UFS·UV
UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

This work is licensed under a Creative Commons Attribution-NonCommercial-ShareAlike 4.0 International License.

How to cite this thesis / dissertation (APA referencing method):

Surname, Initial(s). (Date). *Title of doctoral thesis* (Doctoral thesis). Retrieved from http://scholar.ufs.ac.za/*rest of thesis URL on KovsieScholar*

Surname, Initial(s). (Date). *Title of master's dissertation* (Master's dissertation). Retrieved from http://scholar.ufs.ac.za/*rest of thesis URL on KovsieScholar*

Screening for PALB2 mutations in South African women with BRCA negative familial breast cancer

By

Mpoi F Makhetha

2010079537

Submitted in fulfilment of the requirements in respect of
the Magister Scientiae in Medical Science degree qualification
(M.Med.Sc) in the Faculty of Health Sciences, Division of Human
Genetics, University of the Free State, Bloemfontein

Supervisor: Dr NC van der Merwe

Co-supervisor: BK Dajee

February 2018

Declaration

I declare that the master's research dissertation or interrelated, publishable manuscripts/published articles that I herewith submit at the University of the Free State, is my independent work and that I have not previously submitted it for a qualification at another institution of higher education.

I hereby declare that I am aware that the copyright is vested in the University of the Free State.

I hereby declare that all royalties as regards to intellectual property that was developed during the course of and/or in connection with the study at the University of the Free State, will accrue to the University.

_____

Mpoi Makhetha

Dedication

*This dissertation is dedicated to **my immediate family** for their support throughout my studies. Without you, it would not have been possible.*
*Most especially to my sister **Mamello Makhetha** (and other little girls), to inspire you to be whatever it is your mind can imagine. It is indeed true that girls are made of sugar and spice and everything nice. So sprinkle a little bit of yourself in whatever you do and you will see the sparkle in everything around you.*

Acknowledgements

The success of this study could not have been possible without the contribution of several individuals and institutions. I would like to express my deepest gratitude to the following:

My study-leaders: **Dr NC van der Merwe and BK Dajee**
For your patient guidance, encouragement and advice provided throughout my time as a student. I have been extremely lucky to have a supervisor who cared so much about my work, and who responded to my questions and queries so patiently. I am truly grateful.

The **breast cancer patients**
It is through your participation that this study was possible.

The various **Institutions**
The Division of Human Genetics (UFS) and the National Health Laboratory Services for resources and facilities, and the National Research Foundation for financial support towards my studies.

My colleagues
At the Department of Human Genetics for their contribution to my understanding of the field and their assistance to the success of this project.

All **my friends**
For walking this journey with me, it is your unwavering support and words of encouragement that carried me through.

My loving **family**
For the endless love and many years of support, I will forever be indebted to you.

To the **Almighty God**
For the unmerited favour; the amazing grace, and the fulfilment of all promises.
Ebenezer Father, You have always been with me till this far.

Summary

PALB2 (Partner and Localizer of BRCA2) plays a key role in the repair of damaged DNA by localizing BRCA2 and initiating the repair process. The gene has been recognized as a third breast cancer (BC) predisposition gene, together with *BRCA1* and *BRCA2*. The absolute risk for the development of BC for mutation carriers with a familial predisposition, represents 58% at 50 years of age, similar to that of *BRCA2*. Although germline loss-of-function *PALB2* mutations is rare, the mutation spectrum internationally and in South African (SA) is still unknown.

Eighty-six SA breast- and/or ovarian cancer (OVC) patients were selected. Patients had to be affected with either breast- and/or OVC, have a positive family history of BC (two or more affected family members, excluding the index), tested negative for pathogenic mutations in *BRCA1/2*, represent one of the four main SA ethnic groups and had sufficient amounts of good quality DNA stored to complete a comprehensive screen of the entire gene. The complete coding sequence together with the intron-exon boundaries were screened using qPCR-based high resolution melt analysis (HRMA). All samples deviating from the baseline was Sanger sequenced. *In silico* analysis using a combination of prediction programmes was performed for single base changes present in the coding regions of the gene.

A total of 20 variants was identified, of which two represented loss-of-function mutations (*PALB2* c.424A>T, p.Lys142Ter and *PALB2* c.508A>T, p.Arg170Ter). These mutations were situated in exon 4. Although the first mutation has been detected before, *PALB2* c.508A>T, p.Arg170Ter was novel. These mutations were detected for the SA Indian (single patient), Coloured (two patients) and Afrikaner (single patient) populations. The two patients carrying these Class 5 pathogenic mutations had an extensive family history of BC and other ca types.

Two of the 18 missense variants [*PALB2* c.2794G>A, p.Ala915Thr (rs374736398) and *PALB2* c.3434G>T, p.Gly1145Val (novel variant)] were classified as putative Class 4 variants based on their location in the carboxy-terminal WD-40 domain, the binding site for the N-terminus of BRCA2. *PALB2* c.2794G>A, p.Ala915Thr was detected for three SA patients (SA Indian and White population), whereas the novel *PALB2* c.3434G>T, p.Gly1145Val variant was present in a single

Coloured BC patient. In the families of the patients carrying either one of the putative Class 4 variants, BC was the dominant ca type observed. Interpretation of rare missense variants such as *PALB2* c.3434G>T, p.Gly1145Val regarding impact on clinical treatment remains challenging, as no rare missense variant to date has been used to determine the clinical management of patients carrying these variants.

The majority of *PALB2* variants was observed for the Coloured and SA Indian populations, whereas for the other three groups, it was limited. This observation could have been due to the unique admixture of these population groups present in SA, as SA has a multi-faceted colonisation history based on the country's location with respect to major historical trade routes. Another explanation could be the higher percentage of representation of these two population groups in the cohort, as 33.7% of the cohort was Coloured with 25.6% self-identified as SA Indian BC/OVC patients. Although 17.4% of the cohort represented Black patients, the number of variants observed for this group was very low. This indicated that the prevalence of *PALB2* mutations/variants might be different for each of the SA population groups, depending on the genetic heritage of these patients. For this reason, the cohort needs to be increased in order to determine the mutation spectrum. Once this has been determined, case-control studies should be performed to determine the relative risk involved with each variant regarding the development of breast- and other ca types in SA.

Keywords: familial breast cancer, HRMA, mutation screening, *PALB2*, Sanger sequencing, South Africa, computer-based analysis

Table of contents

Summary i
List of figures vi
List of tables ix
Abbreviations and symbols x
Chapter 1: Introduction 1
Chapter 2: Literature review
2.1 Introduction 3
2.1.1 Methods applied for literature searches 3
2.2 Overview of Breast Cancer 4
2.3 Background to the incidence of BC 5
2.4 Genetics of familial BC 5
2.4.1 Breast Cancer Susceptibility gene 1 6
2.4.2 Breast Cancer Susceptibility gene 2 6
2.4.3 Partner and localizer of BRCA2 8
2.4.3.1 Structure of PALB2 protein 9
2.4.3.2 Function of PALB2 11
2.4.3.3 Mechanism of action 11
2.4.3.4 *PALB2* mutations and related diseases 13
2.5 Previous findings 14
2.6 South African studies 21
2.7 Experimental methods 22
2.7.1 qPCR based HRMA 22
2.7.2 DNA Sanger sequencing 24
2.8 Summary 25
2.9 Aim of the study 26
2.10 Objectives 25
Chapter 3: Methodology
3.1 Introduction 27
3.2 Study design 27
3.3 Sampling procedure 28
3.3.1 Samples used for optimisation of conventional and qPCR 28

3.3.2 Research participants 28
3.4 Ethical considerations 33
3.5 Methods 35
3.5.1 DNA extraction 35
3.5.2 DNA quantity and quality 37
3.5.3 Primer sets for *PALB2* 37
3.5.4 Conventional PCR reactions 38
3.5.5 qPCR for HRMA 40
3.5.6 HRMA 40
3.5.7 Automated sanger sequencing 41
3.5.8 Quality control 42
3.5.9 Data analysis 43
Chapter 4: Results and discussion
4.1 Familial breast and/or ovarian cancer patients 45
4.1.1 Various cancer types in families 47
4.1.2 Affected family members 51
4.1.3 Age of diagnosis 53
4.2 *PALB2* primer optimisation 55
4.3 *PALB2* mutation screening results 61
4.3.1 Classification of *PALB2* mutations 61
4.3.2 Mutation screening results for *PALB2* 64
4.3.2.1 Truncation mutations 64
4.3.2.1.1 *PALB2* c.424A>T, p.Lys142* (rs587782005) 64
4.3.2.1.2 *PALB2* c.508A>T, p.Arg170* (novel) 75
4.3.2.2 Missense mutations 79
4.3.2.2.1 *PALB2* c.680C>A, p.Ala227Asp (novel) 79
4.3.2.2.2 *PALB2* c.1395G>A, p.Met465Ile 82
(rs746805049) 82
4.3.2.2.3 *PALB2* c.1508C>A, p.Ala503Glu (novel) 83
4.3.2.2.4 *PALB2* c.2393C>T, p.Pro798Leu (novel) 83
4.3.2.2.5 *PALB2* c.2411, p.Ser804Tyr (novel) 84
4.3.2.2.6 *PALB2* c.2794G>A, p.Val932Met 86
(rs34624036)

4.3.2.2.7 *PALB2* c.2968G>A, p.Glu990Lys (novel) 97
4.3.2.2.8 *PALB2* c.3434, p.Gly1145Val (novel) 98
4.4 Conclusion 102
Chapter 5: Conclusion and recommendations 105
Chapter 6: References 108
6.1 Books and articles 108
6.2 Electronic resources 121
6.3 Databases 122
Appendix A 124
Appendix B 125
Appendix C 126
Appendix D 127
Appendix E 128
Appendix F 129
Appendix G 130
Appendix H 131
Appendix I 135

List of figures

Figure 2.1	Graphical representation of the mean cumulative female risk of developing BC and associated confidence intervals (Antoniou *et al.,* 2014).	7
Figure 2.2	Structure of the PALB2 protein showing the N-terminal, the central region and the C-terminus (Park *et al.,* 2014).	10
Figure 2.3	A demonstration of the roles of PALB2 in homologous recombination (Pauty *et al.,* 2014).	12
Figure 2.4	Schematic representation of known *PALB2* mutations distributed across the coding region of the gene, and the associated influences on the PALB2 protein (Pauly *et al.* 2014).	16
Figure 3.1	Overview of the methodology involved in data collection.	36
Figure 4.1	Inspection of the quantity and quality of the DNA for CAM2792 as determined by the NanoDrop®ND-100 Spectrophotometer v3.01.	46
Figure 4.2	Example of a family pedigree that met the selection criteria for the current study.	48
Figure 4.3	Different cancer types and their prevalence as recorded in the family pedigrees of the 86 participants.	49
Figure 4.4	Prevalence of various cancer types present in the families involved, recorded for each SA ethnic group.	50
Figure 4.5	Family history of patients in different ethnic groups.	52
Figure 4.6	Average age at diagnosis illustrated per ethnic group for the total 86 index patients affected with breast- or OVC.	54
Figure 4.7	Agarose visualization of optimized PCR products for various *PALB2* primer sets.	57
Figure 4.8	Temperature gradient optimisation of the annealing temperature for *PALB2* exon 13_1 used within qPCR based HRMA.	59
Figure 4.9	Evidence of more than one melting domains within *PALB2* exon 13_2.	60

Figure 4.10	Graphs utilized for the visualization of HRMA results for *PALB2*.	62
Figure 4.11	A schematic representation of *PALB2* with all the single nucleotide variants detected during this study superimposed on the exonic structure. The functional and structural domains of the protein are also indicated (Antoniou *et al.,* 2014).	69
Figure 4.12	HRMA results for *PALB2* exon 4 for 23 familial BC patients, each performed in duplicate, of which a single patient (CAM 2338) carried the PALB2 c.424A>T, p.Lys142Ter (rs587782005) mutation.	71
Figure 4.13	Family pedigree for CAM2338 indicating a positive paternal family history of BC and OVC in this Indian family.	72
Figure 4.14	Family pedigree for CAM2735 indicating a positive maternal family history of BC and OVC in this Afrikaner family.	73
Figure 4.15	HRMA results for *PALB2* exon 4 for 23 familial BC patients, each performed in duplicate, of which two patients (CAM 2411 and CAM2755) carried the *PALB2* c.508A>T, p.Arg170Ter* novel mutation.	76
Figure 4.16	Family pedigree for CAM2411, indicating a positive maternal family history of breast- and other cancer types in this Coloured family.	77
Figure 4.17	Family pedigree for CAM2755, indicating a positive paternal family history of mostly BC.	78
Figure 4.18	HRMA results for *PALB2* exon 4 for 23 familial BC patients, each performed in duplicate, of which two patients (CAM 2411 and CAM2987) carried the *PALB2* c.680C>A, p.Ala227Asp novel mutation.	80
Figure 4.19	*In silico* results for the *PALB2* c.680C>A, p.Ala227Asp novel mutation using PolyPhen-2.	81

