moroccan study revealed no pathogenic mutations, this study demonstrated one pathogenic mutation in the *ATM* gene. This highlights the importance of undertaking local research to evaluate inherited breast cancer susceptibility genes as mutation profiles between countries within the same continent may not always be similar. It is also possible that *ATM* gene mutations are uncommon in populations of African ancestry as seen in both studies.

The *RAD50* c.412C>T (p.Arg138Ter) mutation identified in one patient in this study, was previously reported as a novel mutation in an Asian breast cancer patient (Guan *et al*., 2014). The Asian study was performed to determine the efficacy of NGS targeted panels for detecting mutations in hereditary cancer syndromes and the study provided proof that NGS panels are indeed efficient in detecting pathogenic mutations (Guan *et al*., 2014). Panel testing was used in a different study to detect pathogenic mutations in the cancer susceptibility genes in breast cancer patients who had previously tested negative for mutations in the *BRCA1* and *BRCA2* genes (Aloraifi *et al*., 2015). The results from the study indicate that approximately 3% of the identified mutations were in the *RAD50* gene

(Aloraifi *et al*., 2015). This provides further evidence that *RAD50* is a key cancer susceptibility gene as well as the importance of utilising gene panels.

The third pathogenic mutation identified in this study was a deleterious missense mutation in the *BARD1* gene. The mutation identified was a c.3G>A which causes a codon change from a methionine to isoleucine at the first codon position on the amino acid chain (p.Met1Ile). This mutation disrupts the start codon of the mRNA transcript and prevents the start of translation due to the interruption of methionine, which is an essential component of translation initiation. In the presence of such a mutation, the ribosome will not initiate the translation process until another methionine is detected downstream of the original initiation codon (Hartl *et al*., 2011). This may lead to a different protein being produced or no protein being produced at all if a methionine is not detected in the amino acid chain. The absence of a protein product can thus lead to a disease phenotype if one normal allele is not sufficient to produce the amount of protein required by the cell (Hartl *et al*., 2011). The c.3G>A (p.Met1Ile) mutation has also been previously reported to be deleterious in breast cancer patients carrying the said mutation in either *RAD50* (Aloraifi *et al*., 2015) or *BRCA2* (Xia *et al*., 2006).

The pathogenic mutations identified in this study will be validated in a diagnostic setting and the results will be given to the patients by a genetic counsellor or clinician. This may provide some explanations for the presentation of breast cancer in young individuals in these families. Predictive testing can also be offered to relatives of the patients to identify at-risk carriers within the families. This will aid in improving surveillance for the development of tumour growth by performing regular mammograms and breast self-examination. When the cancer is detected at its early stages, relatives may opt for preventative surgery such as mastectomies before the cancer progresses.

4.4.3 Class III variants

Variants of uncertain significance (VUS) are variants that have been identified through genetic testing to be different from the reference sequence but their clinical significance to disease is unknown (Plon *et al*., 2008). For genes that have not been as extensively studied as *BRCA1* and *BRCA2* in the context of inherited cancer syndromes, most variants are categorised as uncertain until more information is available to either classify them as benign or pathogenic (Plon *et al*., 2008).

4.4.3.1 Previously described Class III variants

A total of 17 previously reported VUS and 20 variants of conflicting interpretations were identified in this study. Six of the previously reported VUS and one of the conflicting variants were predicted by at least three *in silico predictions* to be potentially deleterious as highlighted in Sections 3.5.3.1 and 3.5.3.2. All the variants predicted as deleterious are missense mutations and thus change the amino acid where the mutation is present which changes the composition of the protein and possibly function. Further analysis on these variants will need to be performed to reclassify their pathogenicity. The *BARD1* c.1518_1519delTGinsCT and *MSH6* c.3557-4delT variants in Section 3.5.3.1 and the *MEN1* c.1621G>A and the *MSH6* c.4002-11_4002-10delTT variants in Section 3.5.3.2 were observed at high frequencies in this study. Based on the frequencies at which they occur here and their pathogenicity predicted as not deleterious by the *in silico* tools, they can be re-classified to likely benign (Class II) variants until enough evidence is provided in future to re-classify them as benign (Class I). The variants are possibly commonly occurring in the population and not associated to IBC.

4.4.3.2 Previously reported variants classified as Class III in this study

The next set of variants classified as Class III variants in this study are those that were previously reported in ExAC and the 1000 Genomes and dbSNP database but not classified as described in Sections 3.5.3.4, 3.5.3.5 and 3.5.3.6.

The *BLM* c.1536delA and *RAD50* c.2156delT variants described in Section 3.5.3.4 were both predicted as deleterious by the *in silico* prediction tools utilised in the study. Based on the frequency of these variants in this study, it is unlikely that they are deleterious despite them being predicted as such. The c.3292C>T variant in *RAD50* was identified in one individual and it was predicted by four of the five tools as deleterious. Further analyses will need to be performed to determine the pathogenicity of the variant. If it is proven that the variant is indeed pathogenic, the results will need to be reported to the patient and their families, which will significantly aid in improving genetic counselling and management of the disease in the family and other unrelated individuals identified to carry the same variant. The c.20+141G>A variant in the *MRE11A* gene was predicted by the Variant Effect Predictor (VEP) *in silico* tool as a high risk variant. The variant was seen to be located in the splice donor site and it was present in its homozygous state in the individual carrying the variant. The likelihood of identifying a homozygous deleterious variant is extremely low, unless the variant was produced as a sequencing artefact or the parents of the individuals are closely related. Further studies are warranted to correctly classify this variant as it could affect the splicing mechanism, leading to aberrations in transcription and subsequently protein production. If the variant is found to be truly deleterious, consanguinity within the family will need to be investigated.

4.4.3.3 Novel Class III variants

A total of 95 novel class III variants were identified in the study and 13 of the variants were predicted as potentially pathogenic by three of the five prediction tools used to

ascertain the pathogenicity of each variant. The c.281T>C, c. 290T>A, c.302T>A, c.307T>A, c.314T>A, c.323T>A and c.327T>A variants in the *ATM* gene appeared at high frequencies in this study and were predicted as deleterious by the majority of *in silico* tools utilised. These are unlikely to be pathogenic and may have been incorrectly predicted by the tools due to the *ATM* gene not being very well characterised as a breast cancer predisposing gene as the *BRCA1* and *BRCA2* genes. *In silico* prediction tools use available information so regions that are less studied may often result in a less reliable prediction (Kerr *et al*., 2017; Ernst *et al*., 2018). Most of the tools have also been shown to have low specificity, which results in an over-prediction of missense variants as deleterious. This often produces a false-positive or false-negative results which can potentially affect clinical management. As such, it is suggested that in a clinical setting, *in silico* tools should not be used as primary evidence for variant classification and must be supported by segregation and/or functional analyses (Kerr *et al*., 2017; Ernst *et al*., 2018).

The c.286G>A variant in the *ATM* gene and the *MRE11A* c.1887delC variant each appeared once in the 18 individuals whose DNA was sequenced. The *MRE11A* variant is a frame-shift mutation and may potentially lead to the production of a truncated protein. Further investigation and evidence is warranted to correctly re-classify these variant. This will assist in providing the probands and their families with an improved, genetic counselling service. The *PTEN* c.802-2delAinsTT, c.802-1G>T and c.802G>T variants were identified in the same individual and can their presence can be attributed to the sequencing error which occurred in the reverse sequence of the gene as described in Section 3.7. The reverse sequence (Figure 3-6) shows a clear region which omits the presence of the mutation compared to the forward sequence (Figure 3-5) indicating an unclear region where the polymerase could have slipped during sequencing. These variants therefore do not require further analyses.

4.4.3.4 Re-classification of Class III variants

African genomes are genetically diverse and understudied; it is therefore unsurprising that VUS appear more frequently in African populations compared to European populations. These variants pose a great challenge for counsellors and clinicians as their significance cannot be inferred from sequence data alone, and it is thus often difficult to interpret and communicate the findings back to the patients (Cheon *et al*., 2014). The feedback provided by the clinician can be based on the tumour histology of the individuals or the pattern of the cancer that is observed in the family but not the variant identified in the sequence. Predictive testing and management options cannot be offered to the relatives of the individual with the VUS and therefore, relatives at risk of developing disease will not be identified (Cheon *et al*., 2014). This section will, therefore, discuss published recommendations on how to classify VUS.

Multiple recommendations have been made to classify a VUS as either benign or pathogenic. One recommendation is that segregation analysis can be performed to provide sufficient evidence to confirm a variant as pathogenic by performing genetic screening on family members affected with cancer (Eccles *et al*., 2015). If the same variant is seen in cancer-affected members in a family and across different families, the VUS can be reclassified. However, in families where there is limited history or no history of a cancer syndrome, segregation analysis will not provide sufficient evidence for variant reclassification and further analyses would need to be performed.

Segregation analysis is taken into account when utilising the multifactorial likelihood classification model to establish pathogenicity (Lindor *et al*., 2012). The model makes use of multiple parameters to estimate whether a VUS has similar characteristics to previously reported pathogenic variants. The model incorporates amino acid biochemical and evolutionary conservation with clinical data of the patient. In addition, it combines the tumour characteristics and segregation of variants in cancer-affected family members to

calculate probabilities as evidence to classify the variant as pathogenic or not (Lindor *et al*., 2012). When high penetrance cancer susceptibility genes such as *TP53 and CDH1* are extensively studied and more cancer predisposing pathogenic variants are identified, the same model could be applied to estimate pathogenicity of VUS. Due to the complex nature of VUS and their interpretation thereof, guidelines were issued in 2015 to aid in interpreting variants.

The American College of Medical Genetics and the Association for Molecular Pathology (ACMG) issued guidelines that can be used for interpreting and classifying variants (Richards *et al*., 2015). To assess the pathogenicity of novel variants, several lines of evidence would need to be evaluated. The first line would be to perform computational analysis using multiple *in silico* prediction tools to determine the impact of the variant on protein function. This would be followed by functional and segregation analyses and correlated to the phenotype of the proband to classify the variant according (Richards *et al*., 2015).

In a low-resource country such as South Africa, functional analyses for VUS classification may be performed in collaboration with other institutions that have the required resources and expertise. This study has revealed a lot of Class III which most could be classified as Class II (likely benign variants) based on their frequencies in this study as well as the predictions from the computational analyses until further evidence is obtained to correctly classify them. This study further highlights the complexity of African genomes and the high number of variants of uncertain significance and novel variants that are often observed in African populations. There is currently paucity of information about African populations and the results here will, therefore, contribute to adding more knowledge and in future, will enable better characterisation of the variants which pose a great challenge to genetic counsellors and clinicians.

4.4.4 Interesting case identified from the gene panel screening

One of the patients (CANC473) was identified to have two pathogenic mutations upon initial data analysis; a truncating mutation (c.1369C>T) in *ATM* and a splice-site mutation (c.802-2delAinsTT) in the *PTEN* gene. The latter mutation had not been previously reported. However, a similar mutation (c.802-2A>T) which is in the *PTEN* gene has been reported in ClinVar as pathogenic. It was therefore expected that the mutation identified in this study may also have a deleterious effect. Upon inspection of the sequence surrounding the region with the apparent mutation, it was noted that a poly-A stretch was present and thus, polymerase slippage could have occurred. Sanger sequencing was then performed to confirm the presence of the variant. The reverse sequence revealed that no mutation was present, while the forward sequence revealed the presence of slippage. This highlighted the importance of validating NGS variants using Sanger sequencing to avoid incorrect reporting of variants to patients, which could ultimately lead to mismanagement or mistreatment of disease.

In previous years, a study revealed an instance where a female with bilateral breast cancer and an adenocarcinoma of the endometrium was found to carry two pathogenic mutations in *TP53* and *CDH1*; both genes which are high-penetrance genes for inherited cancers (Heitzer *et al*., 2013). Upon discovery of these mutations that were detected by an NGS technique, the patient and her family were provided with appropriate genetic counselling services. Due to the different cancers being observed in the family, the mutations identified in the two genes enabled a better explanation of the disorders. This subsequently enabled the proband and her family to receive improved management and surveilliance strategies for the different cancer syndromes (Heitzer *et al*., 2013). Although the family was assisted and the gene mutations were associated with the various cancers in the family, the identification of two pathogenic mutations poses a great challenge for genetic counselling. This can be difficult in advising patients and their relatives of which surveilliance strategies to opt for as some syndromes have overlapping disease-causing genes. Additionally, there may be difficulties in providing families with probabilities of

future children inheriting either one the mutations or even both. The use of NGS applications is appropriate in detecting multiple pathogenic mutations but also brings forth the challenges of interpretation of results to both the counsellors and the families.

4.4.5 Appropriateness of the gene panel in the black population

The gene panel screen was a pilot study to serve as a proof-of-principle for the employment of gene panels in the black South African population for the study of IBC. The utilisation of the gene panel enabled the detection of deleterious mutations outside of *BRCA1* and *BRCA2* as aimed for. The panel detected pathogenic mutations in the *ATM*, *BARD1* and *RAD50* genes in three out of 18 (accounting for 16.7%) patients who were successfully sequenced. In the previous study conducted where the *BRCA1* and *BRCA2* genes were screened, the positive-mutation pick-up rate was 6.06% (Chen, 2015). This further illustrates the necessity to perform multi-gene testing on the black South African population as an additional mutation pick-up rate in other genes was achieved in this study. Additional studies need be undertaken to determine if there are other genes outside of the 26 genes made available on this commercial kit which should be included in diagnostic testing or indeed if any genes among the 26 can be safely omitted due to a lack of results obtained. This would therefore require additional NGS screening using different panels in a larger sample of black South African women with breast cancer.

4.5 Contribution of study to participants and their families

The *BRCA2* mutation screen and haplotype analysis conducted in the study did not yield any significant results which could potentially aid in improving the participants' lives except for the single individual who was found to have the c.5771_5774delTTCA mutation and subsequently provided with counselling. However, the next generation sequencing study was able to identify pathogenic mutations in at least three of the 18 participants. The families of the affected probands can be offered predictive testing, which in turn could aid in early detection of the disease and thus, improved management. Individuals

identified to carry the mutations could perform regular breast examinations and mammograms to detect tumour growth which can be managed in earlier stages should disease arise.

4.6 Limitations of study

The study had a number of limiting factors which are elaborated on below. The most significant limitation of the study was the small sample size of the breast cancer affected individuals. A larger sample size may have been able to detect the two potential founder mutations in the *BRCA2* gene (c.582G>A and c.5771_5774delTTCA), although these have been shown to occur at similar frequencies as with previous studies (Chen, 2015; Francies *et al*., 2015; Feben *et al*., 2017). One objective of the study was to perform haplotyping to determine the presence of a founder effect for the above-mentioned *BRCA2* mutations. Again, due to the small sample size and also the incorrectly selected SNPs that had low MAF, the haplotypes constructed for the potential *BRCA2* mutations showed no significant association to disease as the SNPs were all determined to be monomorphic. A larger sample size and the inclusion of frequently occurring SNPs in the population may make it possible to make significant inferences to the constructed haplotypes and to either include or exclude the presence of a founder effect.

For the sequencing part of study, a small kit which could take up to a maximum number of 24 samples was purchased due to availability of research funding. A larger kit with a larger set of genes and the ability to sequence approximately 50 individuals might have increased the mutation pick-up rate. Additionally, a number of VUS were identified in the study. This makes it challenging to interpret and feedback the results to the patients and their families and as such, further analysis of these variants is warranted.

4.7 Future prospects

This study indicated that a founder effect was not determined for the two *BRCA2* mutations Additional research to identify if founder mutations exist in other cancer susceptibility genes for the black South African population should be considered as they have been shown to be cost-effective in studies conducted in individuals of other ethnic groups. The study highlighted the effectiveness of a targeted gene panel in detecting deleterious mutations. Furthermore, several VUS were identified in the study. These variants need further studies to be conducted in an attempt to correctly classify them as they pose a major challenge for genetic counselling and management. Potential future studies to improve testing, counselling and management for IBC are proposed and outlined below.

4.7.1 Functional and segregation analysis

Variants of uncertain significance and novel variants are postulated to appear at high frequencies in individuals of African ancestry due to the higher genetic diversity observed in African populations (Eccles *et al*., 2015). This was also observed in this study where a large number of VUS and novel variants in different genes were observed. The pathogenicity of these variants cannot easily be determined based on their sequence information alone and insufficient evidence is unavailable to classify these variants correctly. This evidence can be provided by either performing segregation or functional analyses (Richards *et al*., 2015). Screening of a control cohort may also assist to exclude common variants in the population.

A total of 37 previously described class III variants were identified in this study. An additional 369 variants which have been reported in databases such as ExAC, 1000 Genomes and dbSNP were also identified. However, no clinical data for these variants is available. In this study, all the variants were run across five *in silico* tools to predict pathogenicity. However, the results from the tools in some instances were conflicting. This

is consistent with previously published literature where different *in silico* prediction tools provide inconsistent classification for a specific variant. In a recent publication, the pathogenicity of rare missense variants in *BRCA1* and *BRCA2* was investigated in four populations (African, European, Latino and Asian) using different *in silico* tools (Ernst *et al*., 2018). The prediction by tools showed inconsistencies and reduced reliability in classifying variants across all populations included in the study. As a result, the study suggested that *in silico* tools be used in conjunction with other analyses to provide a more accurate interpretation of variant pathogenicity (Ernst *et al*., 2018).

Therefore, the presence of African genome data and understanding more about the genetic aetiology of IBC in African populations may contribute in the *in silico* prediction tools simulating more accurate results and showing consistency of results between different prediction tools. Nevertheless, computational analysis alone is not sufficient to determine pathogenicity and further functional and segregation analyses may need to be performed to accurately classify these variants clinically.

4.7.2 Next generation sequencing

A next generation sequencing approach was used in this study. A larger gene panel such as the BROCA (breast-ovarian cancer) gene panel (Walsh *et al*., 2010; Shirts *et al*., 2016) may be used to improve mutation pick-up rates. The BROCA panel is currently composed of 66 cancer risk genes and can be utilised to detect mutations in individuals suspected to have an inherited predisposition to breast and ovarian cancer syndromes and other related cancers (i.e. pancreatic, colorectal and endometrial). The sequencing assay can detect variants in exons, introns and certain promoter regions. In addition, the assay is capable of detecting large deletions (Walsh *et al*., 2010; Shirts *et al*., 2016). In a recent study, DNA samples of individuals of African descent from Brazil were evaluated using the BROCA panel (Felix *et al*., 2017). In the study, 173 cases and 119 controls were included for screening. Pathogenic mutations were detected in 37 cases (accounting for 21.4%) and in

one control (accounting for 0.84%). Twelve mutations were identified in *BRCA1* and ten in the *BRCA2* gene. Other mutated genes were *ATM, BRCA1/BARD1, BRIP1, FAM175A, FANCM, NBN, PALB2, RAD51C*, *SLX4* and *TP53*. This study shows the importance of using a gene panel consisting of multiple genes to understand the genetic aetiology of breast cancer more completely (Felix *et al*., 2017) as has been shown.

In addition to targeted gene panels, whole-exome sequencing can be performed to examine genetic susceptibility to disease. Although this type of sequencing focuses on the coding regions of the genome, it could be more affordable in the South African context in comparison to whole-genome sequencing. Exome sequencing was utilised to screen for deleterious mutations in 60 females in the Middle East diagnosed with early-onset breast cancer or with a strong evidence of an inherited breast cancer based on family history (Marafie *et al*., 2015). In the study, novel variants were identified in *BRCA1* and *BRCA2*. However, mutations in *BARD1*, *TP53* and mismatch repair genes appeared to play a more significant role in disease predisposition than *BRCA1* and *BRCA2* (Marafie *et al*., 2015). Exome sequencing also has the potential to reveal novel genes which may be pertinent in disease development. Novel genes would be genomic sequences which show no homology to known sequences but possess regulatory elements and a coding sequence as known genes (Klasberg *et al*., 2016). Additionally, these genes could be previously identified genes which have not yet been associated with cancer syndromes. Exome sequencing can be employed in South Africa to understand breast cancer in the black South African population. Additionally, this type of sequencing will be able to detect mutations in the entire coding region and not only genes limited to those in gene panels. The limitation of utilising exome sequencing is that the variants of uncertain significance observed in gene panels will be multiplied in exomes; posing an even greater challenge for interpretation and genetic counselling.

4.7.3 Epigenetic studies

Epigenetic modifications are changes which occur on the DNA and alter gene expression and activity but are not present in the DNA sequence (Holliday, 2006). These modifications include methylation, acetylation, ubiquitination and phosphorylation. Methylation has been observed in the *MLH1* gene in a proportion of individuals diagnosed with colorectal cancer. In recent studies, methylation of the *BRCA1* promoter region has been studied in breast cancer patients and methylation has been proposed to be a mechanism for disease predisposition. The methylation patterns of the *BRCA1* promoter region studied in 255 women who were diagnosed before the age of 40 and tested negative for germline mutations in *BRCA1* (or *ATM, BRCA2, CHEK2, PALB2* and *TP53*). These women were also included in the study as they had tumours with similar morphological features to those associated with *BRCA1* germline mutations. The methylation patterns in peripheral blood were compared to those observed in the tumours. From the study, methylation was detected in the promoter region of peripheral blood and the associated tumours were highly methylated. The authors propose that epigenetic mutations in *BRCA1* may predispose individuals to breast cancer (Wong *et al*., 2011). A similar study can be considered in a South African population to determine whether epigenetic modifications play a significant in disease susceptibility and how frequently these modifications appear in the said genes.

4.7.4 Genome-wide association studies

Genome-wide association studies (GWAS) are studies which examine genetic variants distributed across the genome in multiple individuals to determine whether a variant/s is associated with a particular trait. Approximately 100 genetic variants have been associated with breast cancer susceptibility (Easton *et al*., 2007; Michailidou *et al*., 2013; Michailidou *et al*., 2015). However, the majority of the variants were identified in women of European descent and are not entirely representative for women of African ancestry. In an attempt to determine association between breast cancer susceptibility and genetic variants, a GWAS study was conducted in women of African ancestry; 6 522 breast cancer

cases and 7 643 controls were included. The study included, for example, Nigerian women, African-American women and women in the African diaspora (Feng *et al*., 2017). Fine mapping was performed to examine 74 variants in 72 regions known to increase cancer risk. From the 74 variants, 68 were found to be frequent in African populations with MAFs greater than 0.05. In six of the 72 regions examined, seven markers in known genes were found to associate with breast cancer in this population (Feng *et al*., 2017). Performing GWAS in black South African breast cancer patients will be important in identifying susceptibility variants specific for this population and the study can subsequently be replicated in other African populations. This will better characterise variants that show association with breast cancer in individuals of African ancestry.

4.7.5 Environmental and genetic contributors to breast cancer

Breast cancer results from the interaction of both genetic and environmental factors. However, it remains unclear how genetic risk loci interact with environmental risk factors and to what extent. To try and understand the complex interaction, a study was performed by examining approximately 35 000 invasive breast cancer cases and 41 000 controls of European ancestry (Nickels *et al*., 2013). Potential gene-environment interactions were investigated between 23 breast cancer susceptibility SNPs and 10 environmental risk factors. The environmental factors included were body mass index, height, the use of oral contraceptives, alcohol consumption, and smoking of cigarettes, age at menarche, parity (number of births) or pregnancies of the female, the use of menopausal hormone therapy, breastfeeding and physical activity. From the study, SNP rs3817198 in *LSP1* was shown to be influenced by parity while SNP rs1045485 in *CASP8* was influenced by alcohol intake (Nickels *et al*., 2013). Performing such studies in African populations could in future aid in reduction of certain environmental exposures based on susceptibility variants carried by individuals, which in turn could reduce risk of disease development. Additionally, this may contribute to the explanation of why individuals of African ancestry are generally diagnosed earlier and have an aggressive tumour phenotype compared to their European counterparts.

5 CONCLUSION

The overall aim of this study was to determine key role players in the development of inherited breast cancer in black South African women. In order to achieve this aim, a number of objectives were undertaken. The suspected *BRCA2* founder mutations (c.582G>A and c.5771_5774delTTCA) investigated in this study together account for a cumulative mutation pick-up rate of approximately 4.46% in all studies conducted in the Division as outlined in Section 3.1. This illustrates that an examination of these two mutations alone would not be sufficient to be implemented for first-line testing for IBC in this population group and that full sequencing of the *BRCA* genes as well as large rearrangement analysis would be a more likely beneficial test offering as is currently being offered.

The results from this study did not yield sufficient evidence to associate haplotypes to disease susceptibility owing to a miscalculation in the selection of SNPs with MAFs of greater than 0.05 and the small sample size of the mutation-positive individuals. This portion of the study would need to be repeated using the correct SNP as described in Section 4.2 above. This would be ideally performed in a larger group of individuals identified to have the mutations under investigation.

In an effort to improve knowledge about the inheritance of breast cancer in the black South African population, next generation sequencing was utilised. Studies that were previously performed at the Division of Human Genetics, Wits University/NHLS showed that approximately 10% of women suspected to have an inherited breast cancer syndrome carried a deleterious mutation in either *BRCA1* or *BRCA2* and the need to expand the scale of screening was warranted. The 26 cancer susceptibility genes screened for in this study identified pathogenic mutations in three genes; *ATM*, *BARD1* and *RAD50*. The mutation pick-up rate of this study was approximately 16.67% above *BRCA1* and *BRCA2* which is higher compared to previous findings although the small sample size

needs to be taken into consideration when evaluating mutation yields. In addition, multiple novel and previously-reported VUS were identified as expected in African genomes which are genetically diverse. Although these VUS create challenges in terms of clinical management and genetic counselling, they provide additional avenues of further research studies. The study further indicates the need to move from single-gene testing to multi-gene testing which promises to be more informative with regards to the genetic aetiology of black South African individuals and at improved costs. The use of NGS applications may aid in identifying pathogenic mutations in African populations which can be used in population-specific gene testing. This can significantly reduce costs associated with testing. In future, genetic counselling and management options provided to the inherited cancer patients and their families may be greatly improved. This pilot study also highlights the importance of performing extensive studies such as the use of a larger targeted gene panel or functional studies which can potentially assist in better characterising VUS and novel variants. The study also highlights the importance of including a control cohort to aid in determining the potential pathogenicity of variants observed in the case cohort. These will contribute significantly towards a better comprehension of IBC in the black population, which will benefit the lives of individuals affected by cancer syndromes.