Figure 4.20	PolyPhen-2 prediction results for the *PALB2* c.2393C>T, p.Pro798Leu novel mutation.	85
Figure 4.21	HRMA results for *PALB2* exon 5 for 23 familial BC patients, each performed in duplicate, of which two patients (CAM 2820 and CAM2821) carried the *PALB2* c.2411C>A, p.Ser804Tyr novel mutation.	87
Figure 4.22	Family pedigrees for two familial BC patients identified with the *PALB2* c.2411C>A, p.Ser804Tyr novel mutation.	88
Figure 4.23	PolyPhen-2 prediction results for the *PALB2* c.2411C>A, p. Ser804Tyr novel mutation.	89
Figure 4.24	HRMA results for *PALB2* exon 8 for 23 familial BC patients, each performed in duplicate, of which a single patient (CAM2854) carried the *PALB2* c.2794G>A, p.Val932Met (rs34624036) mutation.	90
Figure 4.25	Family pedigrees for two familial BC patients identified with the *PALB2* c.2794G>A, p.Val932Met (rs34624036) mutation.	92
Figure 4.26	Family pedigree for CAM2854, indicating a positive maternal family history of BC and OVC in this White family.	93
Figure 4.27	PolyPhen-2 prediction results for *PALB2* c.2794G>A, p.Val932Met.	94
Figure 4.28	PolyPhen-2 prediction results for *PALB2* c.2968G>A, p.Glu990Lys.	99
Figure 4.29	HRMA results for *PALB2* exon 13 for 23 familial BC patients, each performed in duplicate, of which a single patient (CAM2578) carried the *PALB2* c.3434G>T, p.Gly1145Val novel mutation.	100
Figure 4.30	Family pedigree for CAM2578, indicating a positive paternal family history of BC in this Coloured family.	101
Figure 4.31	PolyPhen-2 prediction results for *PALB2* c.3434G>T, p.Gly1145Val novel mutation.	103

List of tables

Table 2.1	*PALB2* mutations published in international literature.	17
Table 3.1	List of Indian patients included in the study.	29
Table 3.2	List of Black patients included in the study.	30
Table 3.3	List of Coloured patients included in the study.	31
Table 3.4	List of Afrikaner patients included in the study.	32
Table 3.5	List of White (non-Afrikaner) patients included in the study.	34
Table 3.6	Oligonucleotides used for qPCR based HRMA of *PALB2.*	38
Table 4.1	Optimized annealing temperatures for qPCR based HRMA of *PALB2*.	56
Table 4.2	*PALB2* variants identified in SA population and designated classification using different global databases	65
Table 4.3	*PALB2* variants identified in SA population and the number of patients for whom they were been detected, their ethnicity and age at diagnosis.	67

Abbreviations and symbols

~	Approximately
‘	Prime
aa	Amino acid
BC	Breast cancer
bp	Base pair
BRCA1	Breast cancer susceptibility gene 1
BRCA2	Breast cancer Susceptibility gene 2
BRIP1	BRCA Interacting Protein C-terminal Helicase 1
ca	Cancer
ChAM	Chromatin Association Motif
Cp	Crossing point
ddNTP	Dideoxynucleotidetriphosphate
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotidetriphosphate
DSB	Double Stranded Break
dsDNA	Double stranded DNA
dx	Age of diagnosis
EDTA	Ethylenediaminetetraacetic acid
ENIGMA	Variant classification breast cancer genes
ER	Oestrogen receptor
ESE	Exonic splicing enhancer
FA	Fanconi anaemia
FoHS	Faculty of Health Science
HER2	HER2/NEU receptors
HGVS	Human Genome Variation Society
HR	Homologous recombination
HRMA	High Resolution Melting Analysis
IR	Ionizing radiation
Kb	Kilo bases
KD	Kilo Dalton

KEAP1	Kelch-like ECH-associated protein 1
MAF	Minor allele frequencies
Mb	Megabases
min	Minutes
MRG15	MRG15 recombinant protein
NHLS	National Health Laboratory Services
NRF2	Nuclear factor (erythroid-derived 2)-like 2,
OVC	Ovarian cancer
PALB2	Partner and Localiser of BRCA2
PCR	Polymerase chain reaction
Pol η	DNA polymerase eta
PR	Progesterone receptors
PRC2	Polycomb Repressive Complex 2
qPCR	Real time PCR
RAD51	RAD51 recombinase gene
SA	South Africa
SASHG	South African Society of Human Genetics
SDS	Sodium dodecyl sulphate
SNP	Single nucleotide polymorphism
SNV	single nucleotide variations
ssDNA	Single stranded DNA
Ta	Annealing temperature
Tm	Melting temperature
UFS	University of Free State
USA	United State of America
VUS	Variant of unknown clinical significance
WD-40 repeats	beta-transducin repeat
XRCC2	X-ray Repair Complimenting Defective Repair In Chinese Hamster Cells-2

Chapter 1

Introduction

Breast cancer (BC) is a pandemic disease that has been identified as one of the leading causes of mortality in women (Klug *et al.,* 2006; Gunnarsson *et al*., 2008). Despite numerous attempts to find better treatments and cure, many women are diagnosed globally every year. In the United States of America (USA), the number of new cases increased by ~ 27% between the year 2000 and 2014 (Greenlee *et al*., 2000; Jemal *et al*., 2010; Siegel *et al*., 2014). Similarly, BC is becoming increasingly common in Africa and other countries as it currently represents the most commonly diagnosed non-dermatological cancer (ca) among women worldwide (Gunnarsson *et al*., 2008). The causes of ca are mainly environmental, biological and physiological factors, accounting for ~ 80% of all BC cases (Klug *et al.,* 2006).

Genetics on the other hand are attributed to ~ 15% of all BC cases (Foulkes *et al.,* 2007; Gunnarsson *et al*., 2008), whereby the high penetrance *BRCA1* and *BRCA2* have been screened to diagnose the disease (Thorslund *et al*., 2010; Roy *et al*., 2012; Anjum *et al*., 2014). Screening of these genes has contributed to more appropriate treatment and management of the disease as well as allowed for pre-symptomatic testing (Thompson & Easton, 2004; Antoniou *et al.,* 2007; Foulkes *et al*., 2007). However, these genes only account for 15 – 20% of familial BC cases, leaving a big proportion undiagnosed (Antoniou *et al.,* 2007; Foulkes *et al*., 2007). In recent years, a number of genes have been identified that have the potential to contribute to the disease with low, medium and recently, high penetrance. Amongst them is *PALB2*, the third gene to impact on familial BC risk (Xia *et al*., 2006; Erkko *et al*., 2007; Rahman *et al*., 2007).

PALB2 protein plays a key role in the repair of damaged DNA by localizing BRCA2 and initiating the repair process (Xia *et al.,* 2006; Pauty *et al.,* 2014; Leyton *et al.,* 2015). Mutations in *PALB2* have been associated with a range of diseases: initially Fanconi anaemia (FA) in the bi-allelic form, and recently with BC in mono-

allelic form (Erkko *et al*., 2008; Wang & Smogorzewska, 2015; Majumbder & Chatterjee, 2017). Other diseases include familial pancreatic-, prostate and male BC (Vietri *et al.,* 2015; Reddy *et al*., 2016).

Although *PALB2* has not been studied well in many countries, mutations in the gene have been found to carry the same risk as *BRCA2*, consequently classifying it as a moderate to high penetrance gene (Hamosh, 2016; Majumbder & Chatterjee, 2017). For this reason it is also included in the BC gene panel in some countries such as Finland (Balia *et al*., 2010; Haanpaa *et al.,* 2013; Majumbder & Chatterjee, 2017). In South Africa (SA), *PALB2* studies have not yielded much results with regards to either the incidence or the impact of mutations on the genetically diverse population. Only two studies have been conducted and resulted in the identification of a single novel mutation, detected in a Caucasian woman from a cohort of White patients only (Sluiter *et al*., 2009; Francies *et al.,* 2015).

The objective of this study was therefore to screen *PALB2* in high risk SA *BRCA1* and *BRCA2* negative BC patients for possible mutations. The study hoped to provide insight into the spectrum of *PALB2* variants found for different SA ethnic groups and the extent to which those variants may contribute to the BC and OVC burden in SA.

Chapter 2

Literature review

2.1 Introduction

The function of this chapter is to build a knowledge base for research by evaluating relevant literature on the genetics of familial BC. The aim of the chapter is to provide broader understanding on the genetics of BC, especially the Partner and Localizer of BRCA2 gene (*PALB2* MIM_610355) gene, to present work that has already been done, and to formulate research questions based on the gaps that have been identified (Taylor & Medina, 2013).

2.1.1 Method applied for literature searches

A systematic literature review was applied. This method does not only evaluate the knowledge existing in literature, but it also investigates the strategies used to gather knowledge and how the information was searched (Taylor & Medina, 2013).

Only scholarly articles found on Google Scholar (https://scholar.google.co.za) were used to gather background knowledge on the topic. Searches were focused on the following: familial BC, Breast cancer susceptibility gene 1 (*BRCA1* MIM_113705), Breast cancer susceptibility gene 2 (*BRCA2* MIM_600185) as well as *PALB2*. For *PALB2*, searches were restricted to 3 areas which are the definition of the gene, the protein, its protein function, and thirdly, the presence and impact of *PALB2* mutations in different populations. The articles on *PALB2* were further restricted to the period between 2006 (year during which the gene was discovered and defined by Xia *et al.,* 2006) and 2018 (year during which this dissertation was presented).

For analysis on population studies, the following areas were key: primarily, the cohort, where the patients' age of diagnosis (dx), the *BRCA1/2* mutations status,

family history, and the ethnicity of patients were highlighted. These factors were important as they coincided with the focus of the current research. The mutations detected in the populations were also taken into consideration. This included their pathogenicity, whether the mutations were novel or had previously been detected in other population groups and whether the mutations were associated with BC in that population group. Lastly, the incidence of mutations in various populations.

This information was applied to the understanding of *PALB2* and the protein it encodes and to confirm the presence and incidence of, or the absence of *PALB2* mutations in populations across the world. Furthermore, to confirm or deny the association of *PALB2* mutations with BC, to investigate the clinical significance of the identified variations, and, to identify the interventions from available knowledge.

2.2 Overview of breast cancer

Similar to most health conditions, BC is a multi-factorial trait and can be classified as either sporadic or familial (Klug *et al*., 2006). Sporadic ca is believed to arise from gene damage acquired from environmental factors such as exposure to large amounts of radiation, tobacco and lifestyle choices (de Martel *et al*., 2012). Biological factors such as virus infections (for instance hepatitis B), bacteria (such as *Helicobacter pylori*), parasites (like candida) and other diseases (such as human immunodeficiency virus infection and acquired immune deficiency syndrome (HIV/AIDS)) that result in immunosuppression also predispose individuals to ca (de Martel *et al*., 2012).

Other influences include hormones and normal aging (Bray *et al*., 2012; de Martel *et al.,* 2012). Together, these factors account for the majority (~ 80%) of all BC cases (Klug *et al.,* 2006). Most of these acquired mutations are neither shared amongst relatives nor passed on to the next generation (van der Groep *et al*., 2006). The mutation driven tumours that develop in these cases are spontaneous and the patients in most cases do not have a positive family history of the disease (Klug *et al*., 2006).

On the other hand, 10 – 15% of BC cases are associated with a genetic predisposition. Research has shown that mutations in predisposing genes in familial cases are passed from generation to generation through Mendelian autosomal

dominant inheritance (Claus *et al*., 1996; Foulkes *et al.,* 2007; Gunnarsson *et al*., 2008; Easton *et al.,* 2015). It has been stated that in Westernised countries, up to 20% of women diagnosed with BC have at least one affected relative (Verhoog *et al.,* 1998; Antoniou *et al.,* 2014). This implies that individuals with a family history of breast and ovarian ca (OVC) are more likely to develop familial BC compared to those without a family history (Lichtenstein *et al.,* 2000; Haanpaa *et al*., 2013).

2.3 Background to the incidence of BC

A significant increase in new BC cases has been observed over the years. In the United States of America (USA), 182 800 new BC cases were recorded in 2000, with 207 090 recorded in 2010 and a total of 232 670 new cases in 2014 (Greenlee *et al*., 2000; Jemal *et al*., 2010; Siegel *et al*., 2014). This increase is not only seen in the USA but worldwide, hence BC became the most commonly diagnosed non-dermatological ca amongst women worldwide (Gunnarsson *et al*., 2008). The disease has also been identified as the second leading cause of ca mortality in women after lung ca (Ferlay *et al*., 2010). The cumulative risk of developing BC is indicated in Figure 2.1, where the graph shows that the female risk of developing BC increases significantly with age, especially from the age of 40 years, (Frank, 2004; Erkko *et al*., 2008; Antoniou *et al*., 2014).

2.4 Genetics of familial BC

For a long time, the diagnosis of familial BC has been based mainly on germline mutations present in two (2) high risk BC genes, namely *BRCA1* and *BRCA2.* These genes account for ~ 20% of all familial BC cases (Thompson & Easton, 2004; Antoniou *et al.,* 2007; Foulkes *et al*., 2007). In patients with four or more BC or OVC affected family members, the odds of detecting a pathogenic *BRCA* mutation increases to ~ 85% (Foulkes *et al*., 2007). As a result, *BRCA1* and *BRCA2* were classified as high penetrance genes and are the most clinically significant genes as far as familial BC is concerned (Wooster *et al*., 1995; Erkko *et al*., 2008; Siegel *et al*., 2014; Vietri *et al.,* 2015).

2.4.1 Breast cancer susceptibility gene 1

BRCA1 is a tumour suppressor gene which functions as a key negative modulator of Polycomb repressive complex 2 (PRC2 MIM_601573), a protein that initiates epigenetic silencing of genes involved in cell fate decisions (Ernst *et al.,* 2010). It is located on chromosome 17q21 and encodes for the BRCA1 protein. The protein plays a major role in promoting the concentration of PALB2 and BRCA2 to sites of Deoxyribonucleic acid (DNA) damage (Xia *et al*., 2006; Zhang *et al*., 2009a). The protein also prevents abnormal proliferation and alteration of the cells during the repair process (Xia *et al*., 2006; Zhang *et al*., 2009a; Anjum *et al*., 2014).