These cumulative results from the study have proven to be useful in reaching the aim of better understanding the genetic aetiology of inherited breast cancer in the black South African population. While some of the genes investigated have been shown to be more or less significant, this study's key benefit has been in highlighting the importance of local research in establishing a comprehensive and targeted genetic testing for this population. It has also shown that NGS technologies should be employed for diagnostic testing even in a limited resource setting as they have proven in the study to be effective. On-going research with regards to IBC in this population group should be encouraged as there are still a lot of avenues open for investigation to help us better understand the genetics of the syndrome in our local population.

6 REFERENCES

6.1 Journal articles

Acs, G., Lawton, T.J., Rebbeck, T.R., Livolsi, V.A. and Zhang, P.J., 2001. Differential expression of E-cadherin in lobular and ductal neoplasms of the breast and its biologic and diagnostic implications. *American Journal of Clinical Pathology*, *115*(1), pp.85-98.

Aloraifi, F., Boland, M.R., Green, A.J. and Geraghty, J.G., 2015. Gene analysis techniques and susceptibility gene discovery in non-BRCA1/BRCA2 familial breast cancer. *Surgical Oncology*, *24*(2), pp.100-109.

Alter, B.P., Greene, M.H., Velazquez, I. and Rosenberg, P.S., 2003. Cancer in Fanconi Anemia. *Blood*, *101*(5), pp.2072-2072.

Alter, B.P., Rosenberg, P.S. and Brody, L.C., 2007. Clinical and molecular features associated with biallelic mutations in FANCD1/BRCA2. *Journal of Medical Genetics*, *44*(1), pp.1-9.

Artandi, S. & Attardi, L. D., 2005. Pathways connecting telomeres and p53 in senescence, apoptosis, and cancer. *Biochemical and Biophysical Research Communications*, 331, pp.881-890.

Auton, A., Abecasis, G., Altshuler, D., Durbin, R., Bentley, D., Chakravarti, A., et al. 2015. A global reference for human genetic variation. *Nature*, 526, pp.68-74.

Bell, D.W., Varley, J.M., Szydlo, T.E., Kang, D.H., Wahrer, D.C., Shannon, K.E., Lubratovich, M., Verselis, S.J., Isselbacher, K.J., Fraumeni, J.F. and Birch, J.M., 1999. Heterozygous germ line hCHK2 mutations in Li-Fraumeni syndrome. *Science*, 286(5449), pp.2528-2531.

Berx, G., Staes, K., van Hengel, J., Molemans, F., Bussemakers, M. J. G., van Bokhoven, A., van Roy, F., 1995. *Cloning and characterization of the human invasion suppressor gene E-cadherin (CDH1)*. *Genomics,* 26, pp.281-289.

Bork, P., Blomberg, N. and Nilges, M., 1996. Internal repeats in the BRCA2 protein sequence. *Nature Genetics*, *13*(1), pp.22-23.

Borresen, A.L., Andersen, T.I., Garber, J., Barbier-Piraux, N., Thorlacius, S., Eyfjörd, J., Ottestad, L., Smith-Sørensen, B., Hovig, E., Malkin, D. and Friend, S.H., 1992. Screening for germ line TP53 mutations in breast cancer patients. *Cancer Research*, 52(11), pp.3234-3236.

Borresen, A.L., Andersen, T.I., Tretli, S., Heiberg, A. and Møller, P., 1990. Breast cancer and other cancers in Norwegian families with ataxia-telangiectasia. *Genes, Chromosomes and Cancer*, 2(4), pp.339-340.

Bourdon, J.-C., Fernandes, K., Murray-Zmijewski, F., Liu, G., Diot, A., Xirodimas, D. P., Saville, M. K., Lane, D. P. p53 isoforms can regulate p53 transcriptional activity. *Genes & Development*, 19, pp.2122-2137.

Brinton, L.A., Figueroa, J.D., Awuah, B., Yarney, J., Wiafe, S., Wood, S.N., Ansong, D., Nyarko, K., Wiafe-Addai, B. and Clegg-Lamptey, J.N., 2014. Breast cancer in Sub-Saharan Africa: opportunities for prevention. *Breast Cancer Research and Treatment*, *144*(3), pp.467-478.

Brown, M.A., Xu, C.F., Nicolai, H., Griffiths, B., Chambers, J.A., Black, D. and Solomon, E., 1996. The 5'end of the BRCA1 gene lies within a duplicated region of human chromosome 17q21. *Oncogene*, *12*(12), pp.2507-2513.

Cavanagh, H. and Rogers, K.M., 2015. The role of BRCA1 and BRCA2 mutations in prostate, pancreatic and stomach cancers. *Hereditary Cancer in Clinical Practice*, *13*(1), p.16.

Cheon, J. Y., Mozersky, J., & Cook-Deegan, R. (2014). Variants of uncertain significance in *BRCA*: a harbinger of ethical and policy issues to come? *Genome Medicine*, *6*(12), pp.121.

Cox, A., Dunning, A.M., Garcia-Closas, M., Balasubramanian, S., Reed, M.W., Pooley, K.A., Scollen, S., Baynes, C., Ponder, B.A., Chanock, S. and Lissowska, J., 2007. A common

coding variant in CASP8 is associated with breast cancer risk. *Nature Genetics*, *39*(3), pp.352-358.

Crawford, B., Adams, S.B., Sittler, T., van den Akker, J., Chan, S., Leitner, O., Ryan, L., Gil, E. and van't Veer, L., 2017. Multi-gene panel testing for hereditary cancer predisposition in unsolved high-risk breast and ovarian cancer patients. *Breast Cancer Research and Treatment*, pp.1-8.

Cunningham, F., Amode, M.R., Barrell, D., Beal, K., Billis, K., Brent, S., Carvalho-Silva, D., Clapham, P., Coates, G., Fitzgerald, S. and Gil, L., 2014. Ensembl 2015. *Nucleic Acids Research*, p.gku1010.

Dalal, S., Beunza, J.J., Volmink, J., Adebamowo, C., Bajunirwe, F., Njelekela, M., Mozaffarian, D., Fawzi, W., Willett, W., Adami, H.O. and Holmes, M.D., 2011. Non-communicable diseases in sub-Saharan Africa: what we know now. *International Journal of Epidemiology*, *40*(4), pp.885-901.

D'Andrea, A.D., 2013. BRCA1: a missing link in the Fanconi anemia/BRCA pathway. *Cancer Discovery*, *3*(4), pp.376-378.

DeSantis, C., Fedewa, S., Goding Sauer, A., Kramer, J., Smith, R. and Jemal, A., 2016. Breast cancer statistics, 2015: Convergence of incidence rates between black and white women. *Cancer Journal for Clinicians*, 66 (1), pp.31-42.

Dodova, R, Mitkova, A., Dacheva, D., Hadjo, L.B., Vlahova, A.I., Hadjieva, M.S.T., Valev, S.S., Caulevska, M.M., Popova, S.D., Popov, I.E. and Dikov, T.I., 2015. Spectrum and frequencies of BRCA1/2 mutations in Bulgarian high risk breast cancer patients. *BMC Cancer*, *15*(1), p.523.

Dong, H., Nebert, D.W., Bruford, E.A., Thompson, D.C., Joenje, H. and Vasiliou, V., 2015. Update of the human and mouse Fanconi anemia genes. *Human Genomics*, *9*(1), p.1.

Easton, D., Pooley, K., Dunning, A., Pharoah, P., Thompson, D., Ballinger, D., et al. 2007. Genome-wide association study identifies novel breast cancer susceptibility loci. *Nature*, 447, pp. 1087–1093.

Easton, D.F., 1999. How many more breast cancer predisposition genes are there?. *Breast Cancer Research*, 1(1), p.14.

Eccles, B., Copson,E., Maishman, T., Abraham, J. and Eccles, D. Understanding of BRCA VUS genetic results by breast cancer specialists. *BMC Cancer*, 15, doi: 10.1186/s12885-015-1934-1.

Ernst, C., Hahnen, E., Engel, C., Nothnagel, M., Weber, J., Schmutzler, R. and Hauke, J. 2018. Performance of in silico prediction tools for the classification of rare *BRCA1/2* missense variants in clinical diagnostics. *BMC Medical Genomics*, 11(35), doi: 10.1186/s12920-018-0353-y.

Feben, C., Spencer, C., Lochan, A., Laing, N., Fieggen, K., Honey, E., Wainstein, T. and Krause, A., 2017. Biallelic BRCA2 mutations in two black South African children with Fanconi anaemia. *Familial Cancer*, pp.1-6.

Felix, G., Zheng, Y., Guindalini, R., Machado-Lopes, T., Zhang, J., Cortes, J., et al. 2017. BROCA gene panel testing in African descendants from northeastern Brazil: Genetic susceptibility profile of an admixed population. *Journal of Clinical Oncology*, 35(15), doi: 10.1200/JCO.2017.35.15_suppl.1572.

Feng, Y., Rhie, S.K., Huo, D., Ruiz-Narvaez, E.A., Haddad, S.A., Ambrosone, C.B., John, E.M., Bernstein, L., Zheng, W., Hu, J.J. and Ziegler, R.G., 2017. Characterizing Genetic Susceptibility to Breast Cancer in Women of African Ancestry. *Cancer Epidemiology and Prevention Biomarkers*, pp.cebp-0567.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D. and Bray, F., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International Journal of Cancer*, *136*(5), pp.E359-E386.

Foley, S., Rios, J., Mgbemena, V., Robinson, L., Hampele, H., Toland, A., et al. 2015. Use of Whole Genome Sequencing for Diagnosis and Discovery in the Cancer Genetics Clinic. *EBioMedicine*, 2(1), pp. 74-81.

Ford, D. and Easton, D.F., 1995. The genetics of breast and ovarian cancer. *British Journal of Cancer*, *72*(4), pp.805-812.

Ford, D., Easton, D.F., Stratton, M., Narod, S., Goldgar, D., Devilee, P., Bishop, D.T., Weber, B., Lenoir, G., Chang-Claude, J. and Sobol, H., 1998. Genetic heterogeneity and penetrance analysis of the BRCA1 and BRCA2 genes in breast cancer families. *The American Journal of Human Genetics*, *62*(3), pp.676-689.

Francies, F.Z., Wainstein, T., De Leeneer, K., Cairns, A., Murdoch, M., Nietz, S., Cubasch, H., Poppe, B., Van Maerken, T., Crombez, B. and Coene, I.L.S.E., 2015. BRCA1, BRCA2 and PALB2 mutations and CHEK2 c. 1100delC in different South African ethnic groups diagnosed with premenopausal and/or triple negative breast cancer. *BMC Cancer*, *15*(1), p.1.

Fuks, F., Milner, J. & Kouzarides, T., 1998. BRCA2 associates with acetyltransferase activity when bound to P/CAF. *Oncogene*, *17*(19), pp.2531-2534.

Gaceb, H., Cherbal, F., Bakour, R., Ould-Rouis, A. and Mahfouf, H., 2017. Clinicopathological and Molecular Study of Triple-Negative Breast Cancer in Algerian Patients. *Pathology & Oncology Research*, pp.1-12.

Gayther, S.A., Russell, P., Harrington, P., Antoniou, A.C., Easton, D.F. and Ponder, B.A., 1999. The contribution of germline BRCA1 and BRCA2 mutations to familial ovarian cancer: no evidence for other ovarian cancer–susceptibility genes. *The American Journal of Human Genetics*, *65*(4), pp.1021-1029.

Ginsburg, O., Bray, F., Coleman, M.P., Vanderpuye, V., Eniu, A., Kotha, S.R., Sarker, M., Huong, T.T., Allemani, C., Dvaladze, A. and Gralow, J., 2016. The global burden of women’s cancers: a grand challenge in global health. *The Lancet*.

Goldgar, D., Healey, S., Dowty, J., Da Silva, L., Chen, X., Spurdle, A., et al. 2011. Rare variants in the ATM gene and risk of breast cancer. *Breast Cancer Research*, 13(4), doi: 10.1186/bcr2919.

Gracia-Aznarez, F.J., Fernandez, V., Pita, G., Peterlongo, P., Dominguez, O., de la Hoya, M., Duran, M., Osorio, A., Moreno, L., Gonzalez-Neira, A. and Rosa-Rosa, J.M., 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), p.e55681.

Gray, P., Dunlop, C., and Elliott, A., 2015. Not All Next Generation Sequencing Diagnostics are Created Equal: Understanding the Nuances of Solid Tumor Assay Design for Somatic Mutation Detection. *Cancers*, 7(3), pp.1313–1332.

Guan, Y., Hu, H., Peng, Y., Gong, Y., Yi, Y., Shao, L., et al. 2014. Detection of inherited mutations for hereditary cancer using target enrichment and next generation sequencing. *Familial Cancer*, 14 (1), 9-18.

Gurdasani, D., Carstensen, T., Tekola-Ayele, F., Pagani, L., Tachmazidou, I., Hatzikotoulas, K., Karthikeyan, S., Iles, L., Pollard, M.O., Choudhury, A. and Ritchie, G.R., 2015. The African genome variation project shapes medical genetics in Africa. *Nature*, *517*(7534), pp.327-332.

Haffty, B.G., Silber, A., Matloff, E., Chung, J. and 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), pp.133-137.

Hall, J.M., Lee, M.K., Newman, B., Morrow, J.E. and Anderson, L.A., 1990. Linkage of Early-Onset Familial Breast Cancer to Chromosome 17q21.*Science*, *250*(4988), pp.1684-1689.

Hedge, M., Chong, B., Biazo, M., Chin, L., Ward, P., Chintagumpala, M., et al. 2005. A homozygous mutation in *MSH6* causes Turcot syndrome. *Clinical Cancer Research*, 11(13), pp.4689-4693.

Heitzer, E., Lax, S., Lafer, I., Muller, S., Pristauz, G., Ulz, P., et al. 2013. Multiplex genetic cancer testing identifies pathogenic mutations in TP53 and CDH1 in a patient with

bilateral breast and endometrial adenocarcinoma. *BMC Medical Genetics*, doi: 10.1186/ 1471-2350-14-129.

Hoche, F., Frankenberg, E., Rambow, J., Theis, M., Harding, J., Qirshi, M., et al. 2014. Cognitive Phenotype in Ataxia-Telangiectasia. *Pediatric Neurology*, 51, pp.297-310.

Holliday, R., 2006. Epigenetics: A Historical Overview. *Epigenetics*, 1(2), pp.76-80.

Howlett, N.G., Taniguchi, T., Olson, S., Cox, B., Waisfisz, Q., de Die-Smulders, C., Persky, N., Grompe, M., Joenje, H., Pals, G. and Ikeda, H., 2002. Biallelic inactivation of BRCA2 in Fanconi anemia. *Science*, *297*(5581), pp.606-609.

Hussain, S., Witt, E., Huber, P.A., Medhurst, A.L., Ashworth, A. and Mathew, C.G., 2003. Direct interaction of the Fanconi anaemia protein FANCG with BRCA2/FANCD1. *Human Molecular Genetics*, *12*(19), pp.2503-2510.

Jackson, S.P. and Bartek, J., 2009. The DNA-damage response in human biology and disease. *Nature*, *461*(7267), pp.1071-1078.

James, P.A., Sawyer, S., Boyle, S., Young, M.A., Kovalenko, S., Doherty, R., McKinley, J., Alsop, K., Beshay, V., Harris, M. and Fox, S., 2015. Large genomic rearrangements in the familial breast and ovarian cancer gene BRCA1 are associated with an increased frequency of high risk features. *Familial Cancer*, *14*(2), pp.287-295.

Kerr, I., Cox, H., Moyes, K., Evans, B., Burdett, B., van Kan, A., et al. 2017. Assessment of in silico protein sequence analysis in the clinical classification of variants in cancer risk genes. *Journal of Community Genetics*, 8(2), pp.87–95.

Kircher, M., Witten, D., Jain, P., O'Roak, BJ., Cooper, GM., Shendure, J., 2014. A general framework for estimating the relative pathogenicity of human genetic variants. *Nature Genetics*. doi: 10.1038/ng.2892.

Klasberg, S., Bitard-Feildel, T., & Mallet, L. (2016). Computational Identification of Novel Genes: Current and Future Perspectives. *Bioinformatics and Biology Insights*, 10, pp.121–131.

Konecny, M., Milly, M., Zavodna, K., Weismanova, E., Gregorova, J., et al. 2011. Comprehensive genetic characterization of hereditary breast/ovarian cancer families from Slovakia. *Breast Cancer Research and Treatment*, 126 (1), pp.119-130.

Kroese, M., Zimmern, R.L. and Sanderson, S., 2004. Genetic tests and their evaluation: can we answer the key questions?. *Genetics in Medicine*, *6*(6), pp.475-480.

Landrum, M., Lee, J., Riley, G., Jang, W., Rubinstein, W., Church, D. and Maglott, D. 2014. ClinVar: public archive of relationships among sequence variation and human phenotype. *Nucleic Acids Research*, doi: 10.1093/nar/gkt1113.

Lawson, J.S. and Heng, B., 2010. Viruses and breast cancer. *Cancers*, *2*(2), pp.752-772.

Lek, M., Karczewski, K., Minikel, E., Samocha, K., Banks, E., Fennell, T., et al. 2016. Analysis of protein-coding genetic variation in 60,706 humans. *Nature*, 536, pp.285-291.

Li, D.M. and Sun, H., 1998. PTEN/MMAC1/TEP1 suppresses the tumorigenicity and induces G1 cell cycle arrest in human glioblastoma cells. *Proceedings of the National Academy of Science*, 95, pp.15406-15411.

LI, F.P. and FRAUMENI, J.F., 1969. Soft-tissue sarcomas, breast cancer, and other neoplasms: a familial syndrome?. *Annals of Internal Medicine*, *71*(4), pp.747-752.

Li, J., Yen, C., Liaw, D., Podsypanina, K., Bose, S., Wang, S.I., Puc, J., Miliaresis, C., Rodgers, L., McCombie, R. and Bigner, S.H., 1997. PTEN, a putative protein tyrosine phosphatase gene mutated in human brain, breast, and prostate cancer. *Science*, *275* (5308), pp.1943-1947.

Lindor, N., Guidugli, L., Wang, X., Vallée, M., Monteiro, A., Tavtigian, S., et al. 2012. A review of a multifactorial probability-based model for classification of BRCA1 and BRCA2 variants of uncertain significance (VUS). *Human Mutation*, 33(1), pp. 8-21.

Liu, N., Lamerdin, J. E., Tucker, J. D., Zhou, Z.-Q., Walter, C. A., Albala, J. S., Busch, D. B., Thompson, L. H., 1997. The human XRCC9 gene corrects chromosomal instability and

mutagen sensitivities in CHO UV40 cells. *Proceedings of the National Academy of Sciences*, 94, pp.9232-9237.

Loubser, F., Edge, J. and Fieggen, K., 2008. Epidemiology, risk factors and genetics in breast cancer. *Continuing Medical Education*, *26*(10), p.497.

Manke, I.A., Lowery, D.M., Nguyen, A. and Yaffe, M.B., 2003. BRCT repeats as phosphopeptide-binding modules involved in protein targeting. *Science*, *302*(5645), pp.636-639.

Mansouri, A., Spurr, N., Goodfellow, P. N., Kemler, R., 1988. Characterization and chromosomal localization of the gene encoding the human cell adhesion molecule uvomorulin. *Differentiation*, 38, pp.67-71.

Mantere, T., Haanpää, M., Hanenberg, H., Schleutker, J., Kallioniemi, A., Kähkönen, M., et al. 2015. Finnish Fanconi anemia mutations and hereditary predisposition to breast and prostate cancer. *Clinical Genetics,* pp. 88(1), 68-73.

Marafie, M., Al-Temaimi, R., Megarbane, A. and Al-Mulla, F., 2015. Germline mutations in early-onset or hereditary breast cancer from the Middle East. *Journal of Clinical Oncology,* 33, doi: 10.1200/jco.2015.33.28_suppl.20.

Maree, J., Wright, S. and Lu, X., 2013. Breast cancer risks and screening practices among women living in a resource poor community in Tshwane, South Africa. *The Breast Journal*, *19*(4), pp.453-454.

Margulies, M., Egholm, M., Altman, W., Attiya, S., Bader, J., Bemben, L., et al. 2005. Genome sequencing in microfabricated high-density picolitre reactors. *Nature*, 437(7057), pp. 376-380.

Marouf, C., Hajji, O., Tazzite, A., Jouhadi., Benider, A., Nadifi, S. 2017. Germline variants in the ATM gene and breast cancer susceptibility in Moroccan women: A meta-analysis. *Egyptian Journal of Medical Human Genetics* 18 (4), 2017, pp.329-334.

McGuigan, F.E. and Ralston, S.H., 2002. Single nucleotide polymorphism detection: allelic discrimination using TaqMan. *Psychiatric Genetics*, *12*(3), pp.133-136.

McLaren, W., Gil, L., Hunt, S., Riat, H., Ritchie, G., Thormann, A., Flicek, P., Cunningham, F., 2016. The Ensembl Variant Effect Predictor. *Genome Biology* 17:122.

Melchor, L. and Benítez, J., 2013. The complex genetic landscape of familial breast cancer. *Human Genetics*, 132(8), pp.845-863.

Meyer, S., Tischkowitz, M., Chandler, K., Gillespie, A., Birch, J.M. and Evans, D.G., 2014. Fanconi anaemia, BRCA2 mutations and childhood cancer: a developmental perspective from clinical and epidemiological observations with implications for genetic counselling. *Journal of Medical Genetics*, *51*(2), pp.71-75.

Meza, J.E., Brzovic, P.S., King, M.C. and Klevit, R.E., 1999. Mapping the functional domains of BRCA1 interaction of the ring finger domains of BRCA1 and BARD1. *Journal of Biological Chemistry*, *274*(9), pp.5659-5665.

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M., et al. 2015. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. *Nature Genetics*, 47, pp.373–380.

Michailidou, K., Hall P., Gonzalez-Neira A., Ghoussaini M., Dennis J., Milne R., et al. 2013. Large-scale genotyping identifies 41 new loci associated with breast cancer risk. *Nature Genetics*, *45*(4), pp.353-361.

Michailidou, K., Lindström, S., Dennis, J., Beesley, J., Hui, S., Kar, S., et al. 2017. Association analysis identifies 65 new breast cancer risk loci. *Nature*, 551, pp.92–94.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P., Harshman, K., Tavtigian, S., et al. (1994). A Strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene *BRCA1*. *Science*, 266, pp.66-71.

Miller, J. & Pearce, D. 2014. Nonsense-mediated decay in genetic disease: Friend or foe?. *Mutation Research/Reviews in Mutation Research*, 762, 52-64.

Milner, J., Ponder, B., Hughes-Davies, L., Seltmann, M. and Kouzarides, T., 1997. Transcriptional activation functions in BRCA2. *Nature*, *386*(6627), p.772.

Morgan, N.V., Essop, F., Demuth, I., de Ravel, T., Jansen, S., Tischkowitz, M., Lewis, C.M., Wainwright, L., Poole, J., Joenje, H. and Digweed, M., 2005. A common Fanconi anemia mutation in black populations of sub-Saharan Africa. *Blood*, *105*(9), pp.3542-3544.

Neuhausen, S., Mazoyer, S., Friedman, L., Stratton, M., Offit, K., Caligo, A., et al., 1996. Haplotype and phenotype analysis of six recurrent *BRCA1* mutations in 61 families: results of an international study. *American Journal of Human Genetics*, 58, pp.271–280.

Nickels, S., Truong, T., Hein, R., Stevens, K., Buck, K., Behrens, S., et al. (2013) Evidence of Gene–Environment Interactions between Common Breast Cancer Susceptibility Loci and Established Environmental Risk Factors, *PLoS Genetics*, 9(3): e1003284. https://doi.org/10.1371/journal.pgen.1003284

Noh, J.M., Kim, J., Cho, D.Y., Choi, D.H., Park, W. and Huh, S.J., 2015. Exome sequencing in a breast cancer family without BRCA mutation. *Radiation Oncology Journal*, *33*(2), pp.149-154.

Oluwagbemiga, L.A., Oluwole, A. and Kayode, A.A., 2012. Seventeen years after BRCA1: what is the BRCA mutation status of the breast cancer patients in Africa?–a systematic review. *SpringerPlus*, 1(1), p.1.

Pal, T., Bonner, D., Kim, J., Monteiro, A.N., Kessler, L., Royer, R., Narod, S.A. and Vadaparampil, S.T., 2013. Early onset breast cancer in a registry-based sample of African-American women: BRCA mutation prevalence, and other personal and system-level clinical characteristics. *The Breast Journal*, *19*(2), pp.189-192.

Peshkin, B.N., Alabek, M.L. and Isaacs, C., 2010. *BRCA1/2* mutations and triple negative breast cancers. *Breast Disease*, *32*(1, 2), pp.25-33.

Petersen, D., Libiger, O., Tindall, E., Hardie, R., Hannick, L., Glashoff, H., et al. 2013 Complex Patterns of Genomic Admixture within Southern Africa. *PLoS Genetics,* 9(3): e1003309. https://doi.org/10.1371/journal.pgen.1003309

Petridis, C., Shinomiya, I., Kohut, K., Gorman, P., Caneppele, M., Shah, V., Troy, M., Pinder, S.E., Hanby, A., Tomlinson, I. and Trembath, R.C., 2014. Germline CDH1 mutations in bilateral lobular carcinoma in situ. *British Journal of Cancer*, *110*(4), pp.1053-1057.

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

Pharoah, P.D., Guilford, P., Caldas, C. and International Gastric Cancer Linkage Consortium, 2001. Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. *Gastroenterology*, 121(6), pp.1348-1353.

Plon, S., Eccles, D., Easton, D., Foulkes, W., Genuardi, M., Greenblatt, M., et al. 2008. Sequence variant classification and reporting: recommendations for improving the interpretation of cancer susceptibility genetic test results. *Human Mutation*, 29(11), pp.1282-1291.

Ramathuba, D.U., Ratshirumbi, C.T. and Mashamba, T.M., 2015. Knowledge, attitudes and practices toward breast cancer screening in a rural South African community. *Curationis*, *38*(1), pp.1-8.