Pathogenic mutations in *BRCA1* account for ~ 40 – 50% of the site-specific familial BC cases (Verhoog *et al*., 1998; Xia *et al*., 2007). These mutations carry an ~ 65% increase in risk for developing BC in females before the age of 70 years (Antoniou *et al.,* 2003; Thompson & Easton, 2004; Erkko *et al.,* 2008; Anjum *et al.,* 2014). Clinically, these mutations tend to result in early-onset BC which is, in most cases, more aggressive, frequently bilateral and have worse histo-prognostic features such as a high degree of aneuploidy, higher grade tumours and higher proliferation indices (Verhoog *et al*., 1998; Wang, 2007; Anjum *et al*., 2014).

2.4.2 Breast cancer susceptibility gene 2

BRCA2, which encodes for the BRCA2 protein, is located on chromosome 13q12 – q13 (Wooster *et al*., 1994; Haley, 2016). The protein is essential for homologous recombination (HR) repair where it recruits and directs the RAD51 recombinase (MIM_179617) to the site of the double stranded break (DSB) (Jensen *et al*., 2010; Liu *et al*., 2010; Thorslund *et al*., 2010; Roy *et al*., 2012).

BRCA2 pathogenic mutations account for 5 – 7% of all BC cases and are associated with ~ 45% of *BRCA*-related BC cases for females under the age of 70 years (Antoniou *et al.,* 2003; Thompson & Easton, 2004; Xia *et al*., 2007; Erkko *et al.,* 2008). These mutations often represent late-onset BCs (Foulkes, 2008; Roy *et al*., 2012). Additionally, bi-allelic mutations in this gene are known to predispose individuals to Fanconi anaemia (FA) (Moldovan & D'Andrea, 2009; Antoniou *et al*., 2014).

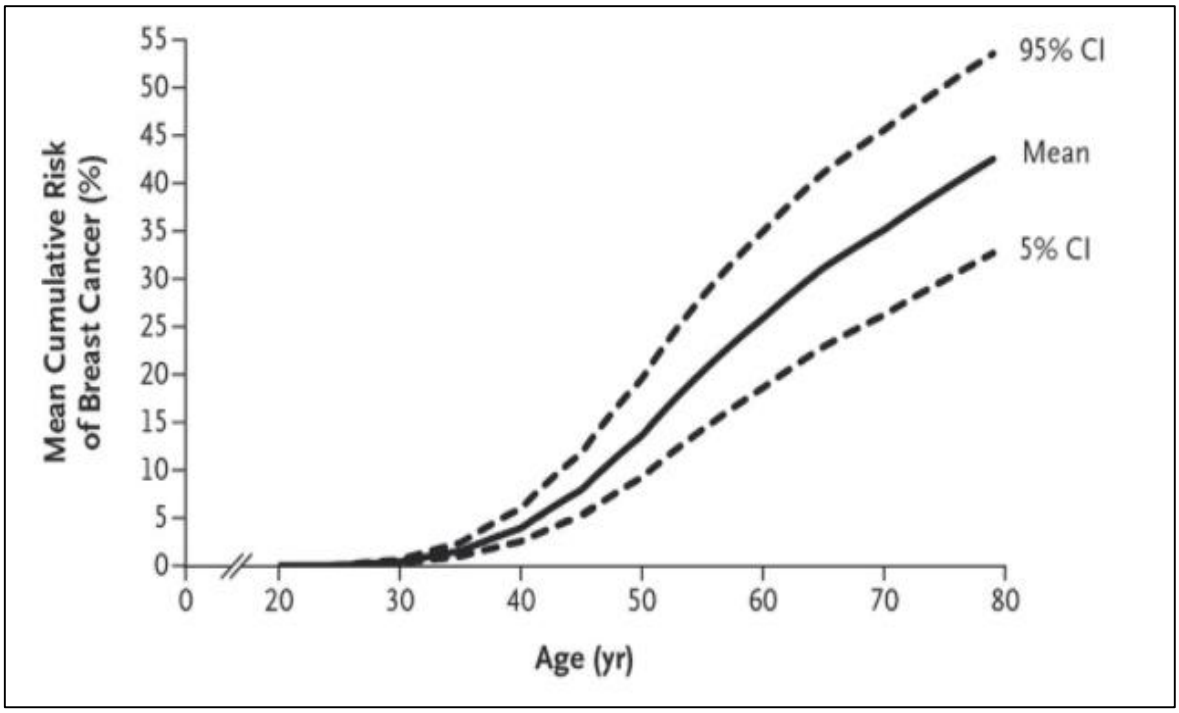


Figure 2.1 Graphical representation of the mean cumulative female risk of developing BC and associated confidence intervals. The graph demonstrates that the percentage risk of developing BC increases with age significantly from the age of 40 years (Antoniou *et al.,* 2014; viewed on 01/12/2017 at http://www.nejm.org/doi/full/10.1056/NEJMoa1400382#t=article).

Screening of *BRCA1* and *BRCA2* has aided in the treatment and management of the disease, however, these genes only account for ~ 20% of all familial BC cases, leaving the rest undefined (Gunnarsson *et al*., 2008; Stratton & Rahman, 2008). As a result, more possible familial BC causes were investigated and various high, medium and low risk genes were identified as possible contributors to the familial BC burden (Ziyaie *et al*., 2000; Xia *et al*., 2006; Thompson *et al.,* 2012; Majumbder & Chatterjee, 2017). This network of genes includes *PALB2,* BRCA1 Interacting Protein C-Terminal Helicase 1 *(BRIP1* MIM_605882*),* RAD51 paralog C *(RAD51C* MIM_602774) and X-Ray Repair Complementing Defective Repair In Chinese Hamster Cells 2 *(XRCC2* MIM_600375), just to mention a few (Xia *et al.,* 2006; Erkko *et al.,* 2007; Rahman *et al*., 2007; Mimitou & Symington, 2009; Meindl *et al.*, 2010; Suwaki *et al*., 2011; Scully & Xie, 2013).

The newly identified genes have a number of features in common, which include their critical role in cellular resistance to ionizing radiation (IR) and their involvement in DNA repair and in strand invasion during HR (Abbott *et al.,* 1998; Moynahan *et al,* 2001b; Takata *et al*., 2001; Zhang & Hannink, 2003; Klug *et al.,* 2006; Mimitou & Symington, 2009; Suwaki *et al.,* 2011; Park *et al.,* 2013; Scully & Xie, 2013; Ghosh *et al.,* 2014).

2.4.3 Partner and localizer of BRCA2

PALB2 is amongst the more recently identified BC genes (Xia *et al.,* 2006; Hamosh, 2016). This gene was discovered in 2006 by Xia *et al. (*2006). It has 13 exons, with exons 2 to exon 13 coding for the PALB2 protein. It is located on chromosome 16p12.3 (Xia *et al*., 2006; Erkko *et al*., 2007; Rahman *et al*., 2007). The gene was initially associated with N-subtype FA, then as a moderate BC penetrance gene, and has recently been classified as a high penetrance gene (Xia *et al.,* 2006; Burke *et al.,* 2017; Majumbder & Chatterjee, 2017). It encodes a ~ 130 kDa protein which interacts with BRCA1 and BRCA2. This protein is a major determinant for the nuclear localisation of BRCA2 to the site of repair (Foulkes *et al*., 2007; Frio *et al.,* 2010; Tischkowitz & Xia, 2010; Antoniou *et al*., 2014).

2.4.3.1 Structure of PALB2 protein

The PALB2 protein is divided into 3 distinct sections, with each section having a unique function (Figure 2.2 A). The first section ranges from amino acid (aa) 1 to aa 394, and consists of the coiled-coil binding domains known as the amino-terminal (Figure 2.2 B). The function of this section is to mediate the protein – protein interactions of three various proteins involved in DNA repair, namely BRCA1, Kelch-like ECH-Associated Protein 1 (KEAP1 MIM 606016) and RAD51. This interaction ensures proper localisation of BRCA2 and thus enables DNA repair (Lupas, 1996; Zhang *et al.,* 2009a; Zhang *et al.,* 2009b; Park *et al.*, 2014).

The central region comprises of aa 395 – 790 (Figure 2.2 C) and contains two unique areas that modulate the interaction of PALB2 with other proteins (Park *et al*., 2013). The first region, the chromatin-association motif (ChAM), ranges from aa 395 – 446 (Figure 2.2 C). It promotes the association of PALB2 and BRCA2 with chromatin, which is critical for cellular response to DNA damage (Liu *et al.,* 2002; Xia *et al.,* 2007; Dray *et al.,* 2010). The second region (aa 611 – 764, Figure 2.2 C) is responsible for the binding of PALB2 to the Mortality-factor-related gene 15 protein (MRG15 MIM 607303) (Lupas, 1996; Sy *et al.,* 2009a; Xie *et al.,* 2012). Its interaction with MRG15 not only enables the protein to recognize and pair with the damaged DNA, but also provides a potential interface between chromatin regulation and DNA damage repair by HR (Figure 2.3) (Gospodinov & Herceg, 2013; Smeenk & van Attikum, 2013).

Finally, the carboxy-terminus of the protein ranges from aa 791 – 1186 (Figure 2.2 D). It is dominated by beta-transducin repeats (WD − 40), which plays an important role in the prevention of tumour formation (Park *et al*., 2013). These WD-40 domains mediate protein interactions of numerous key proteins involved in HR, including BRCA2 and RAD51 (Stirnimann *et al.,* 2010; Park *et al.,* 2013). Additionally, binding of these PALB2 WD-40 domains to the DNA polymerase (pol η) mediates recombination-associated DNA synthesis (Buisson *et al.,* 2014).

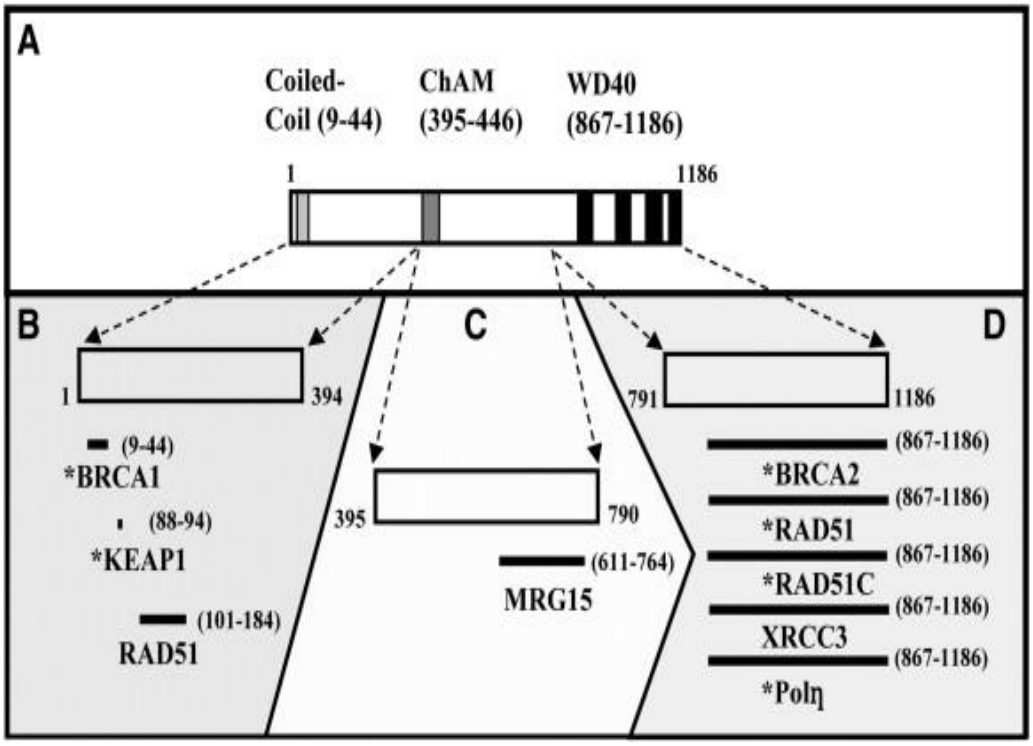


Figure 2.2 Structure of the PALB2 protein. **A.** The three regions of PALB2 protein. **B.** The N-terminal portion of PALB2 (aa 1 – 394), which interacts with BRCA1, KEAP1 and RAD51. **C.** The central region of PALB2 (aa 395 – 790) which interacts with MRG51. **D.** The C-terminus of PALB2 which interacts with BRCA2 (Park *et al.,* 2014; viewed on 01/12/2017 at http://www.sciencedirect.com/science/article/pii/S0304419X14000523).

2.4.3.2 Function of PALB2

The PALB2 protein activates DNA repair after double stranded breaks (DSB) (Fig. 2.3 A). It interacts with BRCA1, BRCA2 and RAD51 and this complex becomes the key regulator of the extension step after strand invasion at replication-dependent DSBs (Xia *et al.,* 2006; Sy *et al.,* 2009a; Zhang *et al.,* 2009a; Zhang *et al*., 2009b; Buisson *et al*., 2010; Buisson *et al.,* 2014; Leyton *et al.,* 2015; Vietri *et al.,* 2015).

Additionally, PALB2 interacts with MRG15 (Figure 2.3 B), a component of various histone modifying acetyltransferase-deacetylase complexes, and together are responsible for chromatin regulation and DNA damage repair by HR (Sy *et al.,* 2009b; Vietri *et al.,* 2015). When PALB2 interacts with KEAP1, it regulates cellular redox homeostasis (Ma *et al*., 2012). It also regulates the rate of Nuclear Factors Erythroid 2-like 2 (NRF2 MIM 600492) accumulation, function and export from the nucleus following induction (Ma *et al*., 2012).