Reeves, M.D., Yawitch, T.M., van der Merwe, N.C., van den Berg, H.J., Dreyer, G. and 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*, *110*(5), pp.677-682.

Rehm, H.L., Bale, S.J., Bayrak-Toydemir, P., Berg, J.S., Brown, K.K., Deignan, J.L., Friez, M.J., Funke, B.H., Hegde, M.R., Lyon, E. and Working Group of the American College of Medical Genetics. 2013. ACMG clinical laboratory standards for next-generation sequencing. *Genetics in Medicine*, *15*(9), pp.733-747.

Riahi, A., Chabouni-Bouhamed, H. and Kharrat, M., 2017. Prevalence of BRCA1 and BRCA2 large genomic rearrangements in Tunisian high risk breast/ovarian cancer families: Implications for genetic testing. *Cancer Genetics*, *210*, pp.22-27.

Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., et al. 2015. Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in Medicine*, 17(5), pp. 405-24.

Richter, L., Norris, S., Hosegood, V., Newell, M., Swart, T. and Ginsburg, C. 2005. "Birth To Twenty": A Study of Health and Development Across Three Generations. *Epidemiology*, 16(5), pp. S152-S153.

Rodriguez, S., Gaunt, T.R. and Day, I.N., 2009. Hardy-Weinberg equilibrium testing of biological ascertainment for Mendelian randomization studies. *American Journal of Epidemiology*, *169*(4), pp.505-514.

Rosenberg, P.S., Alter, B.P. and Ebell, W., 2008. Cancer risks in Fanconi anemia: findings from the German Fanconi Anemia Registry. *Haematologica*,*93*(4), pp.511-517.

Savage, S.A., Ballew, B.J., Giri, N., Chandrasekharappa, S.C., Ameziane, N., de Winter, J. and Alter, B.P., 2016. Novel FANCI mutations in Fanconi anemia with VACTERL association. *American Journal of Medical Genetics Part A*, *170*(2), pp.386-391.

Sawyer, S.L., Tian, L., Kähkönen, M., Schwartzentruber, J., Kircher, M., Majewski, J., Dyment, D.A., Innes, A.M., Boycott, K.M., Moreau, L.A. and Moilanen, J.S., 2015. Biallelic mutations in BRCA1 cause a new Fanconi anemia subtype. *Cancer Discovery*, *5*(2), pp.135-142.

Schaid, D.J. and Jacobsen, S.J., 1999. Blased Tests of Association: comparisons of allele frequencies when departing from Hardy-Weinberg proportions. *American Journal of Epidemiology*, *149*(8), pp.706-711.

Schoeman, M., Apffelstaedt, J., Baatjes, K. & Urban, M. (2013). Implementation of a breast cancer genetic service in South Africa – lessons learned, *South African Medical Journal*, 103(8), 529-533.

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

Schwarzman, M.R., Ackerman, J.M., Dairkee, S.H., Fenton, S.E., Johnson, D., Navarro, K.M., Osborne, G., Rudel, R.A., Solomon, G.M., Zeise, L. and Janssen, S., 2015. Screening for chemical contributions to breast cancer risk: a case study for chemical safety evaluation. *Environmental Health Perspectives (Online)*, *123*(12), p.1255.

Seal, S., Thompson, D., Renwick, A., Elliott, A., Kelly, P., Barfoot, R., Chagtai, T., Jayatilake, H., Ahmed, M., Spanova, K. and North, B., 2006. Truncating mutations in the Fanconi anemia J gene BRIP1 are low-penetrance breast cancer susceptibility alleles. *Nature Genetics*, *38*(11), pp.1239-1241.

Sequeiros, J., Paneque, M., Guimarães, B., Rantanen, E., Javaher, P., Nippert, I., Schmidtke, J., Kääriäainen, H., Kristoffersson, U. and Cassiman, J.J., 2012. The wide variation of definitions of genetic testing in international recommendations, guidelines and reports. *Journal of Community Genetics*, *3*(2), pp.113-124.

Seymour, H.J., Wainstein, T., Macaulay, S., Haw, T. and Krause, A., 2016. Breast cancer in high-risk Afrikaner families: Is BRCA founder mutation testing sufficient?. *SAMJ: South African Medical Journal*, *106*(3), pp.264-267.

Sharrard, R.M. and Maitland, N. J. Alternative splicing of the human PTEN/MMAC1/TEP1 gene. *Biochimica et Biophysica Acta*, 1494, pp.282-285.

Sherry, S., Ward, M., Kholodov, M., Baker, J., Phan, L, Smigielski, E. and Sirotkin, K. 2001. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Research*, 29(1), pp.308-11.

Shirts B., Casadei, S., Jacobson, A., Lee, M., Gulsuner, S., Bennet, R., et al. 2016. Improving performance of multigene panels for genomic analysis of cancer predisposition, *Genetics in Medicine*, 18 (10), pp.974-981.

Sim, N., Kumar, P., Hu., J., Henikoff, S., Schneider, G., Ng, P., 2012. SIFT web server: predicting effects of amino acid substitutions on proteins. *Nucleic Acids Research*, 40 (1), pp.452–457.

Slatkin, M., 2008. Linkage disequilibrium-understanding the evolutionary past and mapping the medical future. *Nature Genetics Reviews,* 9, pp.477-485.

Sluiter, M., Mew, S. and van Rensburg, E.J., 2009. PALB2 sequence variants in young South African breast cancer patients. *Familial Cancer*, *8*(4), p.347.

Smigal, C., Jemal, A. and Ward, E., 2006. Trends in Breast Cancer by Race and Ethnicity: Update 2006. *CA; A Cancer Journal For Clinicians,* 56, 168-183.

Smyth, C., Špakulová, I., Cotton-Barratt, O., Rafiq, S., Tapper, W., Upstill-Goddard, R., MacArthur, B., 2015. Quantifying the cumulative effect of low-penetrance genetic variants on breast cancer risk. *Molecular Genetics & Genomic Medicine*, *3*(3), pp.182–188.

Sohn, J.Y., Park, B., Yoon, K.A., Park, S.J., Lee, M.H., Cho, E.H., Lee, K.S., Lim, M.C., Kong, S.Y. and Lee, E.S., 2016. Prevalence of BRCA1 and BRCA2 small mutation and large genomic rearrangements in breast cancer patients visiting a genetic counseling clinic. *Cancer Research*, *76*(14 Supplement), pp.3474-3474.

Sorrell, A.D., Espenschied, C.R., Culver, J.O. and Weitzel, J.N., 2013. Tumor Protein p53 (TP53) Testing and Li-Fraumeni Syndrome. *Molecular Diagnosis & Therapy*, *17*(1), pp.31-47.

Steck, P. A., Pershouse, M. A., Jasser, S. A., Yung, W. K. A., Lin, H., Ligon A.H., et al. 1997. Identification of a candidate tumour suppressor gene, MMAC1, at chromosome 10q23.3 that is mutated in multiple advanced cancers. *Nature Genetics*, 15: 356-362.

Stephens, M. and Donnelly, P., 2003. A comparison of bayesian methods for haplotype reconstruction from population genotype data. *The American Journal of Human Genetics*, *73*(5), pp.1162-1169.

Stephens, M., Smith, N.J. and Donnelly, P., 2001. A new statistical method for haplotype reconstruction from population data. *The American Journal of Human Genetics*, *68*(4), pp.978-989.

Sy, S.M., Huen, M.S. and Chen, J., 2009. PALB2 is an integral component of the BRCA complex required for homologous recombination repair. *Proceedings of the National Academy of Sciences*, *106*(17), pp.7155-7160.

Tessereau, C., Léoné, M., Buisson, M., Duret, L., Sinilnikova, O.M. and Mazoyer, S., 2015. Occurrence of a non-deleterious gene conversion event in the BRCA1 gene. *Genes, Chromosomes and Cancer*, *54*(10), pp.646-652.

Thai, T.H., Du, F., Tsan, J.T., Jin, Y., Phung, A., Spillman, M.A., Massa, H.F., Muller, C.Y., Ashfaq, R., Mathis, J.M. and Miller, D.S., 1998. Mutations in the BRCAI-Associated RING Domain (BARD1) Gene in Primary Breast, Ovarian and Uterine Cancers. *Human Molecular Genetics*, 7(2), pp.195-202.

Thompson, D. and Easton, D., 2004. The genetic epidemiology of breast cancer genes. *Journal of Mammary Gland Biology and Neoplasia*, *9*(3), pp.221-236.

van der Merwe, N., Hamel, N., Schneider, S-R., Apffelstaedt, J., Wijnen, J. & Foulkes, W. (2012). A founder BRCA2 mutation in non-Afrikaner breast cancer patients of the Western Cape of South Africa. *Clinical Genetics*, 81: 179–184.

Van der Merwe, N.C. and Van Rensburg, E.J., 2009. Hereditary breast/ovarian cancer and BRCA mutations: a South African perspective. *Current Oncology*, *16*, p.347.

van Os, N., Jansen, A., van Deuren, M., Haraldsson, A., van Driel, N., Etzioni, A., van der Flier, M., et al. 2017. Ataxia-telangiectasia: Immunodeficiency and survival. *Clinical Immunology*, 178, pp.45-55.

Wainstein, T., Kerr, R., Mitchell, C., Madaree, S., Essop, F., Vorster, E., et al. (2013). Fanconi anaemia in black South African patients heterozygous for the *FANCG* c.637-643delTACCGCC founder mutation. *South African Medical Journal*, 103(12), pp. 970-973.

Walsh, T., Casadei, S., Lee, M., Pennill, C., Nord, A., Thornton, A., et al. 2011. Mutations in 12 genes for inherited ovarian, fallopian tube and peritoneal carcinoma identified by massively parallel sequencing. *Proceedings of the National Academy of Science*, 108, pp.18032-18037.

Weren, R., Mensenkamp, A., Simons, M., Eijkelenboom, A., Sie, A., Ouchene, H., et al. 2017. Novel *BRCA1* and *BRCA2* Tumor Test as Basis for Treatment Decisions and Referral for Genetic Counselling of Patients with Ovarian Carcinomas. *Human Mutation*, 38 (2), pp. 226-235.

Weymann, D., Laskin, J., Roscoe, R., Schrader, K., Chia, S., Yip, S., et al. 2017. The cost and cost trajectory of whole-genome analysis guiding treatment of patients with advanced cancers. *Molecular Genetics and Genomic Medicine*, 5(3), pp. 251-260.

Wigginton, J.E., Cutler, D.J. and Abecasis, G.R., 2005. A note on exact tests of Hardy-Weinberg equilibrium. *The American Journal of Human Genetics*, *76*(5), pp.887-893.

Winship, I. and Southey, M.C., 2016. Gene panel testing for hereditary breast cancer. *The Medical Journal of Australia, 204*(5), pp.188-190.

Wong, E., Southey, M., Fox, S., Brown, M., Dowty, J., Jenkins, M., et al. 2011. Constitutional methylation of the BRCA1 promoter is specifically associated with BRCA1 mutation-associated pathology in early-onset breast cancer. *Cancer Prevention Research,* 4(1), pp. 23–33.

Wooster, R., Bignell, G., Lancaster, J., Swift, S., Seal, S., Mangion, J., Collins, N., Gregory, S., Gumbs, C., Micklem, G. and Barfoot, R., 1995. Identification of the breast cancer susceptibility gene BRCA2. *Nature*,*378*(6559), pp.789-792.

Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T. and Averill, D., 1994. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. *Science*,*265*(5181), pp.2088-2090.

Xia, B., Sheng, Q., Nakanishi, K., Ohashi, A., Wu, J., Christ, N., Liu, X., Jasin, M., Couch, F.J. and Livingston, D.M., 2006. Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. *Molecular Cell*, *22*(6), pp.719-729.

Yang, H., Jeffrey, P.D., Miller, J., Kinnucan, E., Sun, Y., Thomä, N.H., Zheng, N., Chen, P.L., Lee, W.H. and Pavletich, N.P., 2002. BRCA2 function in DNA binding and recombination from a BRCA2-DSS1-ssDNA structure. *Science*, *297*(5588), pp.1837-1848.

Yano, K.I., Morotomi, K., Saito, H., Kato, M., Matsuo, F. and Miki, Y., 2000. Nuclear localization signals of the BRCA2 protein. *Biochemical and Biophysical Research Communications*, *270*(1), pp.171-175.

Yuan, S.S.F., Lee, S.Y., Chen, G., Song, M., Tomlinson, G.E. and Lee, E.Y.P., 1999. BRCA2 is required for ionizing radiation-induced assembly of Rad51 complex in vivo. *Cancer Research*, *59*(15), pp.3547-3551.

Yurgelun, M.B., Hiller, E. and Garber, J.E., 2015. Population-wide screening for germline BRCA1 and BRCA2 mutations: too much of a good thing?. *Journal of Clinical Oncology*, *33*(28), pp.3092-3095.

Zhang, J., Fackenthal, J.D., Huo, D., Zheng, Y. and Olopade, O.I., 2010. Searching for large genomic rearrangements of the BRCA1 gene in a Nigerian population. *Breast Cancer Research and Treatment*, *124*(2), pp.573-577.

Zhang, J., Fackenthal, J.D., Zheng, Y., Huo, D., Hou, N., Niu, Q., Zvosec, C., Ogundiran, T.O., Hennis, A.J., Leske, M.C. and Nemesure, B., 2012. Recurrent BRCA1 and BRCA2 mutations in breast cancer patients of African ancestry. *Breast Cancer Research and Treatment*, *134*(2), pp.889-894.

6.2 Books

Adzhubei, I., Jordan, D., Sunyaev, SR., 2013. *Predicting functional effect of human missense mutations using PolyPhen-2*. Current Protocols in Human Genetics, Chapter 7, Unit7.20

Bland, K., Copeland, E., Klimberg, S. and Gradishar, W., 2017. *The Breast: the Comprehensive Management of Benign and Malignant Diseases*. 5th ed. Saunders, pp.247-248.

Hartl, D., 2011. *Essential Genetics: A Genomic Perspective*. 5th ed. Jones and Bartlett Publishers, pp. 23-24.

Nussbaum, R., McInnes, R. and Willard, H., 2007. *Thompson & Thompson Genetics in Medicine*. 7th ed. Saunders, Philadelphia, pp. 461-462.

Shetty, M., 2013. Burden of Breast and Gynecological Cancers in Low-Resource Countries, In: *Breast and Gynecological Cancers: An Integrated Approach for Screening and Early Diagnosis in Developing Countries*, by R. Sankaranarayanan & J. Ferlay. Springer, New York, pp 4-10.

Strachan, T. & Read, A., 2010. *Human Molecular Genetics*. 4th ed. Garland Science, USA and UK, pp. 537-539.

6.3 Electronic sources

Breast Cancer Information Core. 2017. *BRCA2* Mutation Database. Retrieved November 22, 2017 from https://research.nhgri.nih.gov/projects/bic/Member/cgi-bin/bic_query_result.cgi?exact_search=1&table=brca2_exons&exon_type=11E

Breast Cancer Information Core. 2017. *BRCA2* Mutation Database. Retrieved November 22, 2017 from https://research.nhgri.nih.gov/projects/bic/Member/cgi-bin/bic_query_result.cgi?exact_search=1&table=brca2_exons&exon_type=7

Census. 2011. Western Cape. Retrieved December 12, 2017 from https://census2011.adrianfrith.com/place/1

Census. 2011. Western Cape. Retrieved December 12, 2017 from https://census2011.adrianfrith.com/place/7

Chen, W. 2016. The molecular aetiology of inherited breast cancer in the South African black population (Master's dissertation). University of the Witwatersrand. Retrieved February 17, 2017 from http://wiredspace.wits.ac.za/bitstream/handle/10539/19755/0700425A%20Wenlong%20Chen%20MSc%20Med.pdf?sequence=1

Ensembl genome browser 85. 2016. *BRCA1*. Retrieved August 4, 2016 from http://www.ensembl.org/Homo_sapiens/Gene/Summary?db=core;g=ENSG00000012048;r=17:43044295-43170245

Ensembl genome browser 85. 2016. *BRCA2*. Retrieved August 5, 2016 from http://www.ensembl.org/Homo_sapiens/Gene/Compara_Alignments?db=core;g=ENSG00000139618;r=13:32315474-32400266

Genetics Home Reference. 2018. Genes. Retrieved January 24, 2018 from https://ghr.nlm.nih.gov/gene

National Cancer Registry. 2012. Summary Statistics of Cancer Diagnosed Histologically in 2012. Retrieved October 10, 2017 from http://www.nicd.ac.za/assets/files/NCR%202012%20results.pdf

World Health Organization. 2015. Cancer. Retrieved August 1, 2017 from http://www.who.int/mediacentre/factsheets/fs297/en/

World Health Organization. 2016. Cancer. Retrieved August 1, 2017 from http://www.who.int/cancer/detection/breastcancer/en/index1.html

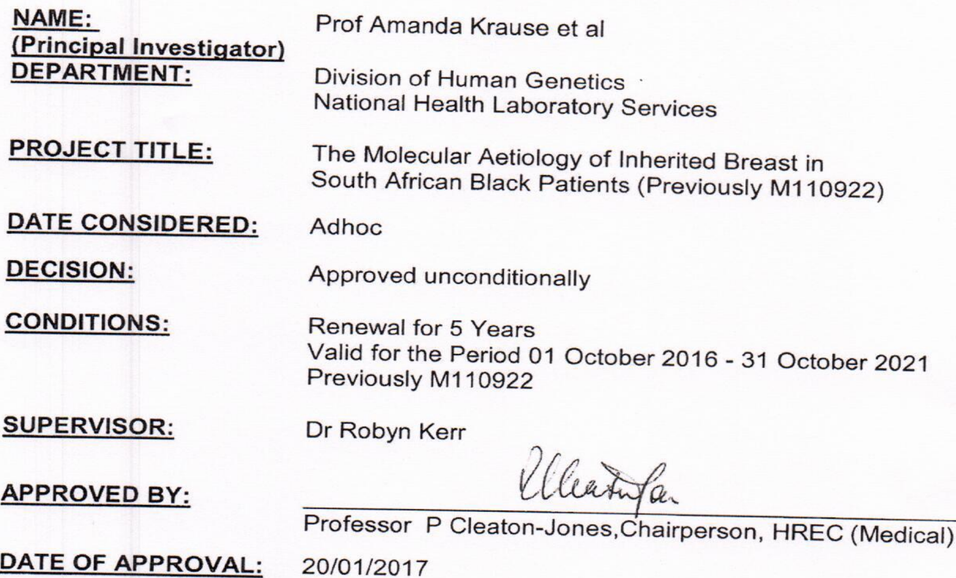
7 APPENDICES

Appendix A: Ethics clearance certificate

R14/49 Prof Amanda Krause et al

HUMAN RESEARCH ETHICS COMMITTEE (MEDICAL)

CLEARANCE CERTIFICATE NO. M16111159

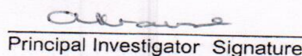


NAME: (Principal Investigator)	Prof Amanda Krause et al
DEPARTMENT:	Division of Human Genetics National Health Laboratory Services
PROJECT TITLE:	The Molecular Aetiology of Inherited Breast in South African Black Patients (Previously M110922)
DATE CONSIDERED:	Adhoc
DECISION:	Approved unconditionally
CONDITIONS:	Renewal for 5 Years Valid for the Period 01 October 2016 - 31 October 2021 Previously M110922
SUPERVISOR:	Dr Robyn Kerr
APPROVED BY:	Professor P Cleaton-Jones, Chairperson, HREC (Medical)
DATE OF APPROVAL:	20/01/2017

This clearance certificate is valid for 5 years from date of approval. Extension may be applied for.

DECLARATION OF INVESTIGATORS

To be completed in duplicate and **ONE COPY** returned to the Research Office Secretary in Room 301, Third Floor, Faculty of Health Sciences, Phillip Tobias Building, 29 Princess of Wales Terrace, Parktown, 2193, University of the Witwatersrand. I/we fully understand the conditions under which I am/we are authorized to carry out the above-mentioned research and I/we undertake to ensure compliance with these conditions. Should any departure be contemplated, from the research protocol as approved, I/we undertake to resubmit the application to the Committee. **I agree to submit a yearly progress report**. The date for annual re-certification will be one year after the date of convened meeting where the study was initially reviewed. In this case, the study was initially reviewed in November and will therefore be due in the month of November each year. Unreported changes to the application may invalidate the clearance given by the HREC (Medical).



Principal Investigator Signature

21/01/2017

Date

PLEASE QUOTE THE PROTOCOL NUMBER IN ALL ENQUIRIES

Appendix B: Consent form

National Health Laboratory Service

University of the Witwatersrand, School Of Pathology

Division Human Genetics

GENETIC COUNSELLING CLINIC

Hospital Street, Johannesburg 2001
Telephone: +27-11-489-9224/9223/9211

PO Box 1038, Johannesburg 2000
Fax: +27-11-489-9226

Prof A Christianson 011 489-9211 Prof A Krause 011 489-9219 Ms T Wessels 011 489-9243

GENETIC TESTING FOR INHERITED CANCER

CONSENT DECLARATION

Having discussed the matter fully with the genetic counselling staff at the Division of Human Genetics, Johannesburg, I wish to have a **diagnostic/predictive** test for ______________________.

I understand that the results of the test should show whether or not I have a mutation in the ______________________gene(s). I further understand that rarely the results may be uninformative or difficult to interpret, but this will be discussed with me at a genetic counselling consultation.

This is a voluntary decision. I realise that, although I have contributed a blood sample for DNA typing, I am free to withdraw from the testing at any time. However, the sample remains the property of the Division of Human Genetics, unless otherwise instructed by me.

Terms and conditions of testing:

- I accept that the policy of the Division of Human Genetics is that results will be given to me in person in a genetic counselling consultation
- I accept that I am responsible for the payment of the genetic test
- In the case of international testing, I accept that results will only be released to me once I have settled the account with the international laboratory

Regarding my DNA sample:

- It may be stored for further **DIAGNOSTIC** genetic testing related to the above condition, the results of which will be communicated to me if relevant **yes / no**
- It may be stored for further genetic testing for **RESEARCH** purposes related to the above condition, the results of which will be communicated to me, if relevant **yes / no**
- It may be stored, to be used as an **ANONAMOUS SAMPLE** for general screening If my DNA sample is used anonymously, I understand that no further information resulting from such tests will be communicated to me **yes/no**

Regarding my genetic test result:

- In order to clarify the genetic risk of the above condition in my blood relatives, my genetic test results may be shared with other family members **yes/no**

FULL NAME ....SIGNATURE..
AT………………………………………………………DATE………...

WITNESS:
FULL NAME.. SIGNATURE…….....................................

Appendix C: Recipes for laboratory solutions

0.5M Ethylenediaminetetraacetic acid (EDTA)

- 93.06 grams (g) of EDTA
- Add 500 mL of distilled water
- Adjust pH to 8 using Sodium Hydroxide (NaOH)

Tris-Borate EDTA (TBE) 10 X buffer:

- 108 g of Tris base
- 58 g of Boric acid
- 7.44 g of EDTA or 9.3g of Na_2EDTA

Make up a final volume of 1 Litre by adding distilled water

Ethidium bromide stained low electroendosmosis (LE) agarose gels:

- 0.8%: 0.8 g of agarose powder in 100millilitres (mL) of 1xTBE and 0.8µl of EtBr
- 2.0%: 2.0 g of agarose powder in 100mL of 1xTBE and 2µl of EtBr
- 3.0%: 3.0 g of agarose powder in 100mL of 1xTBE and 3µl of EtBr

Ficoll based loading dye

- 5 g of 50% sucrose
- 1 mL of 50 nM EDTA
- 0.01 g of 0.1% bromophenol blue

- 1 g of 10% Ficoll

Add distilled water to make up a final volume of 10 mL

Tris-EDTA 1 X buffer

- 10 mL of 10 nM Tris-HCl
- 2 mL of 1 mM EDTA

Make up a final volume of 1 Litre by adding distilled water

Appendix D: Protocols

1. The *BRCA2* c.5771_5774delTTCA mutation PCR protocol and conditions

PCR component	Final concentration	20 µl reaction	50 µl reaction
KapaTaq 2x Ready Mix	1 x	10 µl	25 µl
c5771_Cont_R	10 µM	0.8 µl	2 µl
c5771_Cont_F	10 µM	0.8 µl	2 µl
DNA template	100 ng/µl	2 µl	3 µl
ddH_20	-	6.4 µl	18 µl

Thermal cycling conditions

PCR step	Temperature (℃)	Time	
Initial denaturation	95	2 min	
Denaturation	95	30 sec	30 cycles
Annealing	58	30 sec	30 cycles
Elongation	72	1 min	30 cycles
Final extension	72	2 min	
Hold	12	∞	

The amplicons were electrophoresed in 2% agarose gel.

Expected PCR product size: 324 bp

2. The *BRCA2* c.582G>A mutation PCR protocol and conditions

PCR components	**Final Concentration**	**20 µl reaction**
KapaTaq 2x Ready Mix	1x	10 µl
c582_Control_R	10 µM	0.6 µl
c582_Control_F	10 µM	0.6 µl
c582_inner_F	10 µM	0.8 µl
C582_inner_R	10 µM	0.8 µl
DNA template	100 ng/µl	1 µl
ddH$_2$0	-	6.2 µl

<u>Thermal cycling conditions</u>

PCR step	**Temperature (℃)**	**Time**	
Initial denaturation	95	5 min	
Denaturation	95	30 sec	30 cycles
Annealing	58.8	40 sec	
Elongation	72	1 min	
Final extension	72	5 minutes	
Hold	15	∞	

The amplicons were electrophoresed in 2% agarose gel.

Expected PCR product sizes are 324 bp for the control amplicon, 230 bp for the mutant allele and 164 bp for the wildtype allele.

3. Microsatellite and *FANCG* 7 bp mutation genotyping protocol and conditions

PCR components	Final concentration	25 µl reaction
AmpliTaq Gold polymerase buffer	1x	2.5 µl
AmpliTaq Golg $MgCl_2$	2.5 mM	2.5 µl
dNTPs	1x	2.5 µl
Forward primer	10 µM	0.5 µl
Reverse primer	10 µM	0.5 µl
ddH_20	-	15.3 µl
AmpliTaq Gold polymerase	2.5 Units/µl	0.2 µl
DNA template	100 ng/µl	1 µl

To genotype the 21xTG microsatellite, a concentration of 100 µM for the primers was used. However, the concentrations of other reagents are as outlined in the table above.

<u>Thermal cycling conditions</u>

PCR step	Temperature (℃)	Time
Initial denaturation	94	5 min
Denaturation	94	1 min
Annealing	60	1 min
Elongation	72	1 min
Final extension	72	5 min
Hold	15	∞

30 cycles

To genotype for the *FANCG* c.637-643delTACCGCC mutation, the final extension time was 10 minutes. The other thermal cycling conditions are as outlined above.

4. SNP genotyping

TaqMan SNP genotyping

384-well plate PCR set-up. Genotyping for SNPs rs4986860 and rs1801426

PCR components	Final concentration	5 µl reaction
Probe-primer mix	40X	0.25 µl
TaqMan Genotyping Mastermix	2X	2.5 µl
DNA	10 ng/µl	2.25 µl

Thermal cycling conditions

PCR step	**Temperature (℃)**	**Time**	
Initial denaturation	95	10 min	
Denaturation	95	15 sec	40 cycles
Annealing and elongation	60	1 min	

Restriction fragment length polymorphism

Genotyping for SNP rs9590940

PCR components	**Final concentration**	**20 µl reaction**
KapaTaq Ready Mix	2x	10 µl
Primer 19_F	10 µM	0.8 µl
Primer 19_R	10 µM	0.8 µl
ddH_20	-	7.4 µl
DNA	50 ng/µl	1 µl

Thermal cycling conditions

PCR step	Temperature (℃)	Time
Initial denaturation	95	5 min
Denaturation	95	30 sec
Annealing	60	30 sec
Elongation	72	1 min
Final extension	72	2 min
Hold	12	∞

30 cycles

The PCR products were electrophoresed on 2% agarose gel.

The expected amplicon size: 334 bp

Appendix E: File extraction and input of genotypes for PHASE haplotyping software

File extraction

The software was run on a Linux operating system. PHASE is able to support executable files which need to be extracted using two command lines:

1. **gunzip phase.linux.2.1.1.tar.gz**

The command line "gunzip" is used to decompress files in "gz" format and "tar" refers to an archive or folder of files. This command therefore instructs the program to decompress and extract files from "phase.linux.2.1.1" which is in tar.gz format.

2. **tar -xvf phase.linux.2.1.1.tar**

The second line makes use of the "tar –xvf" command and instructs the program to extract files from an archive "f", which refers to the filename. The filename in this case was "phase.linux.2.1.1".

Input file format

An input file containing the genotype information of all samples to be analysed was prepared using the instructions document included in the software. The input file specifies the number of individuals genotyped, the number of loci, the genomic position (base pair units) and type of locus (SNP, microsatellite, triallelic SNP or HLA alleles), followed by the genotypes for each individual sample. After creating the input file, a basic command line such as the one below is used to run the program:

./PHASE <filename.inp><filename.out>

The mutation-positive patients and controls were analysed separately using the following command lines:

For mutation-positive samples: **./PHASE Controlsamples.inp Controlsamples.out**For control samples: **./PHASE AllsamplesandSNPs3.txt.inp AllsamplesandSNPs3.txt.out**

The software will then input the genotype data from the file, e.g. Controlsamples.inp, run the file and produce an output file i.e. Controlsamples.out, which can then be viewed and interpreted.

Appendix F: Multiplicom protocol

INSTRUCTIONS FOR USE

germline MASTR™ Plus with MID for Illumina MiSeq®

Table 6. Master reaction mix for multiplex PCR

# of Samples	PCR Mix Plex x	Taq DNA polymerase	Taq DNA polymerase 1/10 diluted*
2	20 µl	----	3 µl
4	40 µl	0.6 µl	----
24	240 µl	3.6 µl	----
40	400 µl	6 µl	----

* *Do not pipette less than 0.5 µl of Taq DNA polymerase because the solution is viscous. If less than 0.5 µl of undiluted Taq DNA polymerase is required (< 4 samples), prepare a 1/10 dilution by adding a minimum of 0.5 µl Taq DNA polymerase in molecular biology grade water, followed by vortexing (2-3 s) and centrifuging 10 s at 12,000 x g. Add the required amount to the master reaction mix.*

- Vortex briefly (2-3 s) and centrifuge each master reaction mix vial at 12,000 x g for 10 s. Vortexing at this step is essential for successful amplification.

10.1.2. PCR reaction setup for multiplex PCR

- For every DNA sample, combine 5 µl of the DNA sample (containing an equal amount of 20 to 50 ng genomic DNA; see Section 9 for more details) with 10 µl of each master reaction mix (prepared under Section 10.1.1).
- For each preparation a negative control should be included by adding 5 µl molecular biology grade water to 10 µl of each master reaction mix and close the vial or well immediately.
- Vortex briefly (2-3 s) and centrifuge at 12,000 x g for 10 s.

10.1.3. PCR cycling profile for multiplex PCR

- Run the thermal cycling profile shown in Table 7.

Table 7. PCR cycling profile for multiplex PCR

# of cycles*	Temperature	Time
1	98°C	10 min
20	95°C	45 s
	60°C	45 s
	68°C	2 min
1	72°C	10 min
1	4°C	< 1 h

* *Set the ramp rate of the PCR cycler between 1 and 2°C/s for all heating and cooling steps.*

- Do not store the amplified products for more than 1 h at 4°C (since this would induce primer dimer formation).
  Continue immediately with Sections 10.3 and 10.4.

The resulting set of MASTR Plus-derived amplicons per individual can be controlled for their quality as described below under Section 10.2.

10.2. *Quality control of multiplex PCR by agarose gel or digital alternative*

This section is optional but highly recommended

- Following thermal cycling a quality control step can be performed by:
  - Agarose gel: run 5 µl of each PCR amplified product on a 2% agarose gel containing 0.1 µg/ml ethidium bromide or equivalent (carefully follow all safety instructions applicable in the laboratory), to check for the presence of amplicons.
  - Digital alternative (nanofluidic-based method): check for the presence of amplicons according to manufacturer's instructions.
- Successful amplification is detected as a clearly visible but dispersed band or equivalent digital signal that ranges between 250 bp and 450 bp.
- The negative control should not display any amplification product for the run to be valid. Amplification in the negative control indicates cross-contamination. The complete PCR run should be discarded and the source of contamination should be investigated before attempting a new PCR run. Suitable measures should be taken to avoid reoccurrence of such a problem.

10.3. *Mixing of the multiplex reactions from a single DNA sample*

- The MASTR Plus-derived amplicons are mixed per individual DNA sample in an appropriately labeled 1.5 ml tube, applying the fixed mixing scheme for MASTR Plus resulting in an amplicon library per sample. Keep all PCR products on ice while handling.
- Vortex briefly (2-3 s) and centrifuge at 12,000 x g for 10 s.

The specific fixed mixing scheme is provided in the "MASTR Plus Assay Specification Sheet": Section 2.

Each amplicon library is subsequently purified to remove small residual DNA fragments using magnetic bead-based DNA purification as specified below in Section 10.4.

Proceed immediately with purification of the amplicon libraries.

10.4. Purification of the amplicon libraries

Purify each amplicon library using Agencourt® AMPure® XP beads using the following protocol below.

The volumes of amplicon library versus Agencourt AMPure XP and the number of required purification rounds are specified in the "MASTR Plus Assay Specification Sheet": Section 3.1, indicated as the workflow "MASTR Plus".

Remark: in case the volume of the amplicon library is less than the volume indicated, add molecular biology grade water to reach the required volume.

(1) Gently shake the Agencourt AMPure XP bottle to resuspend any magnetic particles that may have settled.
(2) Combine xx µl amplicon library with xx µl Agencourt AMPure XP.
(3) Mix reagent and amplicon library thoroughly by pipette mixing 10 times. Let this mixture incubate for 5 minutes at room temperature for maximum recovery.
(4) Place the tubes on a magnetic bead separator for 2 minutes to separate beads from the solution. Aspirate the cleared solution from the reaction plate and discard.
(5) Dispense 200 µl of freshly made 70% ethanol to each tube, and incubate for 30 s at room temperature while moving the tubes around on the magnetic bead separator. Aspirate the ethanol and discard. Repeat this wash step once more.
(6) Proceed with the appropriate steps below.

In case the MASTR Plus requires other round(s) of purification:

(7) Remove tubes from the magnetic bead separator, and add 42 µl of molecular biology grade water to each tube; pipette mix 10 times.
(8) Place the tubes on the magnetic bead separator for 1 minute to separate beads from the solution.
(9) Transfer 40 µl eluent to a new appropriately labeled tube; this is the purified amplicon library.
(10) Add xx µl Agencourt AMPure XP beads (if necessary, recalculate the recommended amount of beads specified in the Assay Specification Sheet for a sample volume of 40 µl to ensure the optimal ratio of beads per volume).
(11) Repeat steps 3-6.

In case of the final purification round:

(7) Allow the pellet to dry.
(8) Remove tubes from the magnetic bead separator, and add 20 µl of molecular biology grade water to each tube; pipette mix 10 times.
(9) Place the tubes on the magnetic bead separator for 1 minute to separate beads from the solution.
(10) Transfer the eluent to a new appropriately labeled tube; this is the purified amplicon library.

Proceed immediately with the Universal PCR as described in Section 11.1, or store the purified amplicon libraries immediately in a freezer between -15°C and -25°C.

11. TEST PROTOCOL MID FOR ILLUMINA MISEQ®

The following procedure should be strictly followed and performed by qualified personnel; any deviation from the prescribed protocol can lead to erroneous results.

11.1. *Tagging of amplicons originating from a single specimen with MIDs*

The purified MASTR Plus-derived amplicons are tagged resulting in the incorporation of MIDs and the p5 and p7 adaptors required for sequencing on the MPS system (Figure 2).

The MID kit reagents are required for tagging the MASTR Plus-derived amplicons and are provided as separate MID kits for sequencing on the Illumina MiSeq MPS system, each containing unique MID sequences.

For the complete list of available MID for Illumina MiSeq kits, visit the Section "Complementary MASTR Plus products" on the website: http://www.multiplicom.com.

The precise sequences of the MIDs contained in MID for Illumina MiSeq kits (IFU243) can be obtained from http://www.multiplicom.com/keycode using the KEY-CODE printed on the box label of the used MID kit.

11.1.1. *Preparation of Universal master reaction mixes*

For each sample (including the MASTR Plus negative control sample) to be sequenced in the same sequencing run on an MPS system, a fresh master mix containing a unique combination of MID p7-p5 Primer Mixes should be freshly prepared as specified below.

Remarks:

(1) To avoid issues with low nucleotide diversity in the Illumina MiSeq run, use the Support tools for MID selection.

(2) If aiming to detect variants with an allele frequency of 10% or lower, run-to-run sample carryover can be a confounding factor. To prevent this, it is recommended to avoid sequencing the same MASTR Plus assay with the same MID combination in two consecutive runs.

The Support tools for MID selection (IFU142, IFU143 and IFU240) can be obtained from http://www.multiplicom.com/keycode using the KEY-CODE printed on the box label of the used MID kit.

- Remove the required Universal PCR Mix, Amplification Reagent (AR) and MID p7 and p5 Primer Mixes from the freezer and allow complete thawing on ice. The obligatory AR is included in the MASTR Plus kit.
  Remark: Make sure to use the AR provided with the specific MASTR Plus kit in use.
- After thawing, vortex each vial thoroughly and centrifuge at 12,000 x g for 10 s before use.
- Remove the Taq DNA polymerase from the freezer and put on ice. Centrifuge at 12,000 x g for 10 s before use.
  For each MID combination, label a new PCR tube and mix the necessary reagents according to the scheme exemplified in Table 8 (required volumes are indicated for each reaction).
- Vortex briefly (2-3 s) and centrifuge each Universal master reaction mix vial at 12,000 x g for 10 s. Vortexing at this step is essential for successful amplification.

Table 8. Master reaction mix for Universal PCR

Reagents	Volume
Universal PCR Mix	20 µl
Amplification Reagent	20 µl
MID p7 Primer Mix	4 µl
MID p5 Primer Mix	4 µl
Taq DNA polymerase 1/10 diluted*	2.5 µl

* *Do not pipette less than 0.5 µl of Taq DNA polymerase because the solution is viscous: prepare a 1/10 dilution by adding a minimum of 0.5 µl Taq DNA polymerase in molecular biology grade water, followed by vortexing (2-3 s) and centrifuging 10 s at 12,000 x g. Add the required amount to the master reaction mix.*

11.1.2. PCR reaction setup for Universal PCR

- When this Universal PCR does not follow immediately on the execution of the specific MASTR Plus (and purified amplicon libraries were stored between -15°C and -25°C), remove the purified amplicon libraries from the freezer, allow complete thawing on ice, vortex briefly (2-3 s), centrifuge at 12,000 x g for 10 s and proceed immediately.
- Dilute the purified amplicon library from each sample (including the MASTR Plus negative control sample), using a two-step serial dilution in molecular biology grade water.

The specific dilution factor is provided in the "MASTR Plus Assay Specification Sheet": Section 4.

- Add 2 µl of this dilution to 48 µl of the fresh Universal master reaction mix, and close the vial or well immediately. Vortex briefly (2-3 s) and centrifuge at 12,000 x g for 10 s.

11.1.3. PCR cycling profile for Universal PCR

- Run the thermal cycling profile shown in Table 9.

Table 9. PCR cycling profile for Universal PCR.

# of cycles*	Temperature	Time
1	98°C	10 min
20	95°C	45 s
	64°C	45 s
	68°C	2 min
1	72°C	10 min
1	4°C	Max. 12 h

* *Set the ramp rate of the PCR cycler between 1 and 2°C/s for all heating and cooling steps.*

- Do not store the amplified products for more than 12 h at 4°C before proceeding with the workflow. Continue immediately with Section 11.4, otherwise immediately store amplified products between -15°C and -25°C.

The resulting tagged amplicon library per individual can be controlled for its quality as described below under Sections 11.2 and 11.3. Multiplicom recommends performing at least one of these quality control steps (preferentially Section 11.3).

11.2. *Quality control of Universal PCR by agarose gel or digital alternative*

This section is optional, but highly recommended

- Following thermal cycling, a quality control step can be performed by:
  - Agarose gel: run 5 µl of each amplified sample on a 2% agarose gel containing 0.1 µg/ml ethidium bromide or equivalent (carefully follow all safety instructions applicable in the laboratory), to check for the presence of amplicons.
  - Digital alternative (nanofluidic-based method): check for the presence of amplicons according to manufacturer's instructions.
- Successful amplification of the Universal PCR is detected as a clearly visible but dispersed band or equivalent digital signal between 350 and 550 bp.
- The negative control should not display any amplification product for the run to be valid. Amplification in the negative control indicates cross-contamination. The complete PCR run should be discarded and the source of contamination should be investigated before attempting a new PCR run. Suitable measures should be taken to avoid reoccurrence of such a problem.

11.3. *Quality control of Universal PCR by fluorescent labeling and fragment analysis*

This section is optional, but highly recommended

The quality of the Universal PCR can be reviewed in detail, based on the pattern of the amplicons of a tagged amplicon library of a specific MASTR Plus. The required chromatograms can be obtained using an ABI capillary sequencer with GeneScan (GS) module.

11.3.1. *Preparation of GS-labeling master reaction mix*

- Remove the GS-labeling PCR Mix from the freezer and allow complete thawing on ice.
- After thawing, vortex each vial thoroughly and centrifuge the vials at 12,000 x g for 10 s before use.
- Remove the Taq DNA polymerase from the freezer and put on ice. Centrifuge at 12,000 x g for 10 s before use.
- Prepare a GS-labeling master reaction mix for all samples that will be labeled with volumes per reaction as indicated in Table 10.

Table 10. Master reaction mix for GS-labeling PCR

Reagents	Volume
GS-labeling PCR Mix	13 µl
Taq DNA Polymerase*	0.075 µl

* *Do not pipette less than 0.5 µl of Taq DNA polymerase because the solution is viscous. If less than 0.5 µl of undiluted Taq DNA polymerase is required (< 7 reactions), prepare a 1/10 dilution by adding a minimum of 0.5 µl of Taq DNA polymerase in molecular biology grade water, followed by vortexing and centrifuging 10 s at 12,000 x g. Add the required amount to the GS-labeling PCR Mix.*

- Vortex briefly (2-3 s) and centrifuge the vial at 12,000 x g for 10 s.

11.3.2. PCR reaction setup for GS-labeling PCR

- For each Universal PCR reaction that will be analyzed, combine 2 µl of the Universal PCR product (obtained after completion of Section 11.1.3) with 13 µl of the freshly prepared GS-labeling master reaction mix, and close the vial or well immediately.
- Vortex briefly (2-3 s) and centrifuge at 12,000 x g for 10 s.

11.3.3. PCR cycling profile for GS-labeling PCR

- Run the thermal cycling profile shown in Table 11.

Table 11. PCR cycling profile for GS-labeling PCR

# of cycles*	Temperature	Time
1	98°C	10 min
5	95°C	30 s
	65°C	30 s
	68°C	30 s
1	72°C	10 min
1	4°C	∞

* *Set the ramp rate of the PCR cycler between 1 and 2°C/s for all heating and cooling steps.*

11.3.4. Fragment analysis

- Prepare a size standard mix for all reactions using the volumes per reaction indicated in Table 12.

Table 12. Preparation of GS-labeled amplicons for fragment analysis

Reagents	Volume (x no. of samples)
Hi-Di™ Formamide	10 µl
GS600 size standard	0.3 µl

- For each reaction dispense 10 µl of size standard mix per well into a 96-well plate.
- Add 2 µl of the GS labeling PCR-product to a well containing the "Size standard mix".
- Centrifuge the plate at 1,000 x g for 10 s.
- Denature at 95°C for 3 minutes and immediately put on ice.
- Run on the fragment analyzer according to manufacturer's instructions.
- Review the generated data (.fsa files) with MAQ-S software using the required Assay Description file (.enc file) and GS Reference Pattern. The Universal PCR amplification is valid if the analysis pattern is consistent with the GS Reference Pattern.

The MAQ-S software, Assay Description file and GS Reference Pattern can be obtained from http://www.multiplicom.com/keycode using the KEY-CODE located on the box label of the used MASTR Plus assay.

If a valid analysis pattern is obtained, continue immediately with the Section 11.4, otherwise immediately store amplified products between -15°C and -25°C.

If the analysis is not valid, carefully review the assay protocol for deviations. If the analysis is persistently non-valid, contact customer service: customerservice@multiplicom.com.

11.4. Purification of the tagged amplicon libraries

Purify each tagged amplicon library using Agencourt® AMPure® XP beads using the protocol below.

> *The volumes of amplicon library versus Agencourt AMPure XP and the number of required purification rounds are specified in the "MASTR Plus Assay Specification Sheet": Section 3.2, indicated as the workflow "MID for Illumina MiSeq".*

(1) Gently shake the Agencourt AMPure XP bottle to resuspend any magnetic particles that may have settled.
(2) Combine xx µl tagged amplicon library with xx µl Agencourt AMPure XP.
(3) Mix reagent and tagged amplicon library thoroughly by pipette mixing 10 times. Let this mixture incubate for 5 minutes at room temperature for maximum recovery.
(4) Place the tubes on a magnetic bead separator for 2 minutes to separate beads from the solution. Aspirate the cleared solution from the reaction plate and discard.
(5) Dispense 200 µl of freshly made 70% ethanol to each tube, and incubate for 30 s at room temperature while moving the tubes around on the magnetic bead separator. Aspirate the ethanol and discard. Repeat this wash step once more.
(6) Proceed with the appropriate steps below.

In case the MASTR Plus requires other round(s) of purification:

(7) Remove tubes from the magnetic bead separator, and add 42 µl of molecular biology grade water to each tube; pipette mix 10 times.
(8) Place the tubes on the magnetic bead separator for 1 minute to separate beads from the solution.
(9) Transfer 40 µl eluent to a new appropriately labeled tube; this is the purified tagged amplicon library.
(10) Add xx µl Agencourt AMPure XP beads (if necessary, recalculate the recommended amount of beads specified in the Assay Specification Sheet for a sample volume of 40 µl to ensure the optimal ratio of beads per volume).
(11) Repeat steps 3-6.

In case of the final purification round:

(7) Allow the pellet to dry.
(8) Remove tubes from the magnetic bead separator, and add 20 µl of molecular biology grade water to each tube; pipette mix 10 times.
(9) Place the tubes on the magnetic bead separator for 1 minute to separate beads from the solution.
(10) Transfer the eluent to a new appropriately labeled tube; this is the purified tagged amplicon library.

Proceed immediately with Section 11.5, or store the purified tagged amplicon libraries immediately in a freezer between -15°C and -25°C.

11.5. Pooling of different purified tagged amplicon libraries

- Determine the concentration of the obtained purified tagged amplicon libraries by either:
  - a spectrophotometric method,
  - a fluorometric method,
  - qPCR (eg, KAPA Library Quant Kits, KAPA Biosystems).

  Remark: when AR3 is used for the Universal PCR, only a spectrophotometric method can be used.

- Calculate the concentration of each purified tagged amplicon library in nM, using the following equation:

$$Sample\ conc\ [nM] = \frac{sample\ conc\ [ng/\mu l] \times 10^6}{656.6 \times average\ amplicon\ size}$$

The average amplicon size after the Universal PCR is specified in the "MASTR Plus Assay Specification Sheet": Section 5.

- Dilute each purified tagged amplicon library to 10 nM in 1 x TE buffer, vortex briefly (2-3 s) and centrifuge at 12,000 x g for 10 s.
- If multiple amplicon libraries need to be sequenced on an MPS system, pool equal volumes for samples requiring the same minimal coverage (see Remarks 2 and 3). This results in an amplicon pool with a required number of total read pairs that can be calculated by multiplying the number of pooled samples with the minimal coverage per sample.
- Vortex briefly (2-3 s) and centrifuge at 12,000 x g for 10 s.
- In case of combining > 1 amplicon pools, combine those by using volumes that are directly related to the calculated number of read pairs for each of these amplicon pools, eg, an amplicon pool which requires 24,000 read pairs will need a volume that is twice as high as an amplicon pool requiring 12,000 read pairs.
- Vortex briefly (2-3 s) and centrifuge at 12,000 x g for 10 s.

  Remarks:

  (1) Take care to pool only purified tagged amplicon libraries that contain different MID combinations (incorporated in the Universal PCR step of the MID for Illumina MiSeq kit).

  (2) For SNV analysis, it is highly recommended to obtain a certain minimal coverage (number of read pairs) per individual sample, corresponding to a minimal coverage per affected allele of 50 read pairs. Nevertheless, higher coverage may be required to detect low frequency variants.

  (3) For CNV analysis, it is highly recommended to obtain a certain minimal coverage (number of read pairs) per individual sample, corresponding to a minimal coverage per amplicon of 300 read pairs.

  (4) For the calculation of the number of samples per run, take into account that the PhiX control library added to the sample already uses part of the capacity of the run.

The minimal number of read pairs per sample is specified in the "MASTR Plus Assay Specification Sheet": Section 5.

The required percentage of the PhiX control library is specified in the "MASTR Plus Assay Specification Sheet": Section 6.

The number of samples per run can be calculated by using the Calculator at http://www.multiplicom.com/calculator.

Proceed immediately with Section 12, otherwise immediately store the amplicon pool in a freezer between -15°C and -25°C.

12. BRIDGE AMPLIFICATION AND ILLUMINA MISEQ SEQUENCING

Run-to-run sample carryover can be a confounding factor increasing the noise of the sequencing run. To reduce carryover, it is recommended to perform a maintenance wash prior to sequencing. The wash tray should be thoroughly cleaned after each use to eliminate any risk of template molecules remaining in the tray. Additionally, it is recommended to avoid sequencing the same MASTR Plus assay with the same MID combination in two consecutive runs. When still suffering from carryover effects, contact Illumina Technical Support for further assistance.

The amplicon pool is further processed using one of the Reagent Kits provided by Illumina Inc., containing all necessary consumables and a reagent cartridge for sequencing the purified tagged amplicon library on an Illumina MiSeq MPS system following the guidelines described below.

A detailed description on how to set up an Illumina MiSeq run can be found in the Illumina MiSeq system user guide. Make sure always to use the latest version of the Illumina MiSeq user guide.

Remark: When not proceeding immediately with Section 12 after completing Section 11.5, redetermine and recalculate the concentration of the sequencing sample as described in that section.

- Choose the applicable Illumina MiSeq Reagent Kit: Kit v2 or Nano Kit v2 (500 cycles, 2 x 251 bp; MS-102-2003 or MS-103-1003 respectively), or Kit v3 (600 cycles, 2 x 301 bp; MS-102-3003).

The number of cycles for Read1 and Read2 is specified in the "MASTR Plus Assay Specification Sheet": Section 6.

xxx ≤ R1/R2 ≤ yyy, in which R1=Read1 and R2=Read2
% bases sequenced bidirectional, for v2 and v3

Remark (explanation on equation above):

- *xxx* indicates the minimal number of cycles for a specific MASTR Plus to ensure that every base is sequenced (no overlap between R1 and R2).
- *yyy* indicates the theoretical number of cycles required for complete bidirectional sequencing of all amplicons (full overlap between R1 and R2).
- Due to the limited number of cycles of the current Illumina Reagent kits, for most germline MASTR Plus assays the theoretical *yyy* cannot be reached. Therefore, the theoretical percentage of bases sequenced bidirectional is given for the v2 and v3 Reagent kits respectively.
  Note that the following bases were not taken into account: *(1)* bases of the target specific primers, and *(2)* bases of control amplicons (when present).