PALB2 is also essential for the localisation of BRCA2 as well as the association of BRCA2 with chromatin and nuclear structures (Xia *et al.,* 2006). This association is vital for the DNA damage response pathway (Xia *et al*., 2006; Foulkes *et al*., 2007; Zhang *et al.,* 2009a; Zhang *et al.,* 2009b; Frio *et al.,* 2010; Antoniou *et al*., 2014; Buisson *et al.,* 2014). The functions of PALB2 are summarised in Figure 2.3 A – E by Pauty *et al.* (2014).

2.4.3.3 Mechanism of action

In the presence of DNA damage, BRCA1 is recruited to DSBs by sequential protein-protein interactions: ubiquitin-dependent signalling (Moynahan *et al.,* 2001a; Buisson & Masson, 2012). BRCA1 then recruits PALB2, and the two interact via aa 1393 – 1424 of BRCA1 and the coiled-coil domain (aa 9 – 44) of PALB2 (Figure 2.2 B). This interaction is crucial for accumulation and localisation of DNA polymerase at DSBs, and the extension after strand invasion at DSBs (Sy *et al.,* 2009b; Zhang *et al.,* 2009a; Zhang *et al.,* 2009b; Buisson & Masson, 2012). Subsequently, PALB2 recruits and binds to BRCA2 via PALB2 C-terminal WD − 40 domain (Xia *et al.,* 2007; Balia *et al.,* 2010).

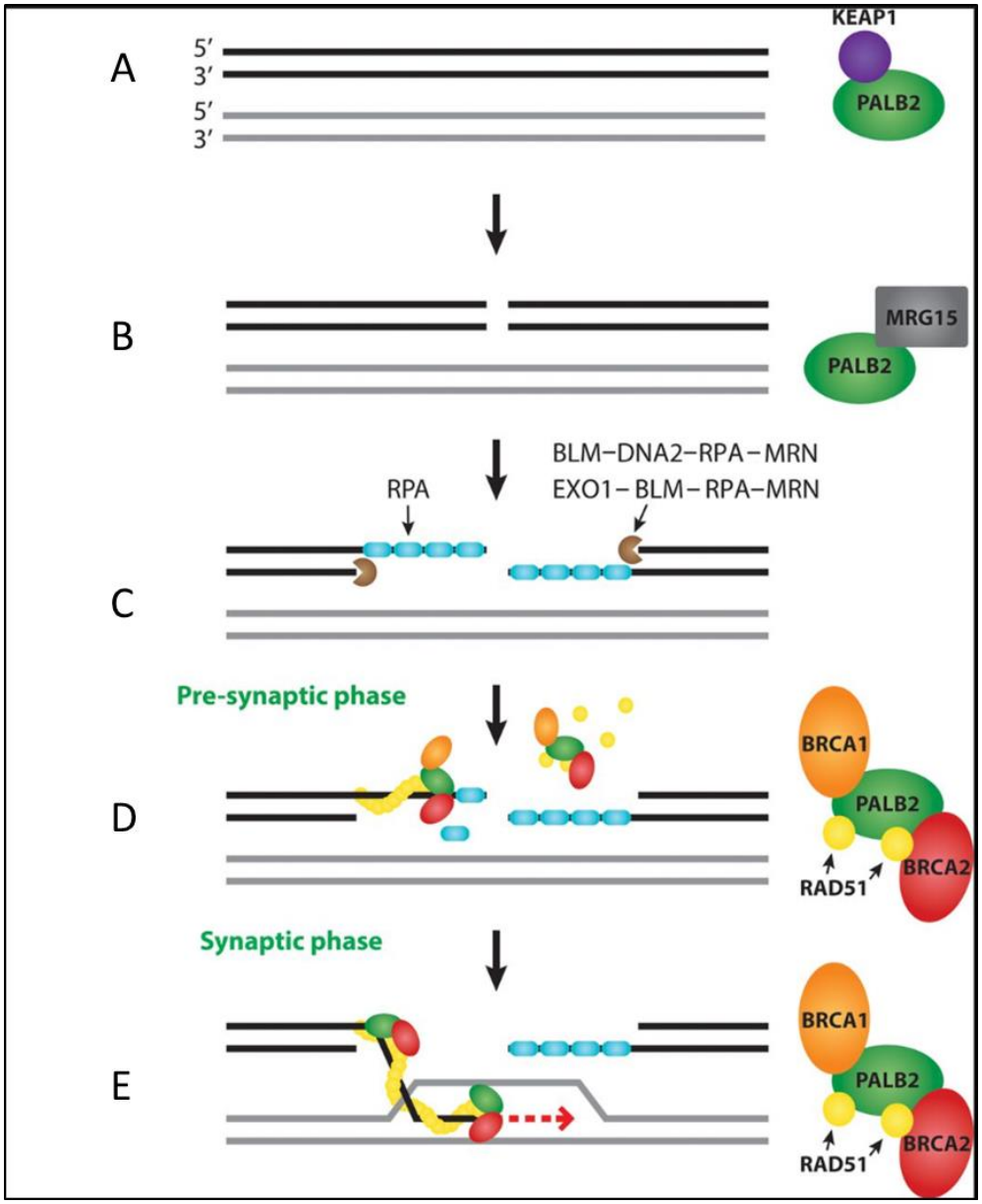


Figure 2.3 The roles of PALB2 in HR**.** **A** Protection of DNA from damage. **B** Double strand break and chromatin remodelling. **C** Resection from 5' to 3' and single – strand DNA protection**.** **D** RAD51 filament assembly. **E** Strand invasion, homologous sequence recognition and DNA synthesis. (Pauty *et al.,* 2014; viewed on 01/12/2017 at http://www.biochemj.org/content/460/3/331.figures-only).

Once BRCA2 is localized, RAD51 is recruited and assembled into a nucleoprotein filament with single-stranded DNA (ssDNA). RAD51 will bind to either the N-terminus or the WD-40 domain of PALB2 (Figure 2.2 D) (Buisson *et al*., 2010; Dray *et al.,* 2010). The role of RAD51 is to coat the processed ssDNA filament, promoting pairing during HR. The protein is also responsible for the initiation of replication at stalled replication forks (Xia *et al*., 2006; Zhang *et al*., 2009b; Buisson *et al.,* 2014; Leyton *et al.,* 2015).

A fifth protein, MRG15, then interacts with the central region of PALB2 (aa 611 – 764). This protein regulates transcription (Figure 2.2 B) and the acetylation of histone H4, which is essential for DSB signalling and repair (Carrozza *et al*., 2005; Rea *et al*., 2007; Xia *et al*., 2007; Wu *et al*., 2011; Vietri *et al.,* 2015). Overall the BRCA1 – PALB2 – BRCA2 – RAD51 – MRG15 complex performs the function of DNA damage response by HR repair in order to maintain genome integrity (Venkitaraman, 2002; Xia *et al*., 2006; Sy *et al.,* 2009a; Vietri *et al*., 2015). The DNA repair process is illustrated in Figure 2.3 by Pauty *et al*. (2014).

2.4.3.4 *PALB2* mutations and related diseases

Alterations in the structure and function of PALB2 that result from acquired mutations have been associated with a number of conditions. Initially, *PALB2* was known as *FANC-N* because of its association with FA (Reid *et al.,* 2007; Xia *et al.,* 2007; Mamrak *et al*., 2017). Affected individuals often presented with congenital abnormalities, bone marrow failure, and ca predisposition (Wang & Smogorzewska, 2015).

Other *PALB2* mutations have been associated with a 3 to 4% risk of developing familial pancreatic ca, increased risk for prostate and male BC and familial lipomatosis (Erkko *et al.,* 2007; Rahman *et al.,* 2007; Tischkowitz *et al.,* 2009; Slater *et al.,* 2010; Fernandes *et al.,* 2013; Antoniou *et al.,* 2014; Vietri *et al.,* 2015; Reddy *et al*., 2016).

Most importantly, *PALB2* mono-allelic mutations have predominantly been linked to early onset familial BC (Rahman *et al.,* 2007; Xia *et al.,* 2007; Erkko *et al*., 2008). These mutations have been associated with about 2 – 4 fold increase in risk of developing BC, a risk comparable to that of *BRCA2* mutations, hence *PALB2* was characterised as a high penetrance gene (Rahman *et al.,* 2007; Tischkowitz *et al.,*

2007; Erkko *et al*., 2008; Balia *et al.,* 2010; Foulkes *et al*., 2010; Tischkowitz & Xia, 2010; Haanpaa *et al.,* 2013; Majumbder & Chatterjee, 2017).

Although all people are at a risk of developing BC, it has been established that carriers of *PALB2* mutations are four (4) times more likely to develop BC than non-carriers (Rahman *et al.,* 2007; Antoniou *et al.*, 2014; Wojcik *et al*., 2016). Actually, the relative risk of developing BC amongst *PALB2* mutation carriers is higher at a younger age whereby it was estimated at 8 − 9 times as high amongst those younger than 40 years of age, 6 − 8 times amongst those 40 − 60 years of age, and 5 times as high amongst those older than 60 years of age (Erkko *et al.,* 2007; Rahman *et al.,* 2007; Antoniou *et al.,* 2014).

Currently, little is known about the clinical outcomes of *PALB2* mutations that may guide clinical management. But, at protein level, studies suggest that these *PALB2* mutations may result in the loss of BRCA2 and RAD51 localisation, in a reduced BRCA2 binding capacity, and in the deficiency of DNA synthesis (Xia *et al*., 2006; Rahman *et al.,* 2007; Tischkowitz *et al.,* 2007; Xia *et al.,* 2007; Erkko *et al.,* 2008; Pylkäs *et al*., 2008; Balia *et al.,* 2010; Buisson *et al*., 2010; Dray *et al*., 2010; Hayakawa *et al.,* 2010; Buisson *et al.,* 2014).

Tumours in *PALB2* mutation carriers have been found to be identical to those occurring in *BRCA2* cases. They often present with an estrogen-receptor-positive phenotype that is associated with aggressive disease and result in reduced survival (Foulkes *et al.,* 2007; Tischkowitz *et al.,* 2007; Heikkinen *et al.,* 2009; Teo *et al*., 2013; Antoniou *et al.,* 2014). The ten-year survival of *PALB2* mutation carriers was estimated to be 48%, falling to 32% in women with a tumour of 2.0 cm or larger as opposed to 74.7% in non-carriers (Heikkinen *et al.,* 2009; Cybulski *et al.,* 2015).

2.5 Previous findings

Multiple international studies have investigated the presence of *PALB2* mutations in association with familial BC across different populations. The majority of these studies reported mutations that spread across the entire coding region of the gene, with potential hotspots being exons 4 and 5, most probably due to the exon size (Figure 2.4) (Rahman *et al.,* 2007; Erkko *et al.,* 2008; Southey *et al.,* 2010; Casadei *et al.,* 2011; Antoniou *et al.,* 2014; Noskowicz *et al*., 2014). The mutations were detected in both male and female BC patients and were most prevalent in patients

who had a positive family history of BC or other ca types (Foulkes *et al*., 2007; Erkko *et al*., 2008; Papi *et al.,* 2009; Dansonka-Mieszkowska *et al.,* 2010). Some of the previously detected mutations are indicated in Figure 2.4, together with the types of ca they are associated with (Pauly, *et al.,* 2014).

The occurrence of *PALB2* mutations range from global single nucleotide variants (SNVs) to rare variants. An example of a global pathogenic mutation is *PALB2* c.509_510delGA, p.Arg170Ilefs (rs515726123). This mutation has been detected by multiple studies in Poland, Germany and Russia (Dansonka-Mieszkowska *et al.*, 2010; Casadei *et al*., 2011; Wojcik *et al*., 2016).

The incidence of the mutation varies; it is low in patients without a family history (as low as ~ 0.6%), but increases to ~ 1.7% for the patients who have one affected family member and to ~ 4.6% in patients who have a history of 3 or more affected family members (Dansonka-Mieszkowska *et al.*, 2010; Cybulski *et al.,* 2015; Wojcik *et al*., 2016). Although this mutation was detected in various countries, it must be noted that these countries are geographically in close proximity, therefore the gene pool may not be significantly different.

Some mutations appear to be more common in a particular population compared to the rest of the world (Table 2.1). For example, *PALB2* 1050_1051delAAinsTCT, p.Gln350Hisfs (rs515726060) is more common in the Chinese population, *PALB2* c.1633G>T, p.Glu545Ter (rs18077103) in the German, Polish and Russian populations and *PALB2* c.2323C>T, p. Gln775Ter (rs180177111) is common amongst the French-Canadians (Foulkes *et al*., 2007; Cao *et al.,* 2008; Bogdanova *et al*., 2011). Additionally, mutations such as *PALB2* c.1571delG, p.Leu451Terfs (rs5845353) and *PALB2* c.2257C>T, p.Arg753Ter (rs180177110) are more common in Italy compared to the rest of the world (Papi *et al.,* 2009; Balia *et al*., 2010; Vietri *et al.*, 2015). Examples of *PALB2* mutations and the populations for which they were detected are listed in Table 2.1.