Appendix G: Assay specification sheet

MASTR™ Plus Assay Specification Sheet

BRCA Hereditary Cancer MASTR Plus

IFU405
v150103

1. ASSAY SPECIFIC INFORMATION

Number of PCR Mixes	5
Number of amplicons	561
Amplification Reagent	3

2. ASSAY SPECIFIC MIXING SCHEME

No. Plex	Volume	No. Plex	Volume
Plex 1	10.5 µl	Plex 4	6 µl
Plex 2	4 µl	Plex 5	6 µl
Plex 3	3 µl		

3. RATIO AMPLICON LIBRARY VERSUS AGENCOURT AMPURE BEADS

	Workflow	Purification Ratio			Number of purifications
		Amplicon Library	Agencourt® AMPure® XP	Ratio "AMPure per library"	
3.1	MASTR™ Plus	20 µl	24 µl	1/1.2	1
3.2	MID for Illumina MiSeq®	30 µl	25.5 µl	1/0.85	1
	Short Read Amplification	30 µl	36 µl	1/1.2	1

4. DILUTION PURIFIED MASTR-DERIVED LIBRARY FOR UNIVERSAL PCR

1/1000

5. AVERAGE AMPLICON SIZE AFTER UNIVERSAL PCR & MINIMAL COVERAGE PER SAMPLE

Workflow	Avg. amplicon size	Min. coverage per sample SNV analysis	Min. coverage per sample CNV analysis
MID for Illumina MiSeq®	473 bp	561 000 read pairs	1 683 000 read pairs
Short Read Amplification*	376 bp	1 055 000 reads	3 164 500 reads

* *For use with Ion PGM Sequencing 200 Kit v2 (Life technologies)*

6. ADDITIONAL INFORMATION FOR MISEQ SEQUENCING

% PhiX Control Library		5 %
No. Cycles Read1 and Read2		196 ≤ R1/R2 ≤ 369
% bases sequenced bidirectionally	v2	65.23 %
	v3	88.46 %

7. REFERENCES

All additional MASTR assay specific files needed for analysis of the data can be downloaded from http://www.multiplicom.com/keycode where you will be requested to enter the product KEY-CODE found on the box label and box insert.

FOR SPECIFIC GUIDELINES ON ANALYSIS OF PMS2 EXONS 11-15 AND CHEK2 EXON 11, REFER TO TECHNICAL GUIDE IFU476

www.multiplicom.com
Revision date: February 23, 2015

© 2015 Multiplicom NV, all rights reserved.

Page 1 of 2

Appendix H: Genotyping results for the SNP and microsatellite screening

Genotype data for the control samples

	Marker 17xAT		SNP rs4986860		SNP rs9590940		SNP rs1801426		Marker 21xTG	
Sample	**Allele 1**	**Allele 2**	**A**	**G**	**A**	**C**	**A**	**G**	**Allele 1**	**Allele 2**
1	14	14	A	A	A	A	A	A	20	21
2	10	10	A	A	A	A	A	A	19	24
3	11	14	A	A	A	A	A	A	21	24
4	9	11	A	A	A	A	A	A	19	22
5	10	21	A	A	A	A	A	A	20	20
6	13	13	A	A	A	A	A	A	19	22
7	10	10	A	A	A	A	A	G	21	21
8	10	12	A	A	A	A	A	G	23	24
9	10	23	A	A	A	A	A	G	19	24
10	3	9	A	A	A	A	A	A	17	19
11	3	15	A	A	A	A	A	A	21	22
12	12	12	A	A	A	A	A	A	18	21
13	13	13	A	A	A	A	A	A	14	22
14	3	3	A	A	A	C	A	G	17	22
15	21	21	A	A	A	C	A	A	20	20
16	3	3	A	A	A	A	A	A	17	17
17	3	3	A	A	A	A	A	G	22	24
18	10	10	A	A	A	A	A	A	18	24
19	11	11	A	A	A	A	A	A	17	17
20	12	22	A	A	A	A	A	G	20	20
21	10	12	A	G	A	A	A	A	18	24
22	11	26	A	A	A	A	A	A	17	20
23	10	11	A	A	A	A	A	A	17	17
24	14	14	A	A	A	A	A	A	20	20
25	19	19	A	G	A	C	A	G	20	24
26	4	11	A	G	A	A	A	A	17	17

	Marker 17xAT		SNP rs4986860		SNP rs9590940		SNP rs1801426		Marker 21xTG	
Sample	**Allele 1**	**Allele 2**	**A**	**G**	**A**	**C**	**A**	**G**	**Allele 1**	**Allele 2**
27	9	15	A	A	A	A	A	A	21	21
28	15	15	A	A	A	A	A	A	22	24
29	15	15	A	A	A	A	A	A	21	21
30	8	12	A	A	A	A	A	A	14	22
31	11	11	A	A	A	A	A	A	17	17
32	12	12	A	A	A	A	A	A	21	24
33	14	26	A	A	A	A	A	A	18	21
34	11	13	A	A	A	A	A	A	18	18
35	14	22	A	A	A	A	A	G	18	22
36	10	10	A	A	A	A	A	A	17	23
37	8	11	A	A	A	A	A	A	18	18
38	10	10	A	A	A	A	A	A	24	24
39	20	28	A	A	A	A	A	A	20	20
40	12	25	A	A	A	A	A	A	21	25
41	10	13	A	A	A	A	A	A	20	20
42	11	32	A	A	A	A	A	A	17	17
43	18	18	A	A	A	A	A	A	22	22
44	12	14	A	A	A	A	A	A	18	21
45	4	4	A	A	A	A	A	A	20	20
46	4	14	A	A	A	A	A	A	14	18
47	9	12	A	A	A	A	A	G	21	24
48	20	20	A	A	A	A	A	A	24	24
49	25	25	A	A	A	A	A	A	18	25
50	8	11	A	A	A	A	A	A	20	20
51	3	3	A	A	A	A	A	A	18	18
52	11	22	A	A	A	A	A	G	17	20
53	12	15	A	A	A	A	A	A	21	21
54	14	27	A	A	A	A	A	A	17	17
55	12	12	A	G	A	A	A	A	17	24
56	22	22	A	A	A	A	A	G	14	20

Sample	Marker 17xAT		SNP rs4986860		SNP rs9590940		SNP rs1801426		Marker 21xTG	
	Allele 1	Allele 2	A	G	A	C	A	G	Allele 1	Allele 2
57	11	14	A	A	A	A	A	A	17	20
58	10	29	A	A	A	A	A	A	17	17
59	12	12	A	A	A	A	A	A	21	21
60	10	10	A	A	A	A	A	A	18	25
61	12	12	A	A	A	A	A	A	17	17
62	11	12	A	A	A	A	A	G	17	22
63	10	24	A	A	A	A	A	G	21	21
64	12	15	A	A	A	A	A	A	22	22
65	2	10	A	A	A	A	A	A	22	24
66	10	12	A	A	A	A	A	G	20	20
67	12	26	A	G	A	A	A	A	20	24
68	11	21	A	A	A	A	A	A	20	20
69	11	17	A	A	A	A	A	A	19	19
70	13	13	A	A	A	A	A	G	19	19
71	11	14	A	A	A	A	A	A	18	18
72	10	22	A	A	A	A	A	G	18	22
73	11	14	A	A	A	A	A	A	19	19
74	10	19	A	A	A	A	A	A	18	24
75	11	11	A	A	A	A	A	A	17	19
76	3	13	A	A	A	A	A	A	17	17
77	13	13	A	A	A	A	A	A	18	18
78	11	26	A	A	A	A	A	A	17	17
79	21	21	A	A	A	A	A	A	18	18
80	11	11	A	A	A	A	A	A	18	20
81	12	12	A	A	A	A	A	G	20	20
82	11	11	A	A	A	C	A	A	19	22
83	11	24	A	A	A	A	A	A	17	20
84	22	23	A	A	A	A	A	G	17	22
85	3	10	A	A	A	A	A	A	17	22
86	3	22	A	A	A	A	A	G	17	19

	Marker 17xAT		SNP rs4986860		SNP rs9590940		SNP rs1801426		Marker 21xTG	
Sample	**Allele 1**	**Allele 2**	**A**	**G**	**A**	**C**	**A**	**G**	**Allele 1**	**Allele 2**
87	13	13	A	A	A	A	A	A	19	22
88	19	28	A	A	A	A	A	A	18	18
89	11	11	A	A	A	A	A	A	20	22
90	14	14	A	A	A	A	A	G	19	28
91	11	22	A	A	A	C	A	G	22	22
92	11	16	A	A	A	A	A	A	18	18
93	3	22	A	A	A	A	A	G	17	19
94	10	14	A	A	A	A	A	A	18	24
95	11	12	A	A	A	A	A	G	17	21
96	27	27	A	A	A	A	A	A	19	19
97	12	12	A	A	A	A	A	A	19	19
98	9	11	A	A	A	A	A	A	17	17
99	9	9	A	A	A	A	A	A	19	24
100	12	20	A	A	A	A	A	A	21	24

Genotype results for the patient samples

	Marker 17xAT		SNP rs4986860		SNP rs9590940		SNP rs1801426		Marker 21xTG	
Sample	**Allele 1**	**Allele 2**	**A**	**G**	**A**	**C**	**A**	**G**	**Allele 1**	**Allele 2**
Patient 1	11	11	A	A	A	A	A	A	21	21
Patient 2	12	22	A	A	A	A	A	A	19	19
Patient 3	10	10	A	A	A	A	A	A	21	21
Patient 4	12	12	A	A	A	A	A	A	18	21
Patient 5	10	11	A	A	A	A	A	A	17	21

Appendix I: Haplotypes generated by PHASE software

Haplotypes constructed for the patient samples

	Haplotype	Number of alleles
1	10 AAA 21	3
2	11 AAA 17	1
3	11 AAA 21	2
4	12 AAA 18	1
5	12 AAA 19	1
6	12 AAA 21	1
7	22 AAA 19	1

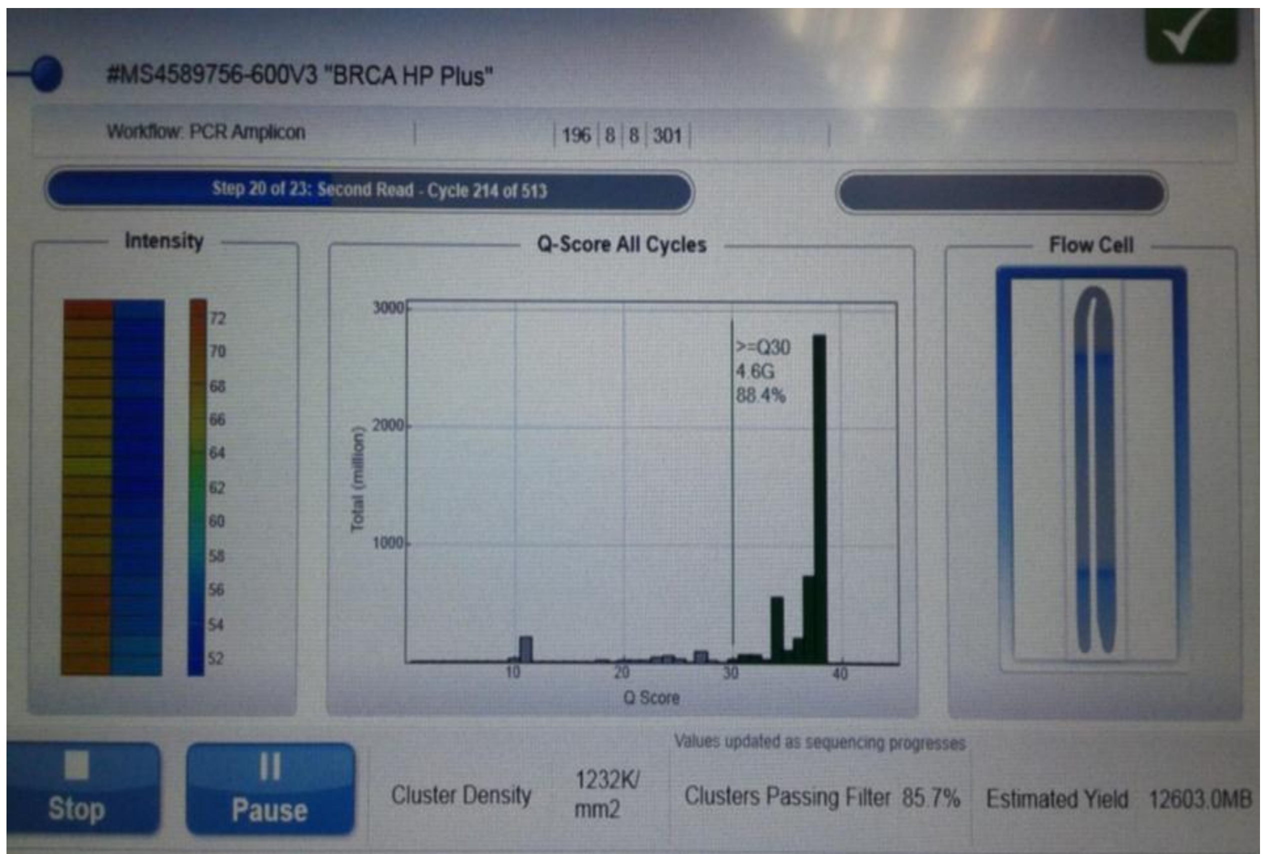
Haplotypes constructed for the control samples

	Haplotype	Number of alleles
1	2 AAA 22	1
2	3 AAA 17	7
3	3 AAA 18	2
4	3 AAA 22	3
5	3 AAG 24	1
6	3 ACG 22	1
7	4 AAA 18	1
8	4 AAA 20	2
9	4 ACA 17	1
10	8 AAA 18	1
11	8 AAA 20	1
12	8 AAA 22	1
13	9 AAA 17	1
14	9 AAA 19	3
15	9 AAA 21	1
16	9 AAA 24	2
17	10 AAA 17	4
18	10 AAA 18	4
19	10 AAA 19	1
20	10 AAA 20	3
21	**10 AAA 21**	**1**
22	10 AAA 23	1
23	10 AAA 24	9
24	10 AAA 25	1
25	10 AAG 21	2
26	**11 AAA 17**	**17**
27	11 AAA 18	5
28	11 AAA 19	3
29	11 AAA 20	4
30	11 AAA 22	2

31	11 AAA 24	1
32	11 ACA 17	1
33	11 ACA 22	1
34	12 AAA 14	1
35	12 AAA 17	3
36	**12 AAA 18**	**1**
37	**12 AAA 19**	**2**
38	12 AAA 20	2
39	**12 AAA 21**	**8**
40	12 AAA 22	1
41	12 AAA 24	1
42	12 AAG 20	2
43	12 AAG 21	2
44	12 AAG 22	1
45	12 AAG 23	1
46	12 GAA 24	3
47	13 AAA 14	1
48	13 AAA 17	1
49	13 AAA 18	3
50	13 AAA 19	3
51	13 AAA 20	1
52	13 AAA 22	3
53	13 AAG 19	1
54	14 AAA 14	1
55	14 AAA 17	1
56	14 AAA 18	4
57	14 AAA 19	2
58	14 AAA 20	4
59	14 AAA 21	3
60	14 AAG 28	1
61	15 AAA 21	5
62	15 AAA 22	2
63	15 AAA 24	1

64	16 AAA 18	1
65	17 AAA 19	1
66	18 AAA 22	2
67	19 AAA 18	2
68	19 AAA 24	1
69	19 GCG 20	1
70	20 AAA 20	1
71	20 AAA 24	3
72	21 AAA 18	2
73	21 AAA 20	3
74	21 ACA 20	1
75	22 AAA 14	1
76	22 AAG 19	2
77	22 AAG 20	3
78	22 AAG 22	4
79	23 AAA 17	1
80	23 AAG 19	1
81	24 AAA 20	1
82	24 AAA 21	1
83	25 AAA 18	1
84	25 AAA 25	2
85	26 AAA 17	1
86	26 AAA 18	1
87	26 AAA 20	2
88	27 AAA 17	1
89	27 AAA 19	2
90	28 AAA 18	1
91	28 AAA 20	1
92	29 AAA 17	1
93	32 AAA 17	1

<u>**Appendix J: Quality metrics obtained from the Illumina MiSeq instrument**</u>



Appendix K: Benign and likely benign variants

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.103C>A	*ATM*	p.Arg35Arg	Silent	rs55861249	-	-	0.0000743	0.0008	Likely benign
c.1176C>G	*ATM*	p.Gly392Gly	Silent	rs1800727	0.019	0.07	0.0045	0.0496	Likely benign
c.1236-18insT	*ATM*	-	Intronic	rs587781356	-	-	0.0077	0.0076	Benign
c.1541G>A	*ATM*	p.Gly514Asp	Missense	rs2235000	0.018	0.064	0.0053	0.0591	Likely benign
c.1636C>G	*ATM*	p.Leu546Val	Missense	rs2227924	0.0184	0.066	0.00509	0.0558	Likely benign
c.186-7C>T	*ATM*	-	Intronic	rs55674039	0.0034	0.011	0.0018	0.0189	Benign
c.2193C>T	*ATM*	p.Tyr731Tyr	Silent	rs2229019	0.018	0.066	0.0051	0.056	Likely benign
c.2614C>T	*ATM*	p.Pro872Ser	Missense	rs3218673	0.016	0.058	0.0046	0.0511	Likely benign
c.2639-17G>T	*ATM*	-	Intronic	rs2234994	0.032	0.069	0.0144	0.0678	Benign
c.2685A>G	*ATM*	p.Leu895Leu	Silent	rs3218687	0.016	0.058	0.0046	0.051	Benign
c.2922-21T>G	*ATM*	-	Intronic	rs149096247	0.011	0.04	0.0767	0.22	Benign
c.3383A>G	*ATM*	p.Gln1128Arg	Missense	rs2229020	0.0076	0.028	0.0022	0.0237	Likely benign
c.3403-15T>A	*ATM*	-	Intronic	rs79701258	0.035	0.11	0.0123	0.0965	Likely benign
c.3747-34A>G	*ATM*	-	Intronic	rs3092840	0.01	0.037	0.0042	0.034	Likely benign
c.378T>A	*ATM*	p.Asp126Glu	Silent	rs2234997	0.062	0.22	0.019	0.1988	Likely benign
c.3993+133A>G	*ATM*	-	Intronic	rs142220799	0.012	0.042	-	-	Likely benign
c.3994-121delC	*ATM*	-	Intronic	rs56013141	-	-	-	-	Likely benign
c.3994-97G>A	*ATM*	-	Intronic	rs4988000	0.028	0.1	-	-	Benign
c.4138C>T	*ATM*	p.His1380Tyr	Missense	rs3092856	0.032	0.069	0.0145	0.067	Likely benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.4777-20A>G	*ATM*	-	Intronic	rs3218678	0.062	0.22	0.0189	0.1988	Benign
c.5557G>A	*ATM*	p.Asp1853Asn	Missense	rs1801516	0.067	0.0083	0.1101	0.0249	Likely benign
c.5948G>A	*ATM*	p.Ser1983Asn	Missense	rs659243	1	1	1	1	Benign
c.6095+15T>C	*ATM*	-	Intronic	rs3212321	0.016	0.058	0.0046	0.0511	Likely benign
c.6235G>A	*ATM*	p.Val2079Ile	Missense	rs1800060	0.0028	0.0076	0.0026	0.018	Benign
c.6437G>C	*ATM*	p.Ser2146Thr	Missense	rs56815840	0.0024	0.0076	0.0013	0.0129	Benign
c.6572+12G>T	*ATM*	-	Intronic	rs3218677	0.016	0.058	0.0044	0.0487	Likely benign
c.657T>C	*ATM*	p.Cys219Cys	Silent	rs2235003	0.028	0.1	0.0083	0.0926	Likely benign
c.6995T>C	*ATM*	p.Leu2332Pro	Missense	rs4988111	0.0072	0.026	0.0022	0.0246	Benign
c.7630-17T>C	*ATM*	-	Intronic	rs116047570	0.0046	0.017	0.0009	0.0088	Likely benign
c.8269-14A>T	*ATM*	-	Intronic	rs114320959	0.0026	0.0091	0.0007	0.0076	Benign
c.G279G>A	*ATM*	p.Lys93Lys	Silent	rs368196317	-	-	0.00000851	9.99E-05	Likely benign
c.1053G>C	*BARD1*	p.Thr351Thr	Silent	rs2070096	0.1903	0.16	0.2051	0.1659	Likely benign
c.1075_1095del21	*BARD1*	-	Nonframeshift substitution	rs28997575	0.0447	-	0.0293	0.0418	Benign
c.1134G>C	*BARD1*	p.Arg378Ser	Missense	rs2229571	0.46	0.38	0.5487	0.397	Likely benign
c.1315-19G>A	*BARD1*	-	Intronic	rs6704780	0.3652	0.36	0.3818	0.3759	Benign
c.1518T>C	*BARD1*	p.His506His	Silent	rs2070093	0.77	0.55	0.8134	0.5989	Likely benign
c.1568+14C>T	*BARD1*	-	Intronic	rs5031011	0.3522	0.33	0.3614	0.3445	Benign
c.1738G>A	*BARD1*	p.Glu580Lys	Missense	rs35306212	0.0052	0.02	0.0016	0.0154	Benign
c.1933T>C	*BARD1*	p.Cys645Arg	Missense	rs2228456	0.025	0.088	0.0064	0.0693	Benign
c.216-14delT	*BARD1*	-	Intronic	rs56130510	-	-	0.0063	0.0047	Likely benign
c.-30G>C	*BARD1*	-	5'UTR variant	rs1129804	0.74	0.76	0.7072	0.7778	Likely benign
c.70C>T	*BARD1*	p.Pro24Ser	Missense	rs1048108	0.3313	0.19	0.4377	0.283	Likely benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.1122T>C	*BLM*	p.His374His	Silent	rs28385009	0.011	-	0.0079	0.00009653	Benign
c.2308-50G>A	*BLM*	-	Intronic	rs17273206	0.1609	0.19	0.1553	0.1724	Benign
c.2555+7T>C	*BLM*	-	Intronic	rs3815003	0.37	0.67	0.2956	0.6001	Benign
c.2603C>T	*BLM*	p.Pro868Leu	Missense	rs2227935	0.051	0.075	0.0559	0.0684	Likely benign
c.3102G>A	*BLM*	p.Tyr1034Tyr	Silent	rs2227933	0.15	0.2	0.1611	0.1882	Likely benign
c.3358+32T>G	*BLM*	-	Intronic	rs17273842	0.1448	0.17	0.15891	0.1628	Benign
c.3961G>A	*BLM*	p.Val1321Ile	Missense	rs7167216	0.0671	0.095	0.06634	0.0923	Likely benign
c.410A>G	*BLM*	p.Lys137Arg	Missense	rs28384988	0.0044	0.015	0.0013	0.0139	Benign
c.419A>G	*BLM*	p.Glu140Gly	Missense	rs35886055	0.0102	0.036	0.0029	0.0317	Benign
c.615G>A	*BLM*	p.Lys205Lys	Silent	rs28903082	0.0056	0.02	0.0016	0.0179	Benign
c.-19-115T>C	*BRCA1*	-	Intronic	rs3765640	0.35	0.22	-	-	Benign
c.1971A>G	*BRCA1*	p.Gln657Gln	Silent	rs28897679	0.0064	0.023	0.0022	0.0237	Benign
c.-20+101C>G	*BRCA1*	-	Intronic	rs799905	0.55	0.89	0.4805	0.7789	Benign
c.2082C>T	*BRCA1*	p.Ser694Ser	Silent	rs1799949	0.34	0.21	0.3483	0.2364	Benign
c.213-161A>G	*BRCA1*	-	Intronic	rs799912	0.55	0.9	-	-	Benign
c.2311T>C	*BRCA1*	p.Leu771Leu	Silent	rs16940	0.34	0.16	0.342	0.1792	Benign
c.2612C>T	*BRCA1*	p.Pro871Leu	Missense	rs799917	0.54	0.89	0.41	0.8193	Benign
c.3113A>G	*BRCA1*	p.Glu1038Gly	Missense	rs16941	0.34	0.16	0.3429	0.1797	Benign
c.3548A>G	*BRCA1*	p.Lys1183Arg	Missense	rs16942	0.35	0.22	0.349	0.2358	Benign
c.3804T>C	*BRCA1*	p.Asn1268Asn	Silent	rs140588714	0.0002	0.0008	0.00002471	0.0003	Likely benign
c.4097-141A>C	*BRCA1*	-	Intronic	rs799916	0.5	0.74	-	-	Benign
c.4185+12_4185+13delGT	*BRCA1*	-	Intronic	rs273900723	-	-	0.0149	0.0142	Benign
c.4308T>C	*BRCA1*	p.Ser1436Ser	Silent	rs1060915	0.34	0.16	0.3431	0.18	Likely benign
c.4485-137T>A	*BRCA1*	-	Intronic	rs2236762	0.5	0.74	-	-	Benign
c.4485-63C>G	*BRCA1*	-	Intronic	rs273900734	0.35	0.22	-	-	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.4675+105G>A	*BRCA1*	-	Intronic	rs8176213	0.0036	0.013	-	-	Benign
c.4837A>G	*BRCA1*	p.Ser1613Gly	Missense	rs1799966	0.36	0.23	0.3496	0.2397	Benign
c.5152+66G>A	*BRCA1*	-	Intronic	rs3092994	0.34	0.22	-	-	Benign
c.548-58elT	*BRCA1*	-	Intronic	rs8176144	-	-	-	-	Benign
c.548-80T>C	*BRCA1*	-	Intronic	rs8176143	0.0068	0.026	-	-	Benign
c.81-13C>G	*BRCA1*	-	Intronic	rs56328013	0.001	0.003	0.0002	-	Benign
c.*105A>C	*BRCA2*	-	3'UTR	rs15869	0.16	0.011	-	-	Benign
c.10234A>G	*BRCA2*	p.Ile3412Val	Missense	rs1801426	0.045	0.12	0.0227	0.1099	Benign
c.1114A>C	*BRCA2*	p.Asn372His	Missense	rs144848	0.25	0.084	0.2779	0.1249	Benign
c.-11C>T	*BRCA2*	-	5'UTR variant	rs76874770	0.0044	0.017	0.0016	0.0178	Benign
c.-26G>A	*BRCA2*	-	5'UTR variant	rs1799943	0.21	0.065	0.2465	0.1005	Benign
c.3264T>C	*BRCA2*	p.Pro1088Pro	Silent	rs36060526	0.0068	0.026	0.0024	0.0264	Benign
c.3396A>G	*BRCA2*	p.Lys1132Lys	Silent	rs1801406	0.27	0.21	0.2945	0.2292	Benign
c.3807T>C	*BRCA2*	p.Val1269Val	Silent	rs543304	0.17	0.18	0.1898	0.205	Benign
c.4090A>C	*BRCA2*	p.Ile1364Leu	Missense	rs56248502	0.0044	0.017	0.0017	0.019	Benign
c.4563A>G	*BRCA2*	p.Leu1521Leu	Silent	rs206075	0.97	0.9	0.9931	0.9264	Benign
c.5418A>G	*BRCA2*	p.Glu1806Glu	Silent	rs34351119	0.0068	0.026	0.0023	0.0262	Benign
c.6220C>A	*BRCA2*	p.His2074Asn	Missense	rs34309943	0.0028	0.0098	0.0009	0.0103	Benign
DNA nomenclature	**Gene**	**Protein nomenclature**	**Mutation type**	**dbSNP**	**1000 Genomes MAF**	**1000 Genomes MAF (AFR)**	**ExAC MAF**	**ExAC MAF (AFR)**	**Mutation Classification**
c.6513G>C	*BRCA2*	p.Val2171Val	Silent	rs206076	0.97	0.9	0.993	0.9253	Benign
c.6513G>T	*BRCA2*	p.Val2171Val	Silent	rs206076	0.0004	0.0015	0.0000579	-	Likely benign
c.681+56C>T	*BRCA2*	-	Intronic	rs2126042	0.19	0.27	-	-	Benign
c.6938-120T>C	*BRCA2*	-	Intronic	rs206080	0.97	0.9	-	-	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.7017G>C	*BRCA2*	p.Lys2339Asn	Missense	rs45574331	0.0068	0.026	0.0023	0.0265	Benign
c.7242A>G	*BRCA2*	p.Ser2414Ser	Silent	rs1799955	0.23	0.21	0.2244	0.2116	Benign
c.7319A>G	*BRCA2*	p.His2440Arg	Missense	rs4986860	0.01	0.039	0.003	0.0341	Benign
c.7397T>C	*BRCA2*	p.Val2466Ala	Missense	rs169547	0.98	0.91	0.9937	0.933	Benign
c.7806-14T>C	*BRCA2*	-	Intronic	rs9534262	0.53	0.58	0.5208	0.5729	Benign
c.8487+19A>G	*BRCA2*	-	Intronic	rs11571743	0.01	0.039	0.0033	0.0349	Benign
c.8755-66T>C	*BRCA2*	-	Intronic	rs4942486	0.51	0.51	-	-	Benign
c.8830A>T	*BRCA2*	p.Ile2944Phe	Missense	rs4987047	0.009	0.033	0.0038	0.0419	Benign
c.9257-143T>A	*BRCA2*	-	Intronic	rs80220521	0.0062	0.023	-	-	Benign
c.9257-83G>A	*BRCA2*	-	Intronic	rs9595456	0.051	0.14	-	-	Benign
c.9730G>A	*BRCA2*	p.Val3244Ile	Missense	rs11571831	0.0068	0.026	0.0023	0.0259	Benign
c.1141-94G>T	*BRIP1*	-	Intronic	rs114901675	0.0068	0.026	-	-	Likely benign
c.1935+11G>A	*BRIP1*	-	Intronic	rs79121306	0.0004	0.0015	0.0000995	0.0012	Likely benign
c.2061G>C	*BRIP1*	p.Val687Val	Silent	rs112414873	0.0014	0.0053	0.0004	0.005	Benign
c.2236A>G	*BRIP1*	p.Ile746Val	Missense	rs111536363	0.0042	0.016	0.0013	0.0145	Likely benign
c.2755T>C	*BRIP1*	p.Ser919Pro	Missense	rs4986764	0.63	0.69	0.5972	0.6702	Benign
c.94-18T>G	*BRIP1*	-	Intronic	rs2138005	0.033	0.12	0.0096	0.103	Benign
c.*54C>T	*CDH1*	-	3'UTR variant	rs1801026	0.1546	0.2	-	-	Benign
c.1138-3C>T	*CDH1*	-	Intronic	rs36103202	0.0012	0.0045	0.0001	0.0015	Benign
c.1896C>T	*CDH1*	p.His632His	Silent	rs33969373	0.023	0.068	0.0114	0.0531	Benign
c.1937-13T>C	*CDH1*	-	Intronic	rs2276330	0.058	0.0083	0.1048	0.0215	Benign
c.2076T>C	*CDH1*	p.Ala692Ala	Silent	rs1801552	0.72	0.94	0.6548	0.886	Benign
c.2253C>T	*CDH1*	p.Asn751Asn	Silent	rs33964119	0.054	0.058	0.0396	0.0493	Benign
c.2439+22C>T	*CDH1*	-	Intronic	rs34751606	0.0026	0.0098	0.001	0.0112	Benign
c.2634C>T	*CDH1*	p.Gly878Gly	Silent	rs2229044	0.018	0.05	0.00963	0.0379	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.324A>G	*CDH1*	p.Arg108Arg	Silent	rs116542018	0.0034	0.011	0.001	0.0083	Benign
c.48+5C>G	*CDH1*	-	Intronic	rs77312180	0.035	0.12	0.0046	0.1576	Benign
c.48+6C>T	*CDH1*	-	Intronic	rs3743674	0.76	0.64	0.8111	0.649	Benign
c.833-16C>G	*CDH1*	-	Intronic	rs33984587	0.0032	0.012	0.0009	0.0101	Benign
c.933C>G	*CDH1*	p.Leu311Leu	Silent	rs35539711	0.0078	0.028	0.0029	0.0318	Benign
c.252A>G	*CHEK2*	p.Glu84Glu	Silent	rs1805129	0.043	0.054	0.0356	0.0647	Benign
c.*265A>T	*EPCAM*	-	3'UTR variant	rs11554292	0.017	0.063	-	-	Likely benign
c.*36G>A	*EPCAM*	-	3'UTR variant	rs77204711	0.017	0.061	0.0049	0.0538	Likely benign
c.1117A>G	*FAM175A*	p.Asp373Asn	Missense	rs13125836	0.021	0.0015	0.0424	0.011	Benign
c.917T>C	*FAM175A*	p.Val306Ala	Missense	rs138986552	0.001		0.001674		Likely benign
c.951C>T	*FAM175A*	p.Leu317Leu	Silent	rs79357787	0.0082	0.03	0.002282	0.0245	Benign
c.1299C>T	*MEN1*	p.His433His	Silent	rs540012	0.98	0.91	0.9927	0.9202	Benign
c.1039-8T>A	*MLH1*	-	Intronic	rs193922367	-	-	0.0001	-	Benign
c.1410-54C>T	*MLH1*	-	Intronic	rs7633154	0.05	0.17	-	-	Benign
c.1558+14G>A	*MLH1*	-	Intronic	rs41562513	0.05	0.12	0.0395	0.1003	Benign
c.1668-19A>G	*MLH1*	-	Intronic	rs9876116	0.31	0.44	0.3712	0.4171	Benign
c.1963A>G	*MLH1*	p.Ile655Val	Missense	rs55907433	0.0026	0.0091	0.001	0.0083	Likely benign
c.2152C>T	*MLH1*	p.His718Tyr	Intronic	rs2020873	0.029	0.11	0.007	0.0784	Benign
c.453+79A>G	*MLH1*	-	Intronic	rs4234259	0.31	0.42	-	-	Benign
c.454-51T>C	*MLH1*	-	Intronic	rs4647255	0.022	0.042	0.0248	0.043	Benign
c.474C>T	*MLH1*	p.Asn158Asn	Silent	rs4647256	0.011	0.039	0.0035	0.0376	Likely benign
c.588+11G>C	*MLH1*	-	Intronic	rs4647258	0.0096	0.034	0.0028	0.031	Benign
c.655A>G	*MLH1*	-	Intronic	rs1799977	0.13	0.039	0.2325	0.0795	Benign
c.-93G>A	*MLH1*	-	5'UTR variant	rs1800734	0.32	0.15	-	-	Benign
c.1002C>G	*MRE11A*	p.Ser334Arg	Missense	rs115244417	0.015	0.054	0.0041	0.0443	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.1098+17T>C	*MRE11A*	-	Intronic	rs1805365	0.054	0.15	0.0273	0.1429	Benign
c.1225+19T>C	*MRE11A*	-	Intronic	rs641936	0.3888	0.36	0.37831	0.3796	Benign
c.1326+93G>T	*MRE11A*	-	Intronic	rs115203193	0.0072	0.023	-	-	Likely benign
c.1413G>A	MRE11A	p.Glu471Glu	Silent	rs560102780	-	-	0.000008241	0.00009615	Likely benign
c.1798G>C	*MRE11A*	p.Glu600Gln	Missense	rs145415033	0.0008	0.003	0.0002	0.0025	Likely benign
c.1994+10G>A	*MRE11A*	-	Intronic	rs1805366	0.0086	0.031	0.00317	0.0353	Benign
c.20+28G>A	*MRE11A*	-	Intronic	rs497763	0.4573	0.34	0.46211	0.3731	Benign
c.2071-53G>T	*MRE11A*	-	Intronic	rs13447742	0.0046	0.017	-	-	Benign
c.2092A>G	*MRE11A*	p.Met698Val	Missense	rs1805362	0.0046	0.017	0.0015	0.0166	Benign
c.315-4delT	*MRE11A*	-	Intronic	rs587781293	-	-	0.0392	0.0216	Benign
c.403-6G>A	*MRE11A*	-	Intronic	rs535801	0.31	0.17	0.3252	0.2028	Benign
c.1511-91G>T	*MSH2*	-	Intronic	rs3732182	0.48	0.67	-	-	Benign
c.1661+12G>A	*MSH2*	-	Intronic	rs3732183	0.48	0.67	0.3501	0.6223	Benign
c.211+9C>G	*MSH2*	-	Intronic	rs2303426	0.63	0.81	0.5047	0.7659	Benign
c.2766T>C	*MSH2*	p.Phe922Phe	Silent	rs55859129	0.014	0.051	0.0039	0.0444	Benign
c.380A>G	*MSH2*	p.Arg127Ser	Missense	rs17217772	0.025	0.087	0.007	0.076	Benign
c.573C>T	*MSH2*	p.Leu191Leu	Silent	rs1800151	0.019	0.07	0.0059	0.0645	Benign
c.793-23G>A	*MSH2*	-	Intronic	rs17224255	0.014	0.048	0.0042	0.0476	Benign
c.819A>G	*MSH2*	p.Val273Val		rs146577635	0.0002	0.0008	0.0003	0.0028	Benign
c.2744A>G	*MSH2,KCNK12*	-	Intergenic	rs2303424	0.53	0.66	0.4579	0.6159	Benign
c.1164C>T	*MSH6*	p.His388His	Silent	rs55708305	0.015	0.051	0.0039	0.0404	Benign
c.116G>A	*MSH6*	p.Gly39Glu	Missense	rs1042821	0.2	0.21	0.2141	0.2661	Benign
c.1345C>T	*MSH6*	p.Leu449Leu	Silent	rs3136333	0.003	0.011	0.0012	0.0142	Benign
c.2253T>C	*MSH6*	p.Asn751Asn	Silent	rs2020913	0.024	0.088	0.0065	0.0697	Benign
c.2272C>T	*MSH6*	p.Leu758Leu	Silent	rs56371757	0.015	0.05	0.0038	0.0397	Benign
c.3173-101G>C	*MSH6*	-	Intronic	rs2072447	0.82	0.95	-	-	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.3173-18T>A	*MSH6*	-	Intronic	rs189672273	0.0024	0.0091	0.0007	0.0079	Likely benign
c.3265T>C	*MSH6*	p.Leu1089Leu	Silent	rs34490141	0.0016	0.0061	0.0003	0.0031	Benign
c.3438+14A>T	*MSH6*	-	Intronic	rs2020911	0.4	0.13	0.4043	0.1787	Benign
c.3646+91T>C	*MSH6*	-	Intronic	rs3136359	0.81	0.95	-	-	Benign
c.3647-51_3647-35delTTTTTGTTTTAATTCCT	*MSH6*	-	Intronic	rs374071326	-	-	-	-	Likely benign
c.3802-40C>G	*MSH6*	-	Intronic	rs3136367	0.81	0.94	0.7426	0.9197	Benign
c.4002-11_4002-10dupTT	*MSH6*	-	Intronic	rs587779307	-	-	0.1323	0.1439	Benign
c.457+52T>A	*MSH6*	-	Intronic	rs3136282	0.19	0.23	0.2715	0.2323	Benign
c.540T>C	*MSH6*	p.Asp180Asp	Silent	rs1800935	0.14	0.15	0.2144	0.1675	Benign
c.A276A>G	*MSH6*	p.Pro92Pro	Silent	rs1800932	0.087	0.084	0.1349	0.1058	Benign
c.1014G>C	*MUTYH*	p.Asp338His	Missense	rs3219489	0.31	0.26	0.2967	0.2846	Benign
c.1544C>T	*MUTYH*	p.Ser515Phe	Missense	rs140118273	0.0046	0.0038	0.0079	0.0078	Benign
c.1601G>A	*MUTYH*	p.Arg534Asp	Missense	rs3219497	0.013	0.048	0.0033	0.0364	Benign
c.504+35A>G	*MUTYH*	-	Intronic	rs3219487	0.94	0.98	0.9255	0.9653	Benign
c.102G>A	*NBN*	p.Leu34eu	Silent	rs1063045	0.392	0.31	0.35272	0.3136	Benign
c.1124+18C>T	*NBN*	-	Intronic	rs2234744	0.353	0.22	0.34459	0.2271	Benign
c.1197T>C	*NBN*	p.Asp399Asp	Silent	rs709816	A = 0.39140	0.15	G = 0.46759	0.7691	Benign
c.1222A>G	*NBN*	p.Lys408Glu	Missense	rs34120922	0.0026	0.0098	0.0009	0.0102	Likely benign
c.1398-19C>T	*NBN*	-	Intronic	rs201495716	0.001	0.003	0.00051	0.0053	Benign
c.1914+9C>T	*NBN*	-	Intronic	rs13312938	0.024	0.084	0.009	0.0769	Benign
c.1915-7A>G	*NBN*	-	Intronic	rs2308962	0.379	0.31	0.3529	0.3151	Benign
c.2016A>G	*NBN*	p.Pro672Pro	Silent	rs1061302	0.3528	0.22	0.3451	0.2278	Benign
c.2071-30A>T	*NBN*	-	Intronic	rs3736639	0.3792	0.31	0.3535	0.3156	Benign
c.2082T>G	*NBN*	p.Pro694Pro	Silent	rs7823648	0.022	0.081	0.0071	0.0755	Benign
c.37+5G>A	*NBN*	-	Intronic	rs116735828	0.009	0.029	0.0027	0.0223	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.553G>C	*NBN*	p.Glu185Gln	Missense	rs1805794	0.36	0.23	0.3453	0.2357	Benign
c.703-18G>A	*NBN*	-	Intronic	rs769418	0.0743	0.003	0.06824	0.0099	Benign
c.1676A>G	*PALB2*	p.Asp559Arg	Missense	rs152451	0.15	0.24	0.1176	0.2251	Likely benign
c.1684+29A>G	*PALB2*	-	Intronic	rs74320059	0.025	0.091	0.0093	0.0863	Benign
c.2586+58C>T	*PALB2*	-	Intronic	rs249954	0.35	0.48	-	-	Benign
c.2586+81C>T	*PALB2*	-	Intronic	rs114710547	0.0082	0.03	-	-	Likely benign
c.2587-38C>G	*PALB2*	-	Intronic	rs180177119	0.0022	-	0.0017	0.0003	Likely benign
c.-47G>A	*PALB2*	-	5'UTR variant	rs8053188	0.066	0.13	0.0421	0.1129	Benign
c.925A>G	*PALB2*	p.Ile309Val	Missense	rs3809683	0.012	0.03	0.0034	0.0297	Benign
c.1041G>A	*PMS2*	p.Glu347Glu	Silent	rs150515238	-	-	0.00000825	0.00009634	Likely benign
c.1408C>T	*PMS2*	p.Pro470Ser	Missense	rs1805321	0.36	0.27	0.3854	0.2804	Benign
c.1488C>T	*PMS2*	p.His496His	Silent	rs1805320	0.032	0.11	0.0074	0.0781	Benign
c.1532C>T	*PMS2*	p.Thr511Met	Missense	rs74902811	0.037	0.11	0.0094	0.0818	Benign
c.-154C>G	*PMS2*	-	Upstream variant	rs3735296	0.16	0.26	-	-	Benign
c.1557T>C	*PMS2*	p.Tyr519Tyr	Silent	rs6972869	0.0098	0.036	0.0031	0.0349	Benign
c.1621A>G	*PMS2*	p.Lys541Glu	Missense	rs2228006	0.88	0.94	0.8514	0.9229	Benign
c.2007-4G>A	*PMS2*	-	Intronic	rs140788589	0.16	0.023	0.1548	0.0443	Benign
c.2007-7C>T	*PMS2*	-	Intronic	rs55954143	0.13	0.28	0.0885	0.2549	Benign
c.2466T>C	*PMS2*	p.Leu822Leu	Silent	rs10000	0.063	0.026	0.2114	0.0848	Benign
c.2570G>C	*PMS2*	p.Gly857Ala	Missense	rs1802683	-	-	0.3702	0.4278	Benign
c.288C>T	*PMS2*	p.Ala96Ala	Silent	rs12532895	0.11	0.011	0.0805	0.0178	Benign
c.705+17A>G	*PMS2*	-	Intronic	rs62456182	0.33	0.18	0.3678	0.1979	Benign
c.706-5_706-4delTT	*PMS2*	-	Intronic	rs531184698 (merged into rs776641246)	-	-	0.0078	0.0087	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.780C>G	*PMS2*	p.Ser260Ser	Silent	rs1805319	0.83	0.86	0.8109	0.8459	Benign
c.-93G>T	*PMS2*	-	Upstream variant	rs6976537	0.032	0.12	0.0079	-	Likely benign
c.1026+32T>G	*PTEN*	-	Intronic	rs555895	0.43	0.53	0.3699	0.5007	Benign
c.165-13_165-10delGTTT	*PTEN*	-	Intronic	rs531173559 (merged into rs786204877)	-	-	0.0005	0.0048	Likely benign
c.802-4_802-3delTT	*PTEN*	-	Intronic	rs398123327	-	-	0.0089	0.01	Likely benign
c.2025C>T	*RAD50*	p.Asp675Asp	Silent	rs34147298	0.027	0.096	0.0075	0.0825	Benign
c.214-44G>A	*RAD50*	-	Intronic	rs74769721	0.024	0.075	0.0103	0.0799	Benign
c.3846T>C	*RAD50*	p.Tyr1282Tyr	Silent	rs1804670	0.047	0.17	0.0151	0.1652	Benign
c.-38A>G	*RAD50*	-	5'UTR variant	rs4526098	0.92	0.74	0.9752	0.7598	Benign
c.551+19G>A	*RAD50*	-	Intronic	rs17166050	0.17	0.12	0.1988	0.117	Benign
c.756+7delT	*RAD50*	-	Intronic	rs377720482	-	-	0.0006	0.0071	Benign
c.3476-12_3476-10delTTC	*RAD50/TH2LCRR*	-	Intronic	rs147883712 (merged into rs730881922)	-	-	0.0016	0.0144	Benign
c.186A>G	*RAD51C*	p.Gln62Gln	Silent	rs28363303	0.0038	0.013	0.001	0.0114	Benign
c.-26C>T	*RAD51C*	-	5'UTR variant	rs12946397	0.16	0.14	0.1797	0.142	Likely benign
c.572-17G>T	*RAD51C*	-	Intronic	rs193023469	0.0036	0.0068	0.0041	0.0067	Likely benign
c.837+13T>C	*RAD51C*	-	Intronic	rs188613030	0.0014	0.0053	0.0004	0.0042	Benign
c.90G>A	*RAD51C*	p.Ala30Ala	Silent	rs115414895	0.014	0.052	0.0036	0.0409	Benign
c.144+581C>T	*RAD51D*	-	Intronic	rs9901455	0.19	0.35	0.1184	0.3018	Benign
c.-7G>T	*RAD51L3-RFFL*	-	ncRNA_intronic	rs553444507	-	-	-	-	Likely benign
c.264C>A	*STK11*	p.Ile88Ile	Silent	rs56354945	0.024	0.083	0.0089	0.0934	Benign

DNA nomenclature	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	Mutation Classification
c.290+36G>T	*STK11*	-	Intronic	rs3764640	0.35	0.27	0.3247	0.3219	Benign
c.369G>A	*STK11*	p.Gln123Gln	Silent	rs140112347	0.0062	0.023	0.0019	0.0216	Benign
c.374+24G>T	*STK11*	-	Intronic	rs2075604	0.27	0.41	0.1618	0.3753	Benign
c.375-49G>A	*STK11*	-	Intronic	rs34928889	0.53	0.47	0.4661	0.4768	Benign
c.464+40_464+46dupGGGGGCC	*STK11*	-	Intronic	rs58579265	0.25	0.22	0.2067	0.1275	Likely benign
c.816C>T	*STK11*	p.Tyr272Tyr	Silent	rs9282859	0.024	0.086	0.0078	0.0887	Benign
c.108G>A	*TP53*	p.Pro36Pro	Silent	rs1800370	0.0076	0.0045	0.0126	0.005	Benign
c.215C>T	*TP53*	p.Pro72Arg	Missense	rs1042522	0.54	0.33	0.66	0.3842	Likely benign
c.-232C>G	*TP53*	-	5'UTR variant	rs1642785	0.58	0.49	0.6689	0.525	Benign
c.96+16_96+31delCCCCAGCCCTCCAGGT	*TP53*	-	Intronic	rs59758982	-	-	-	-	Benign
c.97-29C>A	TP53	-	Intronic	rs17883323	0.078	0.12	0.0659	0.1087	Benign
c.993+12T>C	*TP53*	-	Intronic	rs1800899	0.016	0.038	0.0117	0.0277	Likely benign
c.40-10C>T	*XRCC2*	-	Intronic	rs3218472	0.009	0.025	0.0069	0.0223	Benign
c.-50G>A	*XRCC2*	-	5'UTR variant	rs139350845	0.0016	0.0061	0.00051	0.0057	Benign

Appendix L: Novel variants

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.282G>A	108100001	ATM	p.Met94Ile	Missense	Tolerated	Benign	19.6	Disease-causing	Moderate
c.285G>A	108100004	*ATM*	p.Gln95Gln	Silent	Tolerated	Benign	6.197	Disease-causing	Low
c.300G>A	108100019	*ATM*	p.Leu100Leu	Silent	Tolerated	Benign	10.22	Disease-causing	Low
g.102767_102768delTT	108195978	*ATM*	-	Intronic	N/A	N/A	4.372	Polymorphism	Modifier
g.102767_102769delTTT	108195979	*ATM*	-	Intronic	N/A	N/A	4.028	Polymorphism	Modifier
g.102767_102770delTTTT	108195980	*ATM*	-	Intronic	N/A	N/A	3.664	Polymorphism	Modifier
g.113773_113777delTAAAG	108224486	*ATM*	-	Intronic	N/A	N/A	17.99	Disease-causing	Low
g.142284_142285delTCinsAA	108195976	*ATM*	-	Intronic	N/A	N/A	0.816	Polymorphism	Modifier
g.26425A>T	108119635	*ATM*	-	Intronic	N/A	N/A	8.977	Polymorphism	Modifier
g.48556A>G	108141766	*ATM*	-	Intronic	N/A	N/A	3.454	Polymorphism	Modifier
g.60414A>G	108153624	*ATM*	-	Intronic	N/A	N/A	1.185	Polymorphism	Modifier
g.67013A>G	108160223	*ATM*		Intronic	N/A	N/A	2.807	Polymorphism	Modifier
g.6877C>A	108100087	*ATM*	-	Intronic	N/A	N/A	2.643	Polymorphism	Modifier
g.6896_6897delTT	108100107	*ATM*	-	Intronic	N/A	N/A	5.843	Polymorphism	Modifier
g.93804delA	108187014	*ATM*	-	Intronic	N/A	N/A	5.096	Polymorphism	Modifier
g.97300G>T	108190510	*ATM*	-	Intronic	N/A	N/A	5.641	Polymorphism	Modifier
g.57018C>G	215617411	*BARD1*	-	Intronic	N/A	N/A	0.942	Polymorphism	Modifier
g.57063T>G	215617366	*BARD1*		Intronic	N/A	N/A	7.798	Disease-causing	Modifier
c.2244T>C	91310190	*BLM*	p.Asn748Asn	Silent	Tolerated		6.822	Disease-causing	Low
g.52105A>G	91312662	*BLM*	-	Intronic	N/A	N/A	12.08	Disease-causing	Low