Some of these mutations carry a higher risk than others. To demonstrate, mutation *PALB2* c.1592delT, p.Leu531Cysfs (rs180177102) has been associated with a 14.3 - fold increased risk of developing BC before 70 years of age, a risk that is similar to that of *BRCA2* mutations in Finland (Erkko *et al*., 2008; Heikkinen *et al.,* 2009; Haanpaa *et al*., 2013). The presence of this mutation results in the absence of the C-terminus region, thus inducing lack of interaction between PALB2 and

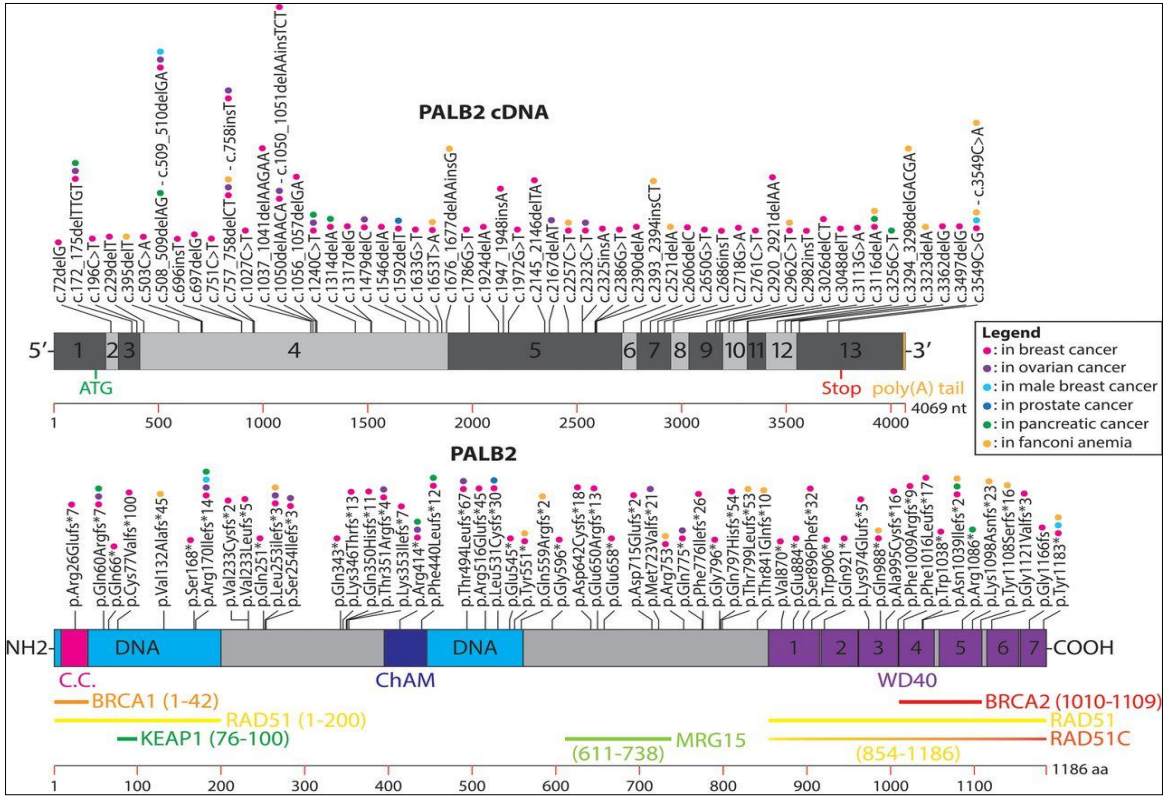


Figure 2.4 A schematic representation of known *PALB2* mutations distributed across the coding region of the gene, and the associated influences on the PALB2 protein. The mutations and aa changes are colour coded, with the colours corresponding to different ca types listed in the key. They are: male and female breast-, ovarian-, prostate-, pancreatic ca, and FA. The colours on the protein also highlight the different regions of the PALB2 protein, and indicates the binding sites of other proteins involved in dbDNA repair (Pauty *et al.* 2014; viewed on 01/12/2017 at http://www.biochemj.org/content/460/3/331.figures-only).

Table 2.1 *PALB2* mutations published in international literature.

Mutation	**Protein change**	**rs number**	**Population/ ethnicity**	**Citation**
PALB2 c.172_175delTTGT	p.Gln60Argfs	rs180177143	European, African American	Casadei *et al.* (2011); Hellebrand *et al.* (2011); Castera *et al.* (2014)
PALB2 c.196C>T	p.Gln66Ter	rs180177083	European, African American	Casadei *et al.* (2011); Wong *et al.* (2011)
PALB2 c.509_510delGA	p.Arg170Ilefs	rs515726123	European, African American, German, Polish, Russian	Dansoka-Mieszkowska *et al.* (2010); Bogdanova *et al.* (2011); Casadei *et al.* (2011); Fernandes *et al.* (2013)
PALB2 c.697delG	p.Val233Leufs	rs180177090	British, SA	Sluiter *et al.* (2009)
PALB2 c.757_758delCT	p.Leu253Ilefs	rs180177092	European, African American	Casadei *et al.* (2011)
PALB2 c.758dupT	p.Ser254Ilefs	rs515726126	African American, Australian	Zheng *et al.* (2012); Wong-Brown *et al.* (2013); Churpek *et al.* (2015)
PALB2 c.1027C>T	p.Gln343Trp	rs63750603	Italian	Catucci *et al.* (2014)
PALB2 c.1037_1041delAAGAA	p.Lys346Thrfs	rs587776410	Malaysian and Singaporean	Pauty *et al.* (2013)
PALB2 c.1050_1051delAAinsTCT	p.Gln350Hisfs	rs180177098	Chinese, Malaysian, Singaporean	Cao *et al.* (2008); Pauty *et al.* (2013)
PALB2 c.1050_1053delAACA	p.Thr351Argfs	rs515726060	Malaysian and Singaporean	Pauty *et al.* (2013)
PALB2 c.1056_1057delGA	p.Lys353Ilefs	rs180177099	Spanish	Garcia *et al.* (2009)
PALB2 c.1240C>T	p.Arg414Cys	rs137854567	European, American, German, Russian, African	Bogdanova *et al.* (2011); Casadei *et al.* (2011); Hellebrand *et al.* (2011)

Table 2.1 Continues

Mutation	**Protein change**	**rs number**	**Population/ ethnicity**	**Citation**
PALB2 c.1317delG	p.Phe440Leufs	rs515726067	Italian	Balia *et al. (*2010)
PALB2 c.1479delC	p.Thr494Leufs	rs515726071	African American	Churpek *et al. (*2015)
PALB2 c.1546delA	p.Arg516Glufs	rs584481560	American	Fernandes *et al. (*2014)
PALB2 c.1571delG	p.Leu451Terfs	rs5845353	Italian	Balia *et al. (*2010)
PALB2 c.1592delT	p.Leu531Cysfs	rs180177102	Finnish	Erkko *et al. (*2007); Tischkowitz *et al. (*2007); Haanpaa *et al. (*2013)
PALB2 c.1633G>T	p.Glu545Ter	rs180177103	American, German, Russians	Bogdanova *et al. (*2011); Fernandes *et al. (*2014)
PALB2 c.1653T>A	p.Tyr551Ter	rs118203997	European, African American	Casadei *et al. (*2011)
PALB2 c.1692A>G	p.Lys564Lys	rs786201842	South American	Leyton *et al. (*2015)
PALB2 c.1919C>A	p.Ser640Ter	rs760094988	French, South American	Damiola *et al. (*2015); Leyton *et al. (*2015)
PALB2 c.1924delA	p.Met642Cysfs	rs730881865	American	Fernandes *et al. (*2014)
PALB2 c.1947_1948dupA	p.Glu650Argfs	rs515726075	Australian	Teo *et al. (*2013)
PALB2 c.1972G>T	p.Glu658Ter	rs762161397	European	Castera *et al. (*2014)
PALB2 c.2145_2146delTA	p.Asp715Glufs	rs515726081	German	Hellebrand *et al. (*2011)

Table 2.1 Continues

Mutation	**Protein change**	**rs number**	**Population/ ethnicity**	**Citation**
PALB2 c.2257C>T	p.Arg753Ter	rs180177110	Italian	Papi *et al. (*2010); Hellebrand *et al. (*2011)
PALB2 c.229delT	p.Cys77Valfs	rs180177084	Jewish, French Canadians and mixed ancestry	Tischkowitz *et al. (*2007)
PALB2 c.2323C>T	p.Gln775Ter	rs180177111	French Canadian	Foulkes *et al. (*2007)
PALB2 c.2325dupA	p.Phe776Ilefs	rs876659997	European	Castera *et al. (*2014)
PALB2 c.2386G>T	p.Gly796Ter	rs180177112	European, African American	Casadei *et al. (*2011); Castera *et al. (*2014); Fernandes *et al. (*2014)
PALB2 c.2390delA	p.Gln797Hisfs	rs515726086	Australian	Wong-Brown *et al. (*2013)
PALB2 c.2559C>T	p.Gly853=	rs180177115	European, African American	Casadei *et al. (*2011)
PALB2 c.2686dupT	p.Ser896Phefs	rs515726091	European, African American	Casadei *et al. (*2011)
PALB2 c.2718G>A	p.Trp906Ter	rs180177122	European, African American	Casadei *et al. (*2011)
PALB2 c.2962C>T	p.Gln988Ter	rs118203999	German	Hellebrand *et al. (*2011)
PALB2 c.2982insT	p.Ala995Cysfs	rs180177127	European	Rahman *et al. (*2007)

Table 2.1 Continues

Mutation	**Protein change**	**rs number**	**Population/ ethnicity**	**Citation**
PALB2 c.2993G>A	p.Gly998Glu	rs45551636	South American	Leyton *et al. (*2015)
PALB2 c.3048delT	p.Phe1016Leu fs	rs515726104	African American	Zheng *et al. (*2012); Churpek *et al. (*2015)
PALB2 c.3116delA	p.Asn1039Ilefs	rs180177133	European	Rahman *et al. (*2007); Garcia *et al. (*2009)
PALB2 c.3362delG	p.Gly1121Vafs	rs515726117	European	Castera *et al. (*2014)
PALB2 c.3549C>G	p.Tyr1183Ter	rs118203998	American	Garcia *et al. (*2009); Ding *et al. (*2011); Tischkowitz *et al. (*2012); Fernandes *et al. (*2014)

MRG15, RAD51 and BRCA2 (Vietri *et al.,* 2015).

On the contrary, there are a number of studies that reported no evidence of an association between BC and *PALB2* mutations regardless of a positive family history of the disease. In the Icelandic population for one, researchers did not find correlation between BC cases and *PALB2* variants detected in the high risk *BRCA1* and *BRCA2* mutation negative patients who took part in the study (Gunnarsson *et al.*, 2008). Correspondingly, the large genomic rearrangements in *PALB2* in the Northern Finish population were not associated with BC in high risk *BRCA1* and *BRCA2* negative patients (Pylkäs *et al.,* 2008). Also, research on 25 Australian male patients with invasive BC did not detect any disease-causing mutations in the *PALB2* gene (De Chalon *et al*., 2010).

Furthermore, similar results were attained in studies on BC patients in Korea, Ireland and Chile respectively (Kim *et al*., 2010; McInerney *et al*., 2010; Leyton *et al*., 2015). These studies dispute the hypothesis that *PALB2* mutations are associated with BC in all populations. In fact, it is known that factors such as epigenetic variation contributes to inter-individual variation in gene expression, and may contribute to the observed variation in ca susceptibility (Cui *et al.,* 2003; Calmon *et al.,* 2015).

2.6 South African studies

In SA, BC is the most common ca among women (Sluiter *et al*., 2009, CANSA, 2012). SA women generally tend to develop BC at a younger age, but the diagnosis only occurs at a later stage due to the lack of awareness, access to diagnostic centres available and limited screening services (Walker *et al*., 2004). The overall life-time risk of developing BC in SA has been increasing over time. In 2004, 1:13 White and 1:81 Black SA females were at risk of developing BC. Currently, the risk is 1:15 for White and 1:53 for Black women (Reeves *et al*., 2004; Francies *et al.,* 2015). To date, the diagnosis had been centred on *BRCA* genes as *PALB2* has not been well studied.

Sluiter *et al.* (2009) investigated the presence of *PALB2* variants in 48 white SA BC patients attending the Oncology clinic in Pretoria. The cohort was early-onset BC cases where 35% of patients reported family history and 65% did not have a family history of BC. The study detected a novel truncating mutation, *PALB2* c.697delG, p.Val233Leufs (rs180177090) in one patient of British descent who was diagnosed with infiltrating ductal carcinoma at the age of 44. This patient did not have a family history of BC but of pancreatic, colon and prostate ca (Sluiter *et al.,* 2009). The drawbacks of the study was that it investigated the variants in White patients only, an ethnic group that forms only 8.7% of SA population (Francies *et al.,* 2015), as a result, it was not a complete reflection of SA patients.

Another study was conducted by Francies *et al.* (2015) on a total of 108 BC patients attending Johannesburg hospitals. The study screened *BRCA1, BRCA2* and *PALB2* mutations and *CHEK2* c.1100delC in different SA ethnic groups diagnosed with pre-menopausal and/or triple negative BC. The selection criteria was inclusive of triple-negative BC tumours and early diagnosis. The study did not take into consideration the family history of patients. It included 78.7% Black, 14.8% White, 4.6% Indian and 1.9% Coloured patients. None of the patients had deleterious *PALB2* mutations. Although the study had a fair representation of SA ethnic groups, it did not focus on high risk patients and specifically for *PALB2*, it did not consider *BRCA* status (Francies *et al*., 2015).

2.7 Experimental methods

The study will make use of Real Time-PCR based High Resolution Melting Analysis (qPCR based HRMA) and Automated DNA Sanger sequencing.

2.7.1 qPCR based HRMA

qPCR based HRMA is the combination of polymerase chain reaction (PCR) and melting analysis performed in a qPCR instrument. It involves the precise monitoring of the change in fluorescence caused by the release of an intercalating DNA dye from a DNA duplex as it melts (Krypuy *et al*., 2006). The fluorescent double stranded DNA (dsDNA) intercalating dyes are normally added to the reaction mixture prior to

amplification, and are incorporated in dsDNA amplicons during the PCR annealing step (Ruskovaa & Raclavskya 2011).

At the end of a PCR reaction, the amplicons are normally in a double stranded form and have associated with molecules of intercalating dye. The DNA molecules are then melted by slowly increasing the temperature to as high as 95ºC. During this step, the DNA strands dissociate at a rate and pattern that correlates with the length and nucleotide sequence of the DNA molecules (Ruskovaa & Raclavskya 2011). This is observed as a sudden decrease of fluorescence near melting temperature (Tm), and can be plotted as a melting curve of unique shape. Based on the nucleotide sequences, wild-type DNA sequence can easily be discriminated from different mutations present (Krypuy *et al.,* 2006; Ruskovaa & Raclavskya 2011).

There are many advantages of HRMA. It is a simple, relatively easy to use assay which can be performed by individuals with no specific skills or those who had no intensive training. It is cost effective, fast and has high sensitivity/ specificity (Vossen *et al*., 2009). It is flexible and has a non-destructive nature so the products can be used in other methods such as gel electrophoresis or Sanger sequencing. HRMA also requires very small amounts of DNA for analysis, and this allows for better temperature control and fast melting rate (Ruskovaa & Raclavskya 2011). It is an in-tube method in which the analysis is performed immediately after the amplification and it thus prevents crossover of PCR products (Krypuy *et al.,* 2006).

Although HRMA is an attractive method, it has its own drawbacks. HRMA cannot easily be applied in multiplex mode, meaning that it cannot be used to type variants in several different fragments at the same time. Also, the sensitivity decreases with an increase in the size of an amplicon (Vossen *et al*., 2009). In addition, HRMA falls short in the detection of homozygous variants, for example, the subtle differences for some variants such as A-T to T-A changes, can easily be missed (Vossen *et al*., 2009; Ruskovaa & Raclavskya 2011). However, it is believed that the advantages of HRMA greatly outweighs the drawbacks, hence it is still the most widely used method of identification and detection in both research and diagnostics.

2.7.2 DNA Sanger Sequencing

Sanger sequencing is a technique used to determine the sequence of nucleotide bases on a DNA molecule (Ewing & Green, 1998). It provides information on the size of the sequence, the organization of genes in a sequence, the size of the genes and where the mutation is found (Kircher & Kelso, 2010). The reaction mixture comprises of DNA polymerase, a primer which flanks the DNA of interest, the deoxynucleotides (dNTPs), dideoxynucleotides (ddNTPs) and ssDNA which serves as a template (Kircher & Kelso, 2010).

Unlike the dNTPs that have a hydroxyl group, the ddNTPs have a hydrogen ion group at the 3' terminal. Because of this hydrogen ion, when a ddNTP binds to a growing stand during the elongation step of DNA polymerase, non-reversible termination occurs. As this random choice is made each time a nucleotide is added to the template, and there are thousands of template strands in the reaction, new strands of every possible length accumulate in the reaction mixture. The resulting molecules are sorted by their molecular weight and the label attached to the terminating ddNTP is read out sequentially in the order created by the sorting step, which is carried out by capillary electrophoresis (Kircher & Kelso, 2010). The laser then scans the fragments to identify the end nucleotides by the fluorescent colour. To analyze the results, a computer detects and records the colour of each ddNTP at the end of each fragment and thus forms a sequence (Kircher & Kelso, 2010).

Automated Sanger sequencing has a wide range of uses. It is applicable in earlier detection of genetic predisposition to diseases, which enables the assessment of risks associated with the development of certain diseases such as ca (Ewing & Green, 1998; Kircher & Kelso, 2010). It is also used in rational drug design whereby drugs are designed to target specific gene products that cause diseases (Ewing & Green, 1998; Kircher & Kelso, 2010).

There are advantages to using Sanger sequencing technology. Firstly, the sequencing machine is highly efficient; up to 384 sequences of between 600 and 1000 nucleotides in length can be sequenced simultaneously. The procedure is fast and has a low error rate (Ewing & Green, 1998; Kircher & Kelso, 2010). However, Sanger sequencing is a relatively expensive method and has proven impractical for larger sequencing projects. It has also been associated with errors which result from a number of reasons such as, errors in the amplification step, natural variance

as well as contamination in the sample used. Furthermore, lower intensities and missing termination variants tend to lead to sequencing errors accumulating toward the ends of long sequences (Ewing & Green 1998; Kircher & Kelso 2010).

Although the disadvantages of the above mentioned techniques are highlighted, and can compromise the integrity of the results of the study, these methods are still effective and applied in many laboratories. In fact, the majority of studies on *PALB2* used these methods to detect mutations and polymorphisms in their studies. Other reasons for our choice was financial and mechanical constraint as well as lack of skills in methods such as Next Generation Sequencing.

2.8 Summary

BC is the most commonly diagnosed female ca in SA (Thompson & Easton, 2004; CANSA, 2012). Familial BC forms ~ 20% of all BC cases and have been defined by *BRCA1/2*, the high risk genes whose mutations are associated with BC (Foulkes *et al*., 2007). Although screening these two genes forms the basis of familial BC diagnosis, their small percentage contribution left room to further explore other possible contributors.

Many studies have established that PALB2 is an important protein in the repair of damaged DNA, and that mutations in *PALB2* are associated with various conditions including BC. However, variants in this gene have conflicting interpretations depending on the populations in which they are detected. For example, while some variants are described as pathogenic in some populations, the same variants are classified as benign in others (Xia *et al.,* 2006; Foulkes *et al*., 2007; Zhang *et al.,* 2009b; Frio *et al.,* 2010; Antoniou *et al*., 2014).

The incidence of *PALB2* mutations also varies from one population to another. For example, in Finland, *PALB2* mutations are as common as *BRCA2* mutations amongst BC patients, hence *PALB2* screening has been incorporated into the Finnish BC screen. This has not only aided in the treatment and management of the disease, but is also used for pre-symptomatic testing and making informed decisions regarding reproduction and lifestyle. In some countries such as SA however, *PALB2* mutations have not been studied extensively, which leaves a lot of unresolved cases being undiagnosed and receiving unspecific treatment.

2.9 Aim of the study

The primary aim of this study was to screen for the presence of pathogenic *PALB2* mutations in the genetically diverse SA *BRCA1/2* negative BC and OVC patients, using HRMA and DNA Sanger sequencing as screening methods. Additionally, to assess the gene's potential role in BC predisposition in patients with a positive family history.

2.10 Objectives

a) Optimize qPCR-based HRMA assays for each of the 12 coding exons and splice site boundaries of *PALB2* (30 different primer sets) in order to screen for potential disease-causing mutations.

b) Comprehensively screen a total of 86 *BRCA1/2* negative breast- or OVC patients who had an extensive family history of breast- and other ca types representing each of the various SA ethnic groups.

c) Perform automated bidirectional Sanger sequencing analysis for amplicons exhibiting variation based on HRMA results (<-2.5 – >2.5 deviation from the base line in the subtractive difference plot).

d) Determine the impact of all variants identified on the protein and disease phenotype.

e) Correlate the mutations found (if any), with that reported internationally.

Chapter 3

Methodology

3.1 Introduction

In the previous chapter, the structure of *PALB2*, the function of the protein and localisation of the various functional domains were discussed in detail. The chapter also highlighted the latest findings regarding the presence of various mutations in distinct populations and their associated effect as far as BC is concerned. This chapter therefore outlines the study methodology that was utilized to screen for mutations in the previously unstudied SA population. The chapter describes the systematic and theoretical framework applied in the study which allowed for the discovery, collection and analysis of data in order to acquire knowledge in an ethical manner (Vogt *et al.,* 2017).

3.2 Study design

The study made use of an experimental quantitative design, which involved selection of a group of individuals upon which a variable was tested without any random pre-selection processes (Cook, 2015). In this case, archived DNA samples of SA female BC cases were screened for the presence of disease-causing mutations in *PALB2* to identify the variables (possible deleterious mutations) in the SA population. For interpretation of data, the study integrated individual case studies and experimental results generated to reinforce the findings. It included the patient's family history, their negative *BRCA1/2* mutation status as well as the *PALB2* mutation screening results gathered from various laboratory experiments. This approach was advantageous in that it saved time and exploited natural phenomena that allowed for a deep understanding of possible variables without inflicting the participants any harm (Cook, 2015).

3.3 Sampling procedure

3.3.1 Samples used for optimisation of conventional and qPCR

DNA samples were obtained from three postgraduate students who initially gave blood voluntarily after signing consent. These individuals neither had a positive family history of breast and/or other ca types, nor were they affected by any disease. These DNA samples were used for the initial optimisation of the conventional PCR as well as qPCR reactions of the various *PALB2* primer sets.

3.3.2 Research participants

From a pool of approximately 4 000 archived DNA samples that were banked at the Molecular Laboratory of the Division of Human Genetics, National Health Laboratory Services (NHLS) in Bloemfontein, a total of 86 unrelated BC/OVC patients were retrospectively selected based on specific selection criteria. For inclusion, patients should have been born in SA, be affected with either breast or ovarian ca, had an extensive family history of breast and other ca types (minimum of three affected individuals including the index) and have tested negative for the presence of pathogenic mutations in the familial BC genes *BRCA1* and *BRCA2*. An additional critical factor for inclusion was the availability of sufficient DNA of a good quality to perform a complete screen of *PALB2* using HRMA and Sanger sequencing. Provincial approval was obtained in advance for the use of their patients (Appendixes A and B).

For the purposes of this study, a first degree relative was defined as a close blood relative and included the patient's parents, full siblings or children. Second degree relatives were defined as the patient's grandparents, grandchildren, aunts, uncles, nephews, nieces or half-siblings. Information regarding family history was extracted from patient request forms, which in most cases were accompanied by a family pedigree drawn by the genetic counsellors/genetic nurses/clinicians from the respective hospital. The selection of patients represented each of the four SA ethnic groups (self-declared by index patient) and are presented in Tables 3.1 to 3.4.

Table 3.1 Indian patients included in the study. Indicated is the patient number, the index patient's unique laboratory DNA number, dx, ethnicity as stipulated on the request form, the extent of the family history and the various ca types present in the family.

Patient number	DNA number	Age at diagnosis	# of affected family members	Ca types present in the family
1	CAM2335	35	4	3 Breast, 1 throat
2	CAM2336	46	3	3 Breast
3	CAM2338	35	4	2 Breast, 2 ovarian
4	CAM2339	38	3	2 Breast, 1 stomach
5	CAM2481	37	3	2 Breast, 1 lung
6	CAM2513	32	4	2 Breast, 1 colon, 1 childhood ca
7	CAM2551	44	4	2 Breast, 1 ovarian, 1 stomach
8	CAM2645	58	4	3 Breast, 1 uterus
9	CAM2646	21	3	3 Breast
10	CAM2708	48	6	5 Breast, 1 lung
11	CAM2758	50	4	3 Breast, 1 uterus
12	CAM2780	66	4	3 Breast, 1 throat
13	CAM2806	57	6	3 Breast, 1 liver, 1 lung, 1 stomach
14	CAM2807	51	4	2 Breast, 1 uterus, 1 prostate
15	CAM2820	67	5	4 Breast, 1 mandible
16	CAM2821	48	5	3 Breast, 1 stomach, 1 leukaemia
17	CAM2826	67	4	3 Breast, 1 mandible
18	CAM2861	28	6	2 Breast, 1 prostate, 1 stomach, 1 throat, 1 endometrial
19	CAM2897	83	3	3 Breast
20	CAM2911	42	8	4 Breast, 1 colon, 2 lung, 1 uterus
21	CAM2960	68	5	4 Breast, 1 tongue
22	CAM2985	48	5	5 Breast

Table 3.2 Black patients included in the study. Indicated is the patient number, the index patient's unique laboratory DNA number, dx, ethnicity as stipulated on the request form, the extent of the family history and the various ca types present in the family.

Patient number	DNA number	Age at diagnosis	# of affected family members	Ca types present in the family
1	CAM2725	30	3	3 Breast
2	CAM2753	58	4	2 Breast, 1 ovarian, 1 throat
3	CAM2754	47	3	2 Breast, 1 ovarian
4	CAM2804	48	2	Breast
5	CAM2815	27	3	2 Breast, 1 trachea
6	CAM2816	25	5	5 Breast
7	CAM2847	46	3	2 Breast, 1 ovarian
8	CAM2869	32	3	3 Breast
9	CAM2912	38	8	4 Breast, 3 cervical, 1 prostate
10	CAM2915	61	4	3 Breast, 1 prostate
11	CAM2950	35	3	1 Breast, 1 stomach, 1 brain
12	CAM2953	33	3	3 Breast
13	CAM2991	35	5	1 Breast, 1 brain, 1 cervical, 2 leukaemia
14	CAM3244	29	3	3 Breast
15	CAM3256	41	3	2 Breast, 1 cervical

Table 3.3 Coloured patients included in the study. Indicated is the patient number, the index patient's unique laboratory DNA number, dx, ethnicity as stipulated on the request form, the extent of the family history and the various ca types present in the family.