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
g.102383delT	41219908	*BRCA1*	-	Intronic	N/A	N/A	9.501	Polymorphism	Modifier
g.102383insT	41219907	*BRCA1*	-	Intronic	N/A	N/A	5.566	Polymorphism	Modifier
g.102384_102385delTT	41219909	*BRCA1*	-	Intronic	N/A	N/A	9.453	Polymorphism	Modifier
g.102401C>A	41219890	*BRCA1*	-	Intronic	N/A	N/A	4.434	Polymorphism	Modifier
g.102406C>A	41219885	*BRCA1*	-	Intronic	N/A	N/A	4.985	Polymorphism	Modifier
g.102413C>T	41219878	*BRCA1*	-	Intronic	N/A	N/A	5.682	Polymorphism	Modifier
g.102415G>T	41219876	*BRCA1*	-	Intronic	N/A	N/A	6.447	Polymorphism	Modifier
g.102417C>A	41219874	*BRCA1*	-	Intronic	N/A	N/A	2.846	Polymorphism	Modifier
g.102440C>T	41219851	*BRCA1*	-	Intronic	N/A	N/A	2.443	Polymorphism	Modifier
g.102442C>T	41219849	*BRCA1*	-	Intronic	N/A	N/A	2.727	Polymorphism	Modifier
g.63701A>C	41258590	*BRCA1*	-	Intronic	N/A	N/A	3.982	Polymorphism	Modifier
g.102394T>G	41219897	*BRCA1*	-	Intronic	N/A	N/A	5.061	Polymorphism	Modifier
g.102417C>T	41219874	*BRCA1*	-	Intronic	N/A	N/A	5.138	Polymorphism	Modifier
c.10191C>T	32972841	*BRCA2*	p.Ser3397Ser	Silent			11.87	Polymorphism	Low
g.83273delACinsT	59857610	*BRIP1*	-	Intronic	N/A	N/A	0.255	Polymorphism	Modifier
g.40578G>A	47612874	*EPCAM*	-	Intronic	N/A	N/A	5.916	Polymorphism	Modifier
g.47616555C>T	47616555	*EPCAM,MSH2*	-	Intergenic	N/A	N/A	1.577	N/A	Modifier
c.-51_-54delTTGCinsCTGG	84406279	*FAM175A*	-	5'UTR variant	N/A	N/A	8.113	Polymorphism	Modifier
g.38097C>G	84406405	*FAM175A*	-	Upstream	N/A	N/A	5.599	Polymorphism	Modifier
g.59653C>T	84384849	*FAM175A*	-	Intronic	N/A	N/A	2.154	Polymorphism	Modifier
g.32273delATATAinsTT	37067099	*MLH1*	-	Intronic	N/A	N/A	5.596	Polymorphism	Modifier
g.32273delATATAinsTTTTT	37067099	*MLH1*	-	Intronic	N/A	N/A	0.04	Polymorphism	Modifier
g.32275delATAinsT	37067099	*MLH1*	-	Intronic	N/A	N/A	5.032	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
g.32275delATAinsTTT	37067099	*MLH1*	-	Intronic	N/A	N/A	0.031	Polymorphism	Modifier
g.32280T>A	37067102	*MLH1*	-	Intronic	N/A	N/A	6.292	Polymorphism	Modifier
g.29823_29831delTTTACCTCAinsCTTAACTCG	94197260	*MRE11A*	-	Intronic	N/A	N/A	7.548	Polymorphism	Modifier
g.58085_58086delGCinsAA	94168991	*MRE11A*	-	Intronic	N/A	N/A	7.86	Polymorphism	Modifier
g.11453_11464delAAAAAAAAAAAA	47641571	*MSH2*	-	Splicing	N/A	N/A	17.82	Disease-causing	Low
g.72346_72347insTTG	47702456	*MSH2*	-	Intronic	N/A	N/A	1.368	Polymorphism	Low
g.72483A>G	47702590	*MSH2*	-	Intronic	N/A	N/A	15.37	Polymorphism	Modifier
g.77699C>T	47707806	*MSH2*	-	Intronic	N/A	N/A	5.352	Polymorphism	Modifier
g.23614G>A	48033834	*MSH6*	-	Intronic	N/A	N/A	0.005	Polymorphism	Modifier
g.23622G>A	48033842	*MSH6*	-	Intronic	N/A	N/A	0.002	Polymorphism	Modifier
g.25055G>A	90990402	*NBN*	-	Intronic	N/A	N/A	12.07	Polymorphism	Modifier
g.37603A>T	23615029	*PALB2*	-	Intronic	N/A	N/A	6.606	Polymorphism	Modifier
g.11861_11861delT	6036895	*PMS2*	-	Intronic	N/A	N/A	7.531	Polymorphism	Modifier
g.5538A>G	6043219	*PMS2*	-	Intronic	N/A	N/A	0.818	Polymorphism	Modifier
g.5529_5531delGTTinsACC	6043228	*PMS2*	-	Intronic	N/A	N/A	2.997	Polymorphism	Modifier
g.5517G>A	6043240	*PMS2*	-	Intronic	N/A	N/A	0.547	Polymorphism	Modifier
g.5514T>A	6043243	*PMS2*	-	Intronic	N/A	N/A	1.34	Polymorphism	Modifier
g.5517delCTGAinsTTGT	6043243	*PMS2*	-	Intronic	N/A	N/A	3.239	Polymorphism	Modifier
g.5504A>G	6043253	*PMS2*	-	Intronic	N/A	N/A	0.099652	Polymorphism	Modifier
g.5517delCTGAAAATAATAATinsTTGTAAATAAC	6043253	*PMS2*	-	Intronic	N/A	N/A	6.974	Polymorphism	Modifier

DNA nomenclature	**Genomic end position**	**Gene**	**Protein nomenclature**	**Mutation type**	**SIFT prediction**	**Polyphen**	**CADD**	**MutationTaster**	**VEP**
g.5491A>G	6043266	*PMS2*	-	Intronic	N/A	N/A	10.85	Polymorphism	Modifier
g.5486T>A	6043271	*PMS2*	-	Intronic	N/A	N/A	11.44	Polymorphism	Modifier
g.5476G>A	6043281	*PMS2*	-	Intronic	N/A	N/A	3.558	Polymorphism	Modifier
g.5458C>T	6043299	*PMS2*	-	Intronic	N/A	N/A	10.48	Polymorphism	Modifier
g.5447A>G	6043310	*PMS2*	-	Intronic	N/A	N/A	13.94	Polymorphism	Modifier
c.342T>G	6043332	*PMS2*	P.Leu114Leu	Silent	Tolerated	Benign	0.079	Disease-causing	Low
c.298C>G	6043376	*PMS2*	p.Gln100Glu	Missense	Tolerated	Benign	6.802	Polymorphism	Moderate
g.5309G>T	6043448	*PMS2*	-	Intronic	N/A	N/A	1.287	Polymorphism	Modifier
g.5301A>G	6043456	*PMS2*	-	Intronic	N/A	N/A	3.634	Polymorphism	Modifier
g.5289A>T	6043468	*PMS2*	-	Intronic	N/A	N/A	5.763	Polymorphism	Modifier
g.5280_5281delAAAinsGG	6043477	*PMS2*	-	Intronic	N/A	N/A	2.356	Polymorphism	Modifier
g.5254delA	6043503	*PMS2*	-	Intronic	N/A	N/A	5.885	Polymorphism	Modifier
g.6786741G>A	6786741	*PMS2CL*	-	ncRNA_exonic	N/A	N/A	14.19	N/A	Modifier
g.6790912G>C	6790912	*PMS2CL*	-	ncRNA_exonic	N/A	N/A	7.591	N/A	Modifier
g.1534A>C	89624403	*PTEN*	-	Intronic	N/A	N/A	4.178	Polymorphism	Modifier
g.23195delT	131914905	*RAD50*	-	Intronic	N/A	N/A	5.289	Disease-causing	Modifier
g.17956G>A	33430586	RAD51L3-RFFL	-	ncRNA_intronic	N/A	N/A	4.688	Polymorphism	Modifier
g.14353C>T	33434189	RAD51L3-RFFL	-	ncRNA_intronic	N/A	N/A	3.035	Polymorphism	Modifier
c.-81C>T	152373245	*XRCC2*	-	5'UTR variant	N/A	N/A	9.048	Polymorphism	Modifier

Appendix M: ExAC reported variants

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	ExAC MAF	ExAC MAF (African)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
g.102786C>T	108195996	*ATM*	-	Intronic	0.0037	0.0353	N/A	N/A	0.527	Polymorphism	Modifier
g.102787A>T	108195997	*ATM*	-	Intronic	0.0029	0.0226	N/A	N/A	0.447	Polymorphism	Modifier
g.149991_149992delAA	108188268	*ATM*	-	Intronic	0.0022	0.0132	N/A	N/A	10.72	Polymorphism	Modifier
g.21452delT	108114662	*ATM*	-	Intronic	0.3494	0.2916	N/A	N/A	0.235	Polymorphism	Modifier
g.28201delT	108121411	*ATM*	-	Intronic	0.0077	0.0076	N/A	N/A	0.769	Polymorphism	Modifier
g.48745delT	108141956	*ATM*	-	Intronic	0.022	0.016	N/A	N/A	2.322	Polymorphism	Modifier
g.48746_48747delTT	108141957	*ATM*	-	Intronic	0.022	0.016	N/A	N/A	2.409	Polymorphism	Modifier
g.58488delT	108151698	*ATM*	-	Intronic	0.0025	0.001	N/A	N/A	2.534	Polymorphism	Modifier
g.6896delT	108100106	*ATM*	-	Intronic	0.4396	0.3647	N/A	N/A	3.735	Polymorphism	Modifier
g.70114delT	108163324	*ATM*	-	Intronic	0.0000338	0.0002	N/A	N/A	6.862	Polymorphism	Modifier
g.72417A>G	108165627	*ATM*	-	Intronic	0.000008254	0.00009733	N/A	N/A	4.059	Polymorphism	Modifier
g.17245_17246delTT	215657184	*BARD1*	-	Intronic	0.0063	0.0047	N/A	N/A	5.256	Polymorphism	Modifier
g.17251A>T	215657178	*BARD1*	-	Intronic	0.0854	0.1016	N/A	N/A	5.154	Polymorphism	Modifier
g.49563T>C	91310120	*BLM*	-	Intronic	0.000008474	0.0001	N/A	N/A	13.31	Polymorphism	Modifier
g.52094delT	91312651	*BLM*	-	Intronic	0.0003	0.0001	N/A	N/A	3.66	Polymorphism	Modifier
g.83283delT	59857600	*BRIP1*	-	Intronic	0.016	0.0154	N/A	N/A	7.368	Polymorphism	Modifier
g.118052delT	37107129	*LRRFIP2*	-	Intronic	0.00006779	0.0014	N/A	N/A	7.234	Polymorphism	Modifier

DNA nomenclature	**Genomic end position**	**Gene**	**Protein nomenclature**	**Mutation type**	**ExAC MAF**	**ExAC MAF (African)**	**SIFT prediction**	**Polyphen**	**CADD**	**MutationTaster**	**VEP**
g.32272_32279delTATATATT	37067101	*MLH1*	-	Intronic	0.007	0.0057	N/A	N/A	5.465	Polymorphism	Modifier
g.32272_32280delTATATATTT	37067102	*MLH1*		Intronic	0.007	0.0057	N/A	N/A	7.65	Polymorphism	Modifier
g.32275_32277delATA	37067099	*MLH1*	-	Intronic	0.0127	0.0096	N/A	N/A	4.979	Polymorphism	Modifier
g.2930_2932delCTT	94224145	*MRE11A*	-	Intronic	0.00004415	0	N/A	N/A	14.42	Polymorphism	Modifier
g.11453_11458delAAAAA	47641564	*MSH2*	-	Splicing	0.0072	0.0448	N/A	N/A	18.56	Disease-causing	Low
g.11453_11466delAAAAAAAAAAAAAA	47641573	*MSH2*	-	Splicing	0.0072	0.0448	N/A	N/A	17.61	Disease-causing	Low
g.5417_5418delTT	47635525	*MSH2*	-	Intronic	0.1511	0.1073	N/A	N/A	6.436	Polymorphism	Modifier
g.72339_72343delGTTTT	47702450	*MSH2*	-	Intronic	0.0065	0.006	N/A	N/A	3.675	Polymorphism	Modifier
g.72349T>G	47702456	*MSH2*	-	Intronic	0.0003	0.0003	N/A	N/A	1.7	Polymorphism	Modifier
c.2418C>T	48027540	*MSH6*	p.Ser806Ser	Silent	0.00001653	0	N/A	N/A	11.27	Disease-causing	Low
c.2493C>T	48027615	*MSH6*	p.Pro831Pro	Silent	0.00000833	0.0001	N/A	N/A	11.62	Disease-causing	Low
g.23662C>A	48033882	*MSH6*	-	Intronic	0.00001225	-	N/A	N/A	0.003	Polymorphism	Modifier
g.23671_23674delTTT	48033893	*MSH6*	-	Intronic	0.1323	0.1439	N/A	N/A	0.004	Polymorphism	Modifier
g.23671_23674delTTTT	48033894	*MSH6*	-	Intronic	0.1323	0.1439	N/A	N/A	0.004	Polymorphism	Modifier
g.8073_8074insTG	48018295	*MSH6*	-	Intronic	0.1412	-	N/A	N/A	3.883	Polymorphism	Modifier
g.8094_8095insGA	48018316	*MSH6*	-	Intronic	0.0014	-	N/A	N/A	1.13	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	ExAC MAF	ExAC MAF (African)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
g.47987delA	90967470	*NBN*	-	Intronic	0.0059	0.0594	N/A	N/A	3.369	Polymorphism	Modifier
g.49528delT	90965929	*NBN*	-	Intronic	0.0011	0.0594	N/A	N/A	7.804	Polymorphism	Modifier
g.37591A>T	23615041	*PALB2*	-	Intronic	0.001	0.0594	N/A	N/A	1.482	Polymorphism	Modifier
c.2349C>T	6017315	*PMS2*	p.Val783Val	Silent	0.00002494	0	Tolerated	Benign	10.31	Disease-causing	Moderate
c.312T>C	6043362	*PMS2*	p.Phe104Phe	Silent	0.00001758	0	Tolerated	N/A	8.477	Disease-causing	Low
g.5323C>G	6043434	*PMS2*	-	Intronic	0.00000868	0	N/A	N/A	5.683	Polymorphism	Modifier
c.2212G>C	6018290	*PMS2*	p.Val738Leu	Missense	0.00005365	-	Tolerated	Benign	7.827	Polymorphism	Moderate
c.269C>G	6043405		p.Ser90Cys	Missense	8.7×10^{-6}	0	Deleterious	Probably damaging	27.6	Disease-causing	Moderate
g.97765delT	89720634	*PTEN*	-	Intronic	0.0089	0.01	N/A	N/A	8.317	Polymorphism	Modifier
g.97765_97765delTTT	89720636	*PTEN*	-	Intronic	0.0089	0.01	N/A	N/A	12.9	Polymorphism	Modifier
g.32637delA	131924347	*RAD50*	-	Intronic	0.0004	0.0003	N/A	N/A	3.491	Polymorphism	Modifier
g.35841delT	131927551	*RAD50*	-	Intronic	0.0235	0.0199	N/A	N/A	1.536	Polymorphism	Modifier
g.35841insT	131927551	*RAD50*	-	Intronic	0.0235	0.0199	N/A	N/A	0.132	Polymorphism	Modifier

Appendix N: Variants previously reported in 1000 Genomes and dbSNP

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.-30-79A>G	108098243	*ATM*	-	rs3218693	Intronic	0.03	0.11	N/A	N/A	1.933	Polymorphism	Modifier
c.185+31T>C	108098646	*ATM*	-	rs571866895	Intronic	-	-	N/A	N/A	0.543	Polymorphism	Modifier
c.496+74G>A	108106635	*ATM*	-	rs3218710	Intronic	0.0184	0.066	N/A	N/A	0.806	Polymorphism	Modifier
c.1803-154G>A	108123390	*ATM*	-	rs55760417	Intronic	0.016	0.058	N/A	N/A	7.731	Polymorphism	Modifier
c.2377-56A>G	108129657	*ATM*	-	rs672655	Intronic	0.4764	0.22	N/A	N/A	19.91	Disease-causing	Modifier
c.3402+70A>G	108150405	*ATM*	-	rs3218671	Intronic	0.013	0.047	N/A	N/A	6.61	Disease-causing	Modifier
c.3747-8A>G	108154946	*ATM*	-	rs569612113	Intronic	-	-	N/A	N/A	9.066	Polymorphism	Modifier
c.4910-156T>A	108167858	*ATM*	-	rs56161469	Intronic	0.034	0.076	N/A	N/A	8.747	Polymorphism	Modifier
c.6006+191_6006+192delCT	108183415	*ATM*	-	rs397897624	Intronic	-	-	N/A	N/A	8.36	Polymorphism	Modifier
c.6198+116T>C	108186956	*ATM*	-	rs4988084	Intronic	0.082	0.29	N/A	N/A	1.465	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.6198+130G>A	108186970	*ATM*	-	rs55982799	Intronic	0.016	0.058	N/A	N/A	3.552	Polymorphism	Modifier
c.6573-67_6573-64delTACA	108195973	*ATM*	-	rs202074417	Intronic	-	-	N/A	N/A	6.869	Polymorphism	Modifier
c.6573-42delT	108195977	*ATM*	-	rs398017511	Intronic	0.2744	0.391	N/A	N/A	4.871	Polymorphism	Modifier
c.7927+69T>C	108203696	*ATM*	-	rs140400673	Intronic	0.012	0.042	N/A	N/A	3.26	Polymorphism	Modifier
c.8010+114_8010+115delAT	108204810	*ATM*	-	rs562911886	Intronic	0.0028	0.0098	N/A	N/A	4.617	Polymorphism	Modifier
c.8787-55C>T	108225483	*ATM*	-	rs664982	Intronic	0.54	0.38	N/A	N/A	2.77	Polymorphism	Modifier
c.8850+60A>G	108225661	*ATM*	-	rs664143	Intronic	0.63	0.7	N/A	N/A	0.543	Polymorphism	Modifier
c.1904-67A>G	215595299	*BARD1*	-	rs4986840	Intronic	0.008	0.03	N/A	N/A	7.159	Polymorphism	Modifier
c.1811-69T>C	215609952	*BARD1*	-	rs74531001	Intronic	0.06	0.16	N/A	N/A	4.737	Polymorphism	Modifier
c.1811-87C>G	215609970	*BARD1*	-	rs80039307	Intronic	0.02	0.074	N/A	N/A	0.178	Polymorphism	Modifier
c.1314+99_1314+106delTAACCTTT	215645185	*BARD1*	-	rs141351703	Intronic	0.0433	0.0363	N/A	N/A	10.75	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.159-70T>G	215661911	*BARD1*	-	rs34377639	Intronic	0.052	0.069	N/A	N/A	12.3	Polymorphism	Modifier
c.159-97A>G	215661938	*BARD1*	-	rs76943679	Intronic	0.006	0.021	N/A	N/A	8.006	Polymorphism	Modifier
c.960-79A>C	91297962	*BLM*	-	rs28385005	Intronic	0.0018	0.0068	N/A	N/A	8.305	Polymorphism	Modifier
c.2074+89A>G	91306476	*BLM*	-	rs28385020	Intronic	0.0094	0.035	N/A	N/A	2.144	Polymorphism	Modifier
c.2193+61G>C	91308705	*BLM*	-	rs28385029	Intronic	0.049	0.06	N/A	N/A	5.085	Polymorphism	Modifier
c.2193+84C>T	91308728	*BLM*	-	rs17181698	Intronic	0.049	0.06	N/A	N/A	7.12	Polymorphism	Modifier
c.2824-162A>T	91333717	*BLM*	-	rs79727167	Intronic	0.013	0.046	N/A	N/A	1.35	Polymorphism	Modifier
c.2824-129delA	91333750	*BLM*	-	rs112877330	Intronic	0.1673	0.379	N/A	N/A	0.157	Polymorphism	Modifier
c.2824-91C>T	91333788	*BLM*	-	rs55871856	Intronic	0.0012	0.0045	N/A	N/A	5.867	Polymorphism	Modifier
c.4077-58_4077-57insGAA	91358274	*BLM*	-	rs72524505	Intronic	0.67	0.62	N/A	N/A	2.675	Polymorphism	Modifier
c.5332+78C>T	41203002	*BRCA1*	-	rs273901763	Intronic	0.0002	0.0008	N/A	N/A	0.999	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.212+88C>T	41258385	*BRCA1*	-	rs184664249	Intronic	0.001	0.0038	N/A	N/A	13.1	Disease-causing	Modifier
c.81-176C>T	41267972	*BRCA1*	-	rs112147727	Intronic	-	-	N/A	N/A	5.25	Polymorphism	Modifier
c.6841+79_6841+82delAATT	32915414	*BRCA2*	-	rs276174881	Intronic	-	-	N/A	N/A	8.398	Polymorphism	Modifier
c.2492+80A>C	59793232	*BRIP1*	-	rs9901948	Intronic	0.045	0.16	N/A	N/A	12.78	Polymorphism	Modifier
c.1628+83T>C	59861548	*BRIP1*	-	rs113210204	Intronic	0.0082	0.029	N/A	N/A	3.341	Polymorphism	Modifier
c.1473+21G>A	59870937	*BRIP1*	-	rs113766686	Intronic	-	-	N/A	N/A	3.478	Polymorphism	Modifier
c.1340 + 109G> A	59876352	*BRIP1*	-	rs2191248	Intronic	0.19	0.034	N/A	N/A	9.25	Polymorphism	Modifier
c.1340+64C>T	59876397	*BRIP1*	-	rs113218012	Intronic	-	-	N/A	N/A	6.15	Polymorphism	Modifier
c.1141-129_1141-128delTC	59876786	*BRIP1*	-	rs58808208	Intronic	-	-	N/A	N/A	5.187	Polymorphism	Modifier
c.1140+198C>G	59878416	*BRIP1*	-	rs62068834	Intronic	0.18	0.0091	N/A	N/A	7.468	Polymorphism	Modifier
c.1140+116G>C	59878498	*BRIP1*	-	rs114997860	Intronic	0.0066	0.023	N/A	N/A	0.587	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.1140+83_1140+84insT	59878523	*BRIP1*	-	rs397721717	Intronic	-	-	N/A	N/A	3.785	Polymorphism	Modifier
c.1140+83_1140+84insTT	59878524	*BRIP1*	-	rs397721717	Intronic	0.8	0.92	N/A	N/A	3.613	Polymorphism	Modifier
c.507+67T>C	59926423	*BRIP1*	-	rs73991950	Intronic	0.035	0.13	N/A	N/A	5.999	Polymorphism	Modifier
c.93+72T>G	59938736	*BRIP1*	-	rs4988342	Intronic	0.075	0.27	N/A	N/A	2.799	Polymorphism	Modifier
c.48+58_48+59insGCCCCGTGCCCCAGCCCTGC	68771445	*CDH1*	-	rs552723559	Intronic	0.066	0.22	N/A	N/A	1.64	Polymorphism	Modifier
c.688-72C>T	68844028	*CDH1*	-	rs183943559	Intronic	0.0002	0.0008	N/A	N/A	11.91	Polymorphism	Modifier
c.1321-88C>T	68849330	*CDH1*	-	rs33934457	Intronic	0.0042	0.015	N/A	N/A	8.032	Disease-causing	Modifier
c.1712-130T>C	68855774	*CDH1*	-	rs34597017	Intronic	0.027	0.098	N/A	N/A	1.241	Polymorphism	Modifier
c.*77A>TA	29083809	*CHEK2*	-	rs541513166	3'UTR variant	0.0012	0.0038	N/A	N/A	6.638	Polymorphism	Modifier
c.903+232A>G	47612581	*EPCAM*	-	rs113731108	Intronic	0.0044	0.017	N/A	N/A	1.73	Polymorphism	Modifier
c.903+515_903+516insTG	47612866	*EPCAM*	-	rs5830961	Intronic	0.63	0.86	N/A	N/A	3.024	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.903+529G>A	47612878	*EPCAM*	-	rs3924917	Intronic	0.57	0.74	N/A	N/A	3.314	Polymorphism	Modifier
c.904-657C>A	47613054	*EPCAM*	-	rs144336519	Intronic	0.0012	0.0045	N/A	N/A	4.608	Polymorphism	Modifier
c.*383A>G	47614135	*EPCAM*	-	rs149731012	3'UTR variant	0.0012	0.0045	N/A	N/A	13.24	Disease-causing	Modifier
c.47616256A>G	47616256	*EPCAM,MSH2*	-	rs72812306	Intergenic	0.13	0.07	N/A	N/A	11.87	N/A	Modifier
c.47616268A>G	47616268	*EPCAM,MSH2*	-	rs72812307	Intergenic	0.13	0.066	N/A	N/A	11.48	N/A	Modifier
c.682-93A>C	84384854	*FAM175A*	-	rs72931492	Intronic	0.1	0.38	N/A	N/A	3.699	Polymorphism	Modifier
c.682-107T>G	84384868	*FAM175A*	-	rs79585753	Intronic	0.033	0.12	N/A	N/A	0.399	Polymorphism	Modifier
c.597-73_597-70delTAAC	84388761	*FAM175A*	-	rs370010265	Intronic	-	-	N/A	N/A	8.097	Polymorphism	Modifier
c.596+140A>G	84390045	*FAM175A*	-	rs76055846	Intronic	0.012	0.042	N/A	N/A	0.39	Polymorphism	Modifier
c.596+97T>C	84390088	*FAM175A*	-	rs60638712	Intronic	0.24	0.67	N/A	N/A	2.082	Polymorphism	Modifier
c.282+131A>T	84393244	*FAM175A*	-	rs111692631	Intronic	0.046	0.17	N/A	N/A	9.3	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.215+176C>T	84397620	*FAM175A*	-	rs17352824	Intronic	0.66	0.81	N/A	N/A	2.922	Polymorphism	Modifier
c.-86G>A	84406311	*FAM175A*	-	rs72932960	Upstream	0.031	0.11	N/A	N/A	9.316	Polymorphism	Modifier
c.1699+279G>T	37107022	*LRRFIP2*	-	rs7639375	Intronic	0.3	0.39	N/A	N/A	2.633	Polymorphism	Modifier
c.453+188C>T	37048742	*MLH1*	-	rs4647247	Intronic	0.059	0.2	N/A	N/A	2.009	Polymorphism	Modifier
c.453+192C>T	37048746	*MLH1*	-	rs4647248	Intronic	0.014	0.045	N/A	N/A	7.012	Polymorphism	Modifier
c.546-75A>C	37053236	*MLH1*	-	rs4647257	Intronic	0.038	0.14	N/A	N/A	1.847	Polymorphism	Modifier
c.791-63G>A	37058934	*MLH1*	-	rs4647279	Intronic	0.0078	0.029	N/A	N/A	6.136	Polymorphism	Modifier
c.2103+63G>A	37090571	*MLH1*	-	rs112877263	Intronic	-	-	N/A	N/A	0.526	Polymorphism	Modifier
c.1784-86_1784-85delAG	94170486	*MRE11A*	-	rs369740922	Intronic	-	-	N/A	N/A	7.674	Polymorphism	Modifier
c.1783+1260A>G	94179125	*MRE11A*	-	rs1014666	Intronic	0.52	0.51	N/A	N/A	2.525	Polymorphism	Modifier
c.1783+1166A>G	94179219	*MRE11A*	-	rs13447698	Intronic	0.0022	0.0076	N/A	N/A	1.868	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.1226-53A>G	94194255	*MRE11A*	-	rs115997947	Intronic	0.0072	0.023	N/A	N/A	0.845	Polymorphism	Modifier
c.845+62C>T	94204678	*MRE11A*	-	rs111321945	Intronic	-	-	N/A	N/A	1.407	Polymorphism	Modifier
g.12715_12716insTTT	94219328	*MRE11A*	-	rs199539992	Intronic	0.0078	0.026	N/A	N/A	1.286	Polymorphism	Modifier
c.153+68A>T	94223931	*MRE11A*	-	rs116481709	Intronic	0.0048	0.015	N/A	N/A	2.976	Polymorphism	Modifier
c.942+29delA	47641560	*MSH2*	-	rs11309117	splicing	-	-	N/A	N/A	18.82	Disease-causing	Low
c.1277-67A>G	47672620	*MSH2*	-	rs114494667	Intronic	0.0024	0.0091	N/A	N/A	0.091	Polymorphism	Modifier
c.1759+53A>T	47698254	*MSH2*	-	rs55839838	Intronic	0.0012	0.0045	N/A	N/A	2.276	Polymorphism	Modifier
c.1759+107A>G	47698308	*MSH2*	-	rs3764959	Intronic	0.5	0.74	N/A	N/A	2.852	Polymorphism	Modifier
c.1760-126A>T	47702038	*MSH2*	-	rs17224815	Intronic	0.025	0.089	N/A	N/A	1.23	Polymorphism	Modifier
c.2005+84_2005+85insA	47702493	*MSH2*	-	rs367733465	Intronic	0.0012	0.0045	N/A	N/A	5.974	Polymorphism	Modifier
c.3439-67A>G	48031982	*MSH6*	-	rs187984197	Intronic	0.0016	0.003	N/A	N/A	5.023	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.1476+73C>T	45796115	*MUTYH*	-	rs3219495	Intronic	0.073	0.27	N/A	N/A	0.567	Polymorphism	Modifier
c.2185-122T>C	90949425	*NBN*	-	rs13312969	Intronic	0.052	0.042	N/A	N/A	5.732	Polymorphism	Modifier
c.2071-61A>T	90955655	*NBN*	-	rs3736640	Intronic	0.057	0.058	N/A	N/A	0.756	Polymorphism	Modifier
c.1125-79C>A	90967862	*NBN*	-	rs1805786	Intronic	0.38	0.3	N/A	N/A	1.972	Polymorphism	Modifier
c.1125-86G>A	90967869	*NBN*	-	rs13312922	Intronic	0.022	0.082	N/A	N/A	3.017	Polymorphism	Modifier
c.897-64A>G	90976799	*NBN*	-	rs13312891	Intronic	0.022	0.081	N/A	N/A	1.714	Polymorphism	Modifier
c.702+150G>T	90983251	*NBN*	-	rs1805827	Intronic	0.035	0.13	N/A	N/A	4.973	Polymorphism	Modifier
c.702+149T>C	90983252	*NBN*	-	rs3026271	Intronic	0.027	0.003	N/A	N/A	4.95	Polymorphism	Modifier
c.480+123A>G	90992839	*NBN*	-	rs146183296	Intronic	0.0026	0.0098	N/A	N/A	3.918	Polymorphism	Modifier
c.-118G>T	90996907	*NBN*	-	rs13312847	5'UTR variant	0.004	0.014	N/A	N/A	15.75	Polymorphism	Modifier
c.3201+101A>G	23625224	*PALB2*	-	rs249935	Intronic	0.15	0.24	N/A	N/A	1.608	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.3201+95C>T	23625230	*PALB2*	-	rs76918977	Intronic	0.026	0.042	N/A	N/A	2.674	Polymorphism	Modifier
c.2587-101C>G	23637819	*PALB2*	-	rs113037118	Intronic	0.0002	0.0008	N/A	N/A	1.401	Polymorphism	Modifier
c.2175-117T>C	6018444	*PMS2*	-	rs75973354	Intronic	0.12	0.023	N/A	N/A	1.072	Polymorphism	Modifier
c.1068G>A	6029507	*PMS2*	p.Lys356Lys	rs528499793	Silent	-	-	Tolerated	Benign	16.39	Disease-causing	Low
c.903+94G>A	6035071	*PMS2*	-	rs111435898	Intronic	-	-	N/A	N/A	8.77	Polymorphism	Modifier
c.164-108G>C	6043797	*PMS2*	-	rs12538294	Intronic	0.12	0.012	N/A	N/A	1.781	Polymorphism	Modifier
c.23+72C>T	6048556	*PMS2*	-	rs3735295	Intronic	0.23	0.34	N/A	N/A	6.799	Polymorphism	Modifier
c.6777369T>C	6777369	*PMS2CL*	-	rs35291651	ncRNA_exonic	0.24	0.25	N/A	N/A	2.31	N/A	Modifier
c.6785602C>T	6785602	*PMS2CL*	-	rs376842582	ncRNA_Intronic	0.0044	0.017	N/A	N/A	2.716	N/A	Modifier
c.6785668G>C	6785668	*PMS2CL*	-	rs189897489	ncRNA_Intronic	0.033	0.12	N/A	N/A	1.817	N/A	Modifier
c.6785797C>T	6785797	*PMS2CL*	-	rs141880038	ncRNA_exonic	-	-	N/A	N/A	7.977	N/A	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.6786771C>T	6786771	*PMS2CL*	-	rs200200461	ncRNA_exonic	0.0036	-	N/A	N/A	18.38	N/A	Modifier
c.6790900T>C	6790900	*PMS2CL*	-	rs200453156	ncRNA_exonic	-	-	N/A	N/A	4.289	N/A	Modifier
c.6791004G>C	6791004	*PMS2CL*	-	rs201078635	ncRNA_exonic	-	-	N/A	N/A	2.97	N/A	Modifier
c.802-53_802-16delTTAATTAAATATGTCATTTCATTTCTTTTTCTTTTCTT	89720635	*PTEN*	-	rs557364463	Intronic	-	-	N/A	N/A	11.28	Polymorphism	Modifier
c.2398-115A>G	131939497	*RAD50*	-	rs2706377	Intronic	-	-	N/A	N/A	0.467	Polymorphism	Modifier
c.1026+55A>G	56809960	*RAD51C*	-	rs573302185	Intronic	-	-	N/A	N/A	12.51	Polymorphism	Modifier
c.903+61A>G	33428159	*RAD51L3-RFFL*	-	rs9914109	ncRNA_Intronic	0.013	0.044	N/A	N/A	3.812	Polymorphism	Modifier
c.480+75T>G	33433932	*RAD51L3-RFFL*	-	rs8067688	ncRNA_Intronic	0.025	0.089	N/A	N/A	1.123	Polymorphism	Modifier
c.465-51T>C	1220321	*STK11*	-	rs2075606	Intronic	0.36	0.44	N/A	N/A	2.422	Polymorphism	Modifier
c.862+142C>A	1221481	*STK11*	-	rs572679790	Intronic	-	-	N/A	N/A	3	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	dbSNP ID	Mutation type	1000 Genomes	1000G MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.862+145C>T	1221484	*STK11*	-	rs741764	Intronic	0.29	0.19	N/A	N/A	0.152	Polymorphism	Modifier
c.1100+156G>A	7573771	*TP53*	-	rs373232559	Intronic	-	-	N/A	N/A	4.028	Polymorphism	Modifier
c.782+92T>G	7577407	*TP53*	-	rs12951053	Intronic	0.18	0.11	N/A	N/A	1.045	Polymorphism	Modifier
c.782+72C>T	7577427	*TP53*	-	rs12947788	Intronic	0.18	0.11	N/A	N/A	5.45	Polymorphism	Modifier
c.121+158T>G	152357628	*XRCC2*	-	rs3218475	Intronic	0.026	0.095	N/A	N/A	1.154	Polymorphism	Modifier
c.-69T>G	152373233	*XRCC2*	-	rs3218385	5'UTR variant	0.086	0.11	N/A	N/A	2.501	Polymorphism	Modifier
c.-88G>C	152373252	*XRCC2*	-	rs3218384	Upstream	0.15	0.051	N/A	N/A	8.506	N/A	Modifier
c.-169G>A	152373333	*XRCC2*	-	rs185284326	Upstream	0.0016	0.0061	N/A	N/A	2.118	N/A	Modifier