Patient number	DNA number	Age at diagnosis	# of affected family members	Ca types present in the family
1	CAM482	32	5	1 Breast, 1 brain, 1 stomach, 1 lung, 1 colon
2	CAM1715	52	10	5 Breast, 1 skin, 1 colon, 1 prostate, 1 lung, 1 leukaemia
3	CAM1817	48	4	4 Breast
4	CAM1818	42	4	4 Breast
5	CAM1860	58	3	2 Breast, 1 prostate
6	CAM1890	38	3	3 Breast
7	CAM1892	50	3	2 Breast, 1 ovarian
8	CAM1998	60	3	2 Breast, 1 throat
9	CAM2233	44	7	3 Breast, 1 ovarian, 1 larynx, 1 colon, 1 intestine
10	CAM2319	29	3	3 Breast
11	CAM2321	56	4	3 Breast, 1 prostate
12	CAM2378	69	4	2 Breast, 2 ovarian
13	CAM2411	38	5	2 Breast, 1 liver, 1 stomach
14	CAM2460	55	8	5 Breast, 1 prostate, 1 lung, 1 stomach
15	CAM2578	42	3	3 Breast
16	CAM2703	62	4	3 Breast, 1 thyroid
17	CAM2755	45	5	3 Breast, 1 prostate, 1 colon
18	CAM2761	39	5	1 Breast, 1 cervix, 1 throat, 2 lung
19	CAM2767	53	4	4 Breast
20	CAM2782	50	22	9 Breast, 1 colon, 4 brain, 3 ovarian, 5 prostate
21	CAM2811	57	3	2 Breast, 1 ovarian
22	CAM2848	37	7	3 Breast, 2 colon , 2 uterine
23	CAM2906	26	7	1 Breast,1 colon, 1, stomach, 3 prostate, 1 sarcoma
24	CAM2921	40	6	4 Breast, 2 colon
25	CAM2983	31	3	3 Breast
26	CAM2987	45	4	2 Breast, 1 stomach, 1 stomach
27	CAM3061	38	5	5 Breast
28	CAM3077	51	3	2 Breast, 1 lymphoma
29	CAM3158	44	3	1 Breast, 1 ovarian, 1 prostate

Table 3.4 Afrikaner patients included in the study. Indicated is the patient number, the index patient's unique laboratory DNA number, dx, ethnicity as stipulated on the request form, the extent of the family history and the various ca types present in the family.

Patient number	DNA number	Age at diagnosis	# of affected family members	Ca types present in the family
1	CAM2496	27	5	Not indicated
2	CAM2574	43	4	3 Breast, 1 ovarian
3	CAM2697	65	9	5 Breast, 1 ovarian, 2 prostate, 1 lung
4	CAM2735	59	3	3 Breast
5	CAM2933	43	5	3 Breast, 1 lung, 1 stomach
6	CAM2998	50	8	7 Breast, 1 prostate
7	CAM3043	39	4	4 Breast
8	CAM3050	38	3	3 Breast
9	CAM3055	66	10	9 Breast, 1 prostate
10	CAM3127	36	6	2 Breast, 2 prostate, 1 throat, 1 stomach

Ethnicity in this study was referred to as proposed by the South African Society of Human Genetics (SASHG) released in 2013 (Appendix C). White patients who did not declare an Afrikaner heritage were included in Table 3.5.

3.4 Ethical considerations

The proposed project was presented and approved by the Postgraduate Evaluation Committee of the Faculty of Health Sciences (FoHS), School of Medicine at the University of the Free State (UFS) in Bloemfontein (Appendix D). The approved research proposal was submitted to the Ethics Committee of the Faculty of Health Sciences of the UFS for ethical clearance (ECUFS NR 31/95C, Appendix E). This study formed part of a comprehensive project on the familial BC syndrome that commenced in 1995 (titled “The application of genetic markers for the identification of familial breast and ovarian cancer” - ETOVS 31/95).

The study adhered to appropriate ethical conduct, which were based on the four fundamental ethical principles, namely autonomy (self-rule), non-maleficence (obligation not to harm others), beneficence (take positive steps to prevent harm), and justice (obligation to provide the patients with what they deserve such as confidentiality). All the patients included in the study, were initially counselled prior to familial BC testing by a genetic counsellor/nurse in collaboration with a clinician. The counselling session was based on the propagated protocol of the SA Genetic Counselling Association (http://www.geneticcounselling.co.za). This protocol was constructed for families requesting information regarding familial/hereditary BC and the involvement of genes such as *BRCA1* and *BRCA2*. It aimed to guide the counselling sessions according to a standardized format. Once written informed consent was obtained, blood was drawn and sent to Bloemfontein for *BRCA* analysis. All the patients were given unique internal reference numbers (CAM number) to ensure confidentiality of information (justice).

After the initial genetic testing for mutations in *BRCA1* and *BRCA2*, the extended research study was discussed during a follow-up counselling session in Durban with participants that tested negative for *BRCA* disease-causing mutations (beneficence).

Table 3.5 White (non-Afrikaner) patients included in the study. Indicated is the patient number, the index patient's unique laboratory DNA number, dx, ethnicity as stipulated on the request form, the extent of the family history and the various ca types present in the family.

Patient number	DNA number	Age at diagnosis	# of affected family members	Ca types present in the family
1	CAM702	24	4	3 Breast, 1 colon
2	CAM2580	30	3	3 Breast
3	CAM2722	55	3	3 ovarian
4	CAM2730	56	4	3 Breast, 1 pancreas
5	CAM2742	69	4	2 Breast, 1 brain, 1 liver
6	CAM2854	49	4	3 Breast, 1 ovarian
7	CAM2919	46	3	3 Breast
8	CAM2945	36	13	5 Breast, 1 colon, 2 prostate, 2 lung, 2 stomach, 1 throat
9	CAM3010	60	3	3 Breast
10	CAM3140	66	4	3 Breast, 1 ovarian

Once the patient understood the information and was given ample time to ask questions, additional informed consent was obtained (autonomy) (Appendix F). Only patients with adequate amounts of high quality DNA available were included into the study. No additional blood samples were requested or collected to ensure non-maleficence (avoid danger or harm to an individual). If informative results were obtained during this study, the information was to be portrayed back to the participant via the genetic counsellor.

Permission for the use of the facilities and the apparatus to conduct the research was obtained from NHLS Business Manager of the Universitas Academic Laboratories in Bloemfontein (Appendix G).

3.5 Methods

An overview of the methodology involved in this project is presented in Figure 3.1. The diagram indicates the steps taken in order to successfully screen for mutations in *PALB2* using qPCR based HRMA and automated Sanger sequencing.

3.5.1 DNA extraction

DNA samples used in this study were all extracted previously from ethylenediaminetetraacetic acid (EDTA) blood samples using the salting out method. This method was first described by Miller *et al.* (1988) and adapted for use in the NHLS diagnostic laboratory by Combrink (2016). The method is currently performed according to a standardized protocol as described in the Standard Operating Procedure of this NHLS laboratory, based on the adaptions described by Combrink (2016).

The method briefly entailed the collection of 10 – 20 ml blood in EDTA tubes. Upon arrival, all the blood in the EDTA tube was transferred to labelled Nunc tubes and stored at - 20°C. On the day of extraction, the frozen samples were thawed and cells ruptured using 45 ml cold lysis buffer [0.3 M sucrose, 10 mM 2-amino-2-(hydroxymethyl)-1,3-propanediol (Tris) pH 7.8, 5 mM $MgCl_2$, 1% (v/v) t-octylphenoxypolyethoxyethanol (Trixton X-100)]. The solution was centrifuged for 20 min at 4000 rpm at 4°C, where after the pellet was resuspended in 1x SET buffer (10 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA) containing 10 $\mu g.\mu l^{-1}$

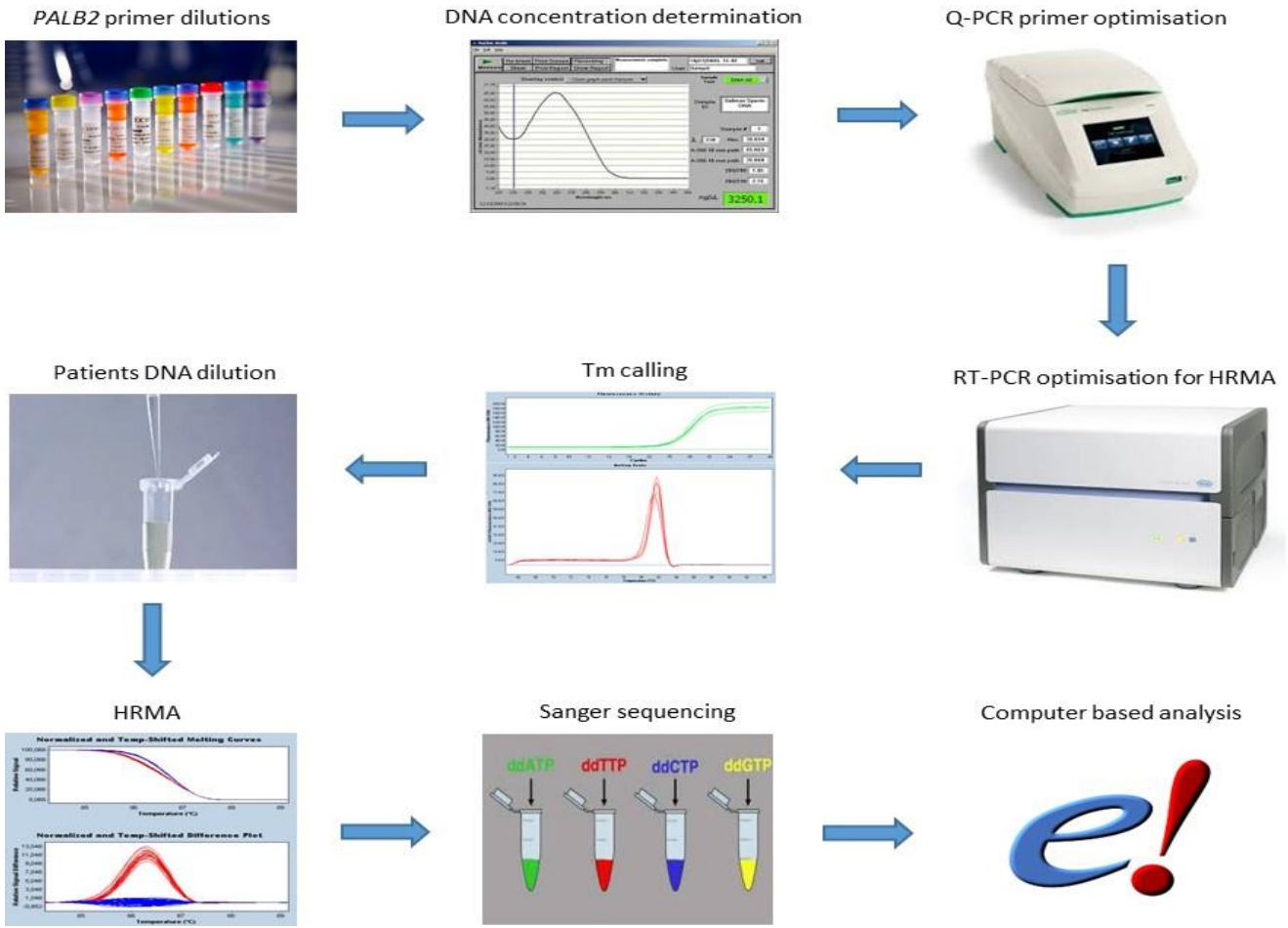


Figure 3.1 Overview of the methodology involved in data collection. The flow diagram indicates the steps taken in order to screen for mutations in *PALB2* using qPCR based HRMA and Sanger sequencing as screening methods. It commenced with primer optimisation and concluded with *in silico* analysis of each variant.

proteinase K and 1% (w/v) sodium dodecyl sulphate (SDS). The solution was incubated overnight at 37°C.

The following day a volume of 1.4 ml saturated NaCl was added to the mixture, where after the tubes were vigorously shaken for 15 sec and centrifuged for 15 min at 4 000 rpm at 4°C. This step was repeated twice. After centrifugation, the supernatant was transferred to a clean tube and the DNA precipitated using 2 volumes 100% (v/v) ethanol. The solution was gently mixed and incubated for 10 min at - 20°C. The precipitated DNA was scooped from the solution, transferred to an Eppendorf tube and washed with 70% (v/v) ethanol. The tubes were briefly centrifuged to pellet the DNA, where after it was air dried and dissolved in $T._1E$ buffer (10 mM Tris-HCl pH 7.6, 0.1 mM EDTA) (Combrink, 2016).

3.5.2 DNA quantity and quality

The quantity and quality of the extracted DNA was determined using the NanoDrop®ND-100 Spectrophotometer (v3.01, NanoDrop Technologies) according to the manufacturer's instructions. The concentration of the sample was expressed as $ng.\mu l^{-1}$. The purity of the extracted DNA has been proven to be critical for successful HRMA (Combrink 2016; Oosthuizen 2016), the DNA of potential research patients were inspected. Two indicators were considered, namely the ratio of absorbance at $A_{230/260}$ nm and the ratio of absorbance at $A_{260/280}$ nm. The ratio at 230 nm and 260 nm (~ 2.0 – 2.2) indicated possible DNA contaminants (Technical Bulletin, NanoDrop - http://www.nanodrop.com/Library/T042-NanoDrop-spectrophotometers-Nucleic-Acid-Purity-Ratios.pdf), whereas the ratio at 260 nm and 280 nm (ideally ~ 1.8) indicated the presence of specific contaminants such as proteins. An initial DNA dilution (with a concentration of 50 ng.µl-1 in $T._1E$) was prepared for the optimisation of the various *PALB2* fragments using PCR, where after the DNA was further diluted to a final concentration of 15 ng.µl-1 for qPCR based HRMA.

3.5.3 Primer sets for *PALB2*

The sequences of 30 primer sets used for the complete screening of *PALB2* were obtained from literature (Bogdanova *et al.*, 2011). The primers were homologous to

the reported *PALB2* gene sequence (NG_007406.1, accession MIM 600185, Genbank accession number U43746). The primer lengths varied from 18 to 29 nucleotides and produced amplicons ranging from 109 to 197 bp (Table 3.6). The primer sets were manufactured by Gibco® (Life Technologies, United Kingdom). Each dried primer was dissolved in a $T_{.1}E$ buffer to a concentration of 20 µM, where after it was further diluted to a final 3 µM for q-PCR as suggested by Combrink (2016) and Oosthuizen (2016).