Appendix O: Variants previously reported in dbSNP, ExAC and 1000 Genomes

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.3284+16T>G	108143595	*ATM*	-	Intronic	rs376541962	-	-	1.651E-05	0.0002	N/A	N/A	4.986	Polymorphism	Modifier
c.3285-9delT	108150208	*ATM*	-	Intronic	rs397779715	-	-	0.1845	0.0413	N/A	N/A	7.431	Polymorphism	Modifier
c.3403-14_3403-13insA	108151708	*ATM*	-	Intronic	rs587781368	-	-	0.5675	-	N/A	N/A	0.114	Polymorphism	Modifier
c.497-48T>G	108114632	*ATM*	-	Intronic	rs2235001	0.032	0.069	0.0286	0.0713	N/A	N/A	0.205	Polymorphism	Modifier
c.5320-48A>G	108173532	*ATM*	-	Intronic	rs149592592	0.0018	0.0068	0.0002	0.0022	N/A	N/A	7.316	Disease-causing	Modifier
c.5762+26C>G	108178737	*ATM*	-	Intronic	rs78330259	0.0012	0.0123	0.0034	0.013	N/A	N/A	0.09	Polymorphism	Modifier
c.5918+22G>T	108181064	*ATM*	-	Intronic	rs368629884	-	-	0.0000334	0.0003	N/A	N/A	9.277	Polymorphism	Modifier
c.662+38T>C	108114883	*ATM*	-	Intronic	rs2235004	0.031	0.11	0.0081	0.0899	N/A	N/A	5.368	Polymorphism	Modifier
c.8010+30_8010+31insT	108204718	*ATM*	-	Intronic	rs397897623	0.018	0.066	0.0048	0.0545	N/A	N/A	11.62	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.901+25T>G	108115778	*ATM*	-	Intronic	rs1800735	0.062	0.22	0.0198	0.2109	N/A	N/A	5.574	Polymorphism	Modifier
g.149992_149992delA	108188267	*ATM*	-	Intronic	rs58978479	-	-	0.0022	0.0132	N/A	N/A	11.8	Polymorphism	Modifier
g.26406_26406delA	108119616	*ATM*	-	Intronic	rs35813860	-	-	0.3608	0.3275	N/A	N/A	4.959	Polymorphism	Modifier
g.48746_48746delT	108141956	*ATM*	-	Intronic	rs373881770	-	-	0.022	0.016	N/A	N/A	2.322	Polymorphism	Modifier
g.5250_5251delAA	108098461	*ATM*	-	Intronic	rs2066734	-	-	0.4253	0.1684	N/A	N/A	8.45	Polymorphism	Modifier
g.58487_58488insT	108151698	*ATM*	-	Intronic	rs3218691	0.06	0.21	0.0025	0.001	N/A	N/A	0.359	Polymorphism	Modifier
g.6895_6896insT	108100106	*ATM*	-	Intronic	rs397713128	0.48	0.36	0.4396	0.3647	N/A	N/A	1.116	Polymorphism	Modifier
c.1568+42T>C	215632164	*BARD1*	-	Intronic	rs5031010	0.043	0.048	0.0507	0.0466	N/A	N/A	2.332	Polymorphism	Modifier
c.158+46A>C	215674090	*BARD1*	-	Intronic	rs35933323	0.74	0.76	0.7387	0.7693	N/A	N/A	5.757	Polymorphism	Modifier
g.29182G>C	215645247	*BARD1*	-	Intronic	rs115533621	0.0014	0.0053	0.0004	0.0048	N/A	N/A	4.323	Polymorphism	Modifier
c.3210+42delT	91337624	*BLM*	-	Intronic	rs397897626	-	-	0.1306	0.1806	N/A	N/A	0.218	Polymorphism	Modifier
c.98+30T>C	91290750	*BLM*	-	Intronic	rs8030	0.0009	0.0078	0.002	0.011	N/A	N/A	2.694	Polymorphis	Modif

					187			8					m	ier
DNA nomenclature	**Genomic end position**	**Gene**	**Protein nomenclature**	**Mutation type**	**dbSNP**	**1000 Genomes MAF**	**1000 Genomes MAF (AFR)**	**ExAC MAF**	**ExAC MAF (AFR)**	**SIFT prediction**	**Polyphen**	**CADD**	**MutationTaster**	**VEP**
g.78875A>G	41243416	*BRCA1*	-	Intronic	rs45620639	0.0006	0.0023	0.0003	0.0022	N/A	N/A	0.271	Polymorphism	Modifier
c.6841+79_6841+82delAATT	32915414	*BRCA2*	-	Intronic	rs276174881	-	-	-	-	N/A	N/A	8.398	Polymorphism	Modifier
g.3588_3588delT	32893198	*BRCA2*	-	Intronic	rs276174878	-	-	0.0063	0.0042	N/A	N/A	5.898	Polymorphism	Modifier
c.2098-22T>C	59821974	*BRIP1*	-	Intronic	rs114505031	0.0082	0.029	0.0018	0.0207	N/A	N/A	2.393	Polymorphism	Modifier
c.2493-50G>T	59770923	*BRIP1*	-	Intronic	rs142108255	0.002	0.0068	0.0005	0.0055	N/A	N/A	2.67	Polymorphism	Modifier
c.2637A>G	59763465	*BRIP1*	p.Glu879Glu	Silent	rs4986765	0.82	0.94	0.7188	0.9048	Tolerated	Benign	5.226	Polymorphism	Low
c.2905+83T>A	59763114	*BRIP1*	-	Intronic	rs4988357	0.34	0.35	0.3937	0.406	N/A	N/A	15.37	Polymorphism	Modifier
c.3411T>C	59760996	*BRIP1*	p.Tyr1137Tyr	Silent	rs4986763	0.62	0.66	0.5941	0.6531	Tolerated	Possibly damaging	2.738	Polymorphism	Low
g.69743G>T	59871140	*BRIP1*	-	Intronic	rs150458297	0.0014	0.0053	0.0003	0.0033	N/A	N/A	4.246	Polymorphism	Modifier
g.86916C>A	59853967	*BRIP1*	-	Intronic	rs139659736	0.0068	0.026	0.0014	0.0162	N/A	N/A	2.136	Polymorphism	Modifier
c.1239C>T	68847317	*CDH1*	p.Tyr413Tyr	Silent	rs36074916	0.0012	0.0045	0.0002	0.0015	Tolerated	Benign	7.987	Polymorphism	Low

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.1712-37T>G	68855867	*CDH1*	-	Intronic	rs35831514	0.004	0.014	0.0017	0.0189	N/A	N/A	6.264	Polymorphism	Modifier
c.-2203G>C	68847443	*CDH1*	-	5'UTR variant	rs33950903	0.023	0.087	0.0071	0.0772	N/A	N/A	1.069	Polymorphism	Modifier
c.48+62_48+63insGCCCCAGCCCCGT	68771431	*CDH1*	-	Intronic	rs74406246	0.69	0.54	0.81	0.662	N/A	N/A	10.82	Polymorphism	Modifier
c.216-34G>A	84393475	*FAM175A*	-	Intronic	rs6535478	0.68	0.9	0.5871	0.8381	N/A	N/A	7.326	Polymorphism	Modifier
c.216-44T>C	84393485	*FAM175A*	-	Intronic	rs75295148	0.099	0.17	0.1191	0.1719	N/A	N/A	9.927	Polymorphism	Modifier
c.87+47_87+48insC	84406091	*FAM175A*	-	Intronic	rs139863124	0.11	0.1	0.052	0.1156	N/A	N/A	5.409	Polymorphism	Modifier
c.1699+171T>G	37107130	*LRRFIP2*	-	Intronic	rs9311151	0.046	0.16	0.0331	0.2455	N/A	N/A	3.955	Polymorphism	Modifier
c.*16G>A	64571790	*MEN1*	-	3'UTR variant	rs140924477	0.0002	0.0008	8.239E-05	0.0007	N/A	N/A	4.208	Polymorphism	Modifier
c.-23-16C>G	64577620	*MEN1*	-	Intronic	rs509606	0.17	0.1	0.3084	0.2095	N/A	N/A	4.443	Polymorphism	Modifier
c.*29_*32delinsG	37092176	*MLH1*	-	3'UTR variant	rs281875167	-	-	0.0179	0.0308	N/A	N/A	6.906	Polymorphism	Modifier
c.1039-27T>A	37067101	*MLH1*	-	Intronic	rs9862158	0.21	0.1	0.0199	0.025	N/A	N/A	4.583	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.1039-29A>T	37067099	*MLH1*	-	Intronic	rs6771325	-	-	0.0377	-	N/A	N/A	0.196	Polymorphism	Modifier
c.1039-31A>T	37067097	*MLH1*	-	Intronic	rs59684491	-	-	0.0613	0.1129	N/A	N/A	0.017	Polymorphism	Modifier
g.55749G>A	37090571	*MLH1*	-	Intronic	rs112877263	-	-	-	-	N/A	N/A	0.526	Polymorphism	Modifier
c.1099-26G>A	94197431	*MRE11A*	-	Intronic	rs115448201	0.0072	0.023	0.0021	0.0209	N/A	N/A	0.851	Polymorphism	Modifier
c.20+141G>A	94225807	*MRE11A*	-	Intronic	rs496797	0.55	0.64	0.5766	0.6693	N/A	N/A	12.26	Polymorphism	High
c.20+167_20+168insT	94225780	*MRE11A*	-	Intronic	rs397942556	0.042	0.11	0.0456	0.1399	N/A	N/A	7.065	Polymorphism	Modifier
c.315-15_315-14insT	94212931	*MRE11A*	-	Intronic	rs71036325	-	-	0.0392	0.0216	N/A	N/A	7.874	Polymorphism	Modifier
g.58103G>T	94168972	*MRE11A*	-	Intronic	rs144976924	0.0004	0.0015	0.0003	0.0034	N/A	N/A	5.009	Polymorphism	Modifier
c.2005+42G>T	47702451	*MSH2*	-	Intronic	rs17218446	-	-	0.0413	-	N/A	N/A	0.48	Polymorphism	Modifier
c.2005+51T>G	47702460	*MSH2*	-	Intronic	rs17218452	0.043	0.15	0.0119	0.1341	N/A	N/A	2.54	Polymorphism	Modifier
c.2005+52T>G	47702461	*MSH2*	-	Intronic	rs529442030	0.0036	0.012	0.0006	0.0067	N/A	N/A	2.045	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
g.11453_11465delAAAAAAAAAAAAA	47641572	*MSH2*	-	Splicing	rs574266201	-		0.0072	0.0448	N/A	N/A	17.72	Polymorphism	Low
g.5417_5417delT	47635524	*MSH2*	-	Intronic	rs201372136	-	-	0.1511	0.1073	N/A	N/A	6.696	Polymorphism	Modifier
c.2822C>T	47739567	*MSH2, KCNK12*	-	Intergenic	rs114185517	0.006	0.023	0.0014	0.0162	N/A	N/A	7.023	Polymorphism	Low
c.3557-17_3557-16insT	48032741	*MSH6*	-	Intronic	rs546821964	-	-	0.1386	0.2299	N/A	N/A	10.28	Polymorphism	Modifier
c.3646+35_3646+38delCTAT	48032878	*MSH6*	-	Intronic	rs3136358	-	-	0.6925	0.8045	N/A	N/A	4.323	Polymorphism	Modifier
c.4001+48_4001+49insAA	48033840	*MSH6*	-	Intronic	rs199687113	-	-	0.0043	0.0426	N/A	N/A	0.002	Polymorphism	Modifier
c.457+50_457+51delTG	48018295	*MSH6*	-	Intronic	rs397839804	-	-	0.0008	0.0015	N/A	N/A	5.35	Polymorphism	Modifier
g.99042_99042delA	48033891	*MSH6*	-	Intronic	rs67094025	-	-	0.1323	0.1439	N/A	N/A	0.004	Polymorphism	Modifier
c.1435-40C>G	45796269	*MUTYH*	-	Intronic	rs3219493	0.94	0.97	0.9245	0.9582	N/A	N/A	1.746	Polymorphism	Modifier
c.1397+46delA	90967465	*NBN*	-	Intronic	rs13312924	-	-	0.0452	0.0784	N/A	N/A	3.601	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
c.321-29G>A	90993150	*NBN*	-	Intronic	rs769417	0.023	0.083	0.0074	0.0806	N/A	N/A	5.544	Polymorphism	Modifier
c.37+11A>G	90996742	*NBN*	-	Intronic	rs115032431	0.0058	0.021	0.0015	0.0179	N/A	N/A	5.514	Polymorphism	Modifier
c.650+27A>C	90982565	*NBN*	-	Intronic	rs114182293	0.013	0.048	0.0043	0.0476	N/A	N/A	3.274	Polymorphism	Modifier
c.896+29_896+30insA	90982563	*NBN*	-	Intronic	rs13312882	0.044	0.16	0.0102	0.1139	N/A	N/A	1.411	Polymorphism	Modifier
c.896+36G>A	90982556	*NBN*	-	Intronic	rs1805826	0.074	0.003	0.0675	0.0101	N/A	N/A	4.751	Polymorphism	Modifier
g.47987_47988insA	90967470	*NBN*	-	Intronic	rs201631775	-	-	0.0059	0.0594	N/A	N/A	0.345	Polymorphism	Modifier
c.3113+28C>T	23632655	*PALB2*	-	Intronic	rs145973806	0.0002	0.0008	8.407E-06	0.00009711	N/A	N/A	2.001	Polymorphism	Modifier
c.3351-48G>T	23615038	*PALB2*	-	Intronic	rs201335445	-	-	0.0002	0	N/A	N/A	5.342	Polymorphism	Modifier
c.3351-66delT	23615043	*PALB2*	-	Intronic	rs397855890	-	-	0.4073	0.375	N/A	N/A	4.537	Polymorphism	Modifier
c.6776903A>G	6776903	*PMS2CL*	-	ncRNA_exonic	rs185608207	0.0004	0.0015	5.781E-05	0.0007	N/A	N/A	4.246	N/A	Modifier
c.6776906A>G	6776906	*PMS2CL*	-	ncRNA_exonic	rs7804416	0.12	0.33	0.0043	0.0278	N/A	N/A	5.166	N/A	Modifier

c.6776956A>G	6776956	*PMS2CL*	-	ncRNA_exonic	rs7804542	0.28	0.33	0.2296	0.319	N/A	N/A	3.718	N/A	Modifier
c.6777003C>T	6777003	*PMS2CL*	-	ncRNA_exonic	rs375311348	-	-	0.0000412	0.0004	N/A	N/A	7.675	N/A	Modifier
c.6777025C>T	6777025	*PMS2CL*	-	ncRNA_exonic	rs180751212	0.0042	0.014	0.0018	0.014	N/A	N/A	9.5	N/A	Modifier
c.6777043C>T	6777043	*PMS2CL*	-	ncRNA_exonic	rs7786232	0.059	0.15	0.0245	0.1262	N/A	N/A	12.08	N/A	Modifier
g.6777218T>C	6777218	*PMS2CL*	-	ncRNA_exonic	rs11974014	0.12	0.32	0.0561	0.244	N/A	N/A	1.631	N/A	Modifier
g.6777236C>T	6777236	*PMS2CL*	-	ncRNA_exonic	rs112753805	-	-	8.318E-06	-	N/A	N/A	1.753	N/A	Modifier
c.254-30delT	89692732	*PTEN*	-	Intronic	rs77494260	0.0154	0.0023	-	-	N/A	N/A	1.31	Polymorphism	Modifier
c.1245+23T>C	131924595	*RAD50*	-	Intronic	rs115100130	0.0088	0.032	0.0028	0.0288	N/A	N/A	2.041	Polymorphism	Modifier
c.1246-42T>A	131925281	*RAD50*	-	Intronic	rs114648433	0.0058	0.022	0.0023	0.0253	N/A	N/A	5.577	Polymorphism	Modifier
c.1969+30A>C	131930766	*RAD50*	-	Intronic	rs2522390	0.03	0.11	0.0085	0.0951	N/A	N/A	2.217	Polymorphism	Modifier
c.2923-10_2923-9insT	131944964	*RAD50*	-	Intronic	rs587781268	0.073	0.25	0.024	0.2395	N/A	N/A	14.16	Polymorphism	Modifier
g.32654_32654delT	131924364	*RAD50*	-	Intronic	rs201800554	-	-	0.0022	0.0017	N/A	N/A	1.754	Polymorphism	Modifier
g.36038T>C	131927748	*RAD50*	-	Intronic	rs10520116	0.013	0.048	0.0035	0.039	N/A	N/A	6.309	Polymorphism	Modifier

DNA nomenclature	Genomic end position	Gene	Protein nomenclature	Mutation type	dbSNP	1000 Genomes MAF	1000 Genomes MAF (AFR)	ExAC MAF	ExAC MAF (AFR)	SIFT prediction	Polyphen	CADD	MutationTaster	VEP
g.53237T>C	131944947	*RAD50*	-	Intronic	rs113053126	0.0026	0.0098	0.0006	0.0072	N/A	N/A	12.05	Polymorphism	Modifier
g.53411C>T	131945121	*RAD50*	-	Intronic	rs10054336	0.054	0.19	0.0171	0.1851	N/A	N/A	9.316	Polymorphism	Modifier
c.904+34T>C	56798207	*RAD51C*	-	Intronic	rs28363318	0.21	0.26	0.28	0.2744	N/A	N/A	1.418	Polymorphism	Modifier
c.278C>T	33444042	*RAD51D*	p.His53His	Silent	rs200538950	0.0004	0.0008	0.0002	0.0016	Deleterious	N/A	0.074	Polymorphism	Moderate
g.20348G>A	33428194	*RAD51L3-RFFL*	-	ncRNA_intronic	rs111462110	-	-	1.649E-05	0.00009649	N/A	N/A	5.322	Polymorphism	Modifier
c.291-32C>T	1218384	*STK11*	-	Intronic	rs57281474	0.043	0.16	0.0116	0.1366	N/A	N/A	3.358	Polymorphism	Modifier
c.464+39_464+40insG	1219452	*STK11*	-	Intronic	rs544282452	0.015	0.05	0.0024	0.0288	N/A	N/A	3.215	Polymorphism	Modifier
c.734+41T>C	1220757	*STK11*	-	Intronic	rs115624397	0.025	0.09	0.0067	0.0811	N/A	N/A	0.069	Polymorphism	Modifier
c.783-33T>C	7577188	*TP53*	-	Intronic	rs113302588	-	-	0.0002	0.0019	N/A	N/A	5.277	Polymorphism	Modifier
c.*1T>C	152345726	*XRCC2*	-	3'UTR variant	rs3218539	0.052	0.19	0.0168	0.1816	N/A	N/A	12.68	Polymorphism	Modifier