3.5.4 Conventional PCR reactions

Optimal amplification of each of the primer sets was tested using a PCR gradient approach. The optimal Ta was calculated using the online tool provided by the primer manufacturer (https://www.idtdna.com). The calculation incorporated the CG content and the length of each primer pair. Once the calculated Ta was determined, optimisation of the qPCR commenced with performing PCR using a gradient. A set of PCR reactions were performed for each of the 30 primer sets across a 10ºC range, ranging from 5ºC below to 5ºC above the calculated Ta.

The components such as the dNTP's and *Taq* DNA polymerase were obtained from Roche Diagnostics (Roche Diagnostics, Mannheim, Germany). Each 50 µl PCR reaction contained 200 ng template DNA of postgraduate students, 20 µM exon specific primers, 250 µM dNTP, 100 mM Tris-HCl (pH 8.3), 1.5 mM $MgCl_2$, 50 mM KCl and 1 U Taq DNA polymerase. A no template control containing all PCR reagents except DNA was included with every run to detect possible DNA contamination of reagents. PCR was performed using the MyCycler™ (Bio-Rad Laboratories Inc., Hercules).

The PCR regime entailed one cycle at 95°C for 1 min, followed by 32 cycles at 94°C for 45 sec, the appropriate annealing temperature (Ta) (ranged from 5°C below to 5°C above the optimal temperature as indicated by Bogdanova *et al*., 2011) for 1 min and 72°C for 45 sec, with a final elongation step at 72°C for 10 min. The quality of the PCR products was visualised using horizontal gel electrophoresis.

Amplification products were separated using 3% (w/v) agarose gels in the presence of 0.5 $\mu g.ml^{-1}$ ethidium bromide (EtBr) and 1 X TBE (0.089 M Tris pH 8, 0.089 M boric acid, 2 mM EDTA), at 95 V for 30 min. The products were sized against a 50 bp DNA ladder (Roche Diagnostics, Mannheim, Germany) that acted

Table 3.6 Oligonucleotides used for qPCR based HRMA of *PALB2*. Indicated are primer names, the 5' – 3' sequence of both the forward and reverse primers and the expected amplicon size.

Exon	Forward Primer cited in the 5' to 3' direction	Reverse Primer cited in the 5' to 3' direction	Expected product size (bp)
Exon 2	ATTGTTGGTGTTTTTCTTCTTC	AATTTGAGAATACGATTCACTTA	109
Exon 3	CTGGGGCTGTTTTTGTCTCC	GTCTAGATTTACCTGAGTGTTT	139
Exon 4_1	TCCTCCTCAGAACCTAAAAAT	ATCCTCCTGGGCCATCTC	145
Exon 4_2	CCTTTAACCCTGGAGATGG	TAAATGTCCTTCTTCTGCTGC	150
Exon 4_3	AAAGCAGAAGCTGCCAAGCA	TTTGCTACTACTGATTTCTTCCTG	142
Exon 4_4	GATTCACTCAGATTGTCTGG	GTTGGGCAGTTGGTGGAAT	187
Exon 4_5	GTTACAGAAATTAATGAAGACA	CCTTTAGGAGGAATGTGTTC	170
Exon 4_6	GTAGTAGTCAGCACCTTGAA	GGTTATCTGTAGAGACAGTCA	138
Exon 4_7	CTGTAAGTTTGGAGGCACAA	GTTTTCATTTGCTGGTAAGTTA	170
Exon 4_8	AAATGAACTCACCTACAATAAC	TTGCTTCCAGGCTAAGACTC	170
Exon 4_9	GTGAGATTCTAAGTCAACCTAA	CTACTTTCCTCTGGCAATTG	156
Exon 4_10	GTTAGAACAACACGAAGCATG	GTGCCAGACATCCCTAATTTCA	176
Exon 4_11	CCAATGAGGAAACTGACCAA	GCAACTGCCTTCCTAGACAA	154
Exon 4_12	GCAGTTGCCCAAGCACCT	ACCTTTCACTTGAATAAATAATT	180
Exon 5_1	GGGTTTTGTTACTATTTTGTGA	CAGCATTCCATCCCTCTGAA	150

Table 3.6 continue

Exon	Forward Primer cited in the 5' to 3' direction	Reverse Primer cited in the 5' to 3' direction	Expected product size (bp)
Exon 5_2	GATGCTTTCACGGCTCCAT	GACTCAAAGGGCTCCACT	164
Exon 5_3	AAGTGAAGTCCTGCTCAGAA	CTAAGTCCTCCATTTCTGTAT	138
Exon 5_4	CTAAACGCATGGATACAGAAA	CTGTGGTAGGCCTGTCATTA	183
Exon 5_5	CGGTTGCGCCTGATGATAAT	GTTGGGGTGTGCAGCAAGT	150
Exon 5_6	GGACGAACTTGCTGCACAC	ACAGAGTCACAGTCACAGGT	161
Exon 5_7	GGACAACCTACCTGTGACT	TGGATTGTACCTGTTCGACG	139
Exon 6	AGACTAATGATGTGACTTTTGT	CGAGACACTGGAAGAGAATA	132
Exon 7	GCCATTGTGTCAGAATCCTT	ATTCCCACTTACCTCTGCGA	187
Exon 8	GCTGCTTTGTTTTATTTAGGTT	GGAATTACATACCTGATCTC	117
Exon 9	TATTTGGCTTAGGGCATTGTT	TTCTTACCCTCCATCTTCTG	181
Exon 10	ATTACAGAGGCAAAGAAAACC	AGCTTACCAAATAACAATGTTGT	131
Exon 11	GTGATCAGCTTATTTATTTTTG	GGTCCCAGCCAGTCATTA	142
Exon 12	GACAGTCTATTTGGGATATTTA	CACACTTGCCTGCCAGCC	197
Exon 13_1	TCTTTGTATGCTATCAGGTTC	AGAGACAGGTGGGAGGAG	132
Exon 13_2	ACTTCTCGGTCAGTGTACTG	GAGGCCCAATATATCCAGAAAA	181

as a sizing standard. The ladder consisted of equimolar mixtures of DNA fragments with a known length. This enabled correct DNA sizing of each amplicon and approximate quantification of the amplicons at the various Ta's. The optimal Ta for each amplicon were selected based on the intensity of the products across the Ta ranges, the specificity and correct size of the amplicon.

3.5.5 qPCR for HRMA

Once the optimal Ta of each amplicon had been determined using PCR, optimisation of the qPCR commenced. Each 9 µl qPCR reaction contained 30 ng genomic DNA, 0.3 µM of each primer and 3 µl LightScanner® Mastermix (BioFire Diagnostics Inc, Salt Lake City, UT) as tested and described by Oosthuizen (2016). The qPCR regime entailed: pre-incubation of one cycle at 95°C for 10 min (ramp rate of 4.4°C/s), followed by amplification steps consisting of 40 cycles at 95°C for 10 sec (ramp rate of 4.4°C/s), primer dependent Ta for 15 sec (ramp rate of 2.2°C/s) and 72°C for 10 - 25 sec, depending on the length of the amplicon (ramp rate of 4.4°C/s) (Combrink 2016). The amplified products were horizontally separated using a 3% (w/v) agarose gel to inspect the quality and specificity of each qPCR amplicon. The gels were run using 1X Tris-Borate-EDTA buffer at 95 V for 30 min.

3.5.6 HRMA

High Resolution Melting Analysis was performed on the LightCycler® 480 II real-time instrument from Roche Diagnostics (Mannheim, Germany). This system offered HRMA-based gene scanning as an integrated solution on a plate-based qPCR instrument. The LightCycler® 480 Gene Scanning Software was based upon a proprietary algorithm for melting curve grouping according to shape and was used for the identification of possible variants in the various amplicons.

For the qPCR and HRMA, the DNA was meticulously diluted to 15 ng.µl^{-1} using a low salt buffer ($T_{.1}E$) in order to eliminate the addition of possible contaminants that could influence the shape of the melting curves. The aliquots were incubated for a minimum of 16 h at 37°C to ensure a homogenously diluted

solution (Combrink, 2016). The concentration of each sample was measured in triplicate using the NanoDrop® to obtain an average concentration. This was performed in order to equalise the starting template for all samples (Oosthuizen, 2016).

An immediate high resolution melt (HRM) followed amplification and consisted of a single cycle starting at 95°C for 1 min (ramp rate of 4.4°C/s) to 40°C for 1 min to allow heteroduplex formation (ramp rate of 2.2°C/s), where after the actual melting (Tm) was achieved by gradually increasing the temperature from 60°C to 95°C. Single acquisitions were recorded during the elongation step of amplification, but were continuous (25 acquisitions per °C) during the melt. The final phase of the HRM consisted of cooling for 10 sec at 40°C (ramp rate of 4.4°C/s).

3.5.7 Automated Sanger sequencing

All HRMA patient samples deviating from the baseline (cut-off beyond -2.5 to 2.5 for the relative signal difference) in the difference plots were bi-directionally sequenced using the BigDye® Terminator v3.1 Cycle Sequencing Kit (Thermo Scientific Corp, Waltham, MA). Cycle sequencing was performed according to the manufacturer's instructions to identify genetic variants. Before sequencing, each qPCR based HRMA product was purified using Illustra™ Exostar® (GE Healthcare, Chicago, Illinois) to ensure the removal of single stranded DNA such as primers and single nucleotides. This step entailed the addition of 2 µl of the enzyme to 5 µl of the qPCR product. The solution was briefly mixed by pipetting, where after it was incubated for 15 min at 37°C. This incubation was followed by a second incubation step at 80°C for 15 min to inactivate the enzyme.

Each 10 µl sequencing reaction contained 3 µl cleaned-up PCR product, 1 µl Ready Reaction master mix, 3.2 µM primer (forward or reverse) and 2 µl BigDye® sequencing buffer. The amplification regime entailed: one cycle at 96°C for 1 min, followed by 25 cycles at 96°C for 10 sec, 54°C for 5 sec and 60°C for 4 min, with a final holding temperature at 4°C. The sequenced products were precipitated by adding 5 µl 125 mM EDTA and 60 µl 100% (v/v) ethanol according to the manufacturer's protocol, followed by a 30 min incubation step at room temperature

in the dark. After centrifugation at 14 000 rpm for 30 min at room temperature, the supernatant was aspirated, the pellet washed once with 70% (v/v) ethanol for a minimum of 30 min and finally air dried in the dark.

Hi-Di™ formamide buffer was added to the dried pellet, where after it was denatured for 5 min at 96°C and the tubes snap-cooled in a cold block (0 – 5°C). The pellets were suspended by pipetting, and analysed on an ABI Genetic Analyser. The electropherograms were analysed by visual inspection using proprietary sequence analysis software (Chromas version 2.31, www.technelysium.com.au). The sequences were aligned to the reference sequence for *PALB2* with LALIGN (www.ch.embnet.org/software/LALIGN), after which translation was performed with the Expasy translate tool (http://au.expasy.org/tools/dna.html).

3.5.8 Quality control

The use of qPCR based HRMA had previously been optimized for use in the laboratory, hence only optimisation of the various *PALB2* primer sets was performed. Optimisation was critical to ensure mutation screening was performed using optimal conditions. Critical factors affecting HRMA were taken into consideration such as the shape and incline of the amplification curve, the crossing point (Cp) of the reactions (should be between 20 – 30 cycles), the total number of fluorescence units obtained at the end of cycling and a single Tm of the amplified product (to ensure amplification specificity). These factors were incorporated to determine the optimal Ta and cycling regime for each of the primer sets.

During HRMA, all samples were performed in duplicate in order to prevent false positive or negative results obtained due to pipetting errors. The first reaction therefore acted as a control for the second. A true positive was considered only if both samples exhibited the same HRMA profile on the difference plot and was identified by the instrument's software as similar and distinctly different from the rest (deviating from the base line).

As HRMA represents a mutation screening technique, the study did not make use of any positive controls. All experiments included a no-template control and an

empty well. The no-template control provided evidence of no reagent cross-contamination during the execution of the laboratory experiments, whereas the empty well reflected a clean plate.

During normalisation of the HRMA data, the gliders were used to limit analysed data only to that surrounding the area of melting of the amplicon. This excluded background noise that could potentially influence the data observed on the difference plot. All samples deviating from the baseline (< - 2.5 and > 2.5) and not grouping with the bulk of the samples, were Sanger sequenced to confirm potential DNA changes. Sequencing was performed in both directions and a variant was only called once both reactions displayed the identical DNA change.

3.5.9 Data analysis

Mutations were described according to the guidelines stipulated by the HGVS nomenclature (http://www.hgvs.org/mutnomen/) and were correlated with that previously reported. Each variant was initially checked for its existence by performing an internet based search. Databases such as the Leiden Open Variant Database (LOVD, http://www.lovd.nl/3.0/home) and ClinVar (http://www.ncbi.nlm.nih.gov/clinvar/) allocated a class (1 – 5) to each of the known variants based on the ENIGMA guidelines (http://enigmaconsortium.org/documents/publications/ENIGMA_Rules_2017-06-29 v2.5.pdf) for variant classification (Appendix H).

Variants that were not found in these databases were searched for in the 1000 Genomes project database by using dbSNP (https://www.ncbi.nlm.nih.gov/projects/SNP/) to investigate whether the variant was novel or allocated to a specific population cluster. On many occasions this search delivered a reference single nucleotide polymorphism (SNP) number for the variant (Reference SNP cluster ID or rs number), indicating that the variant had been reported and peer reviewed but not reported to a certain database. The rs number represented an accession number used by researchers and databases to refer to a specific SNP.

The analysis of variants with an unknown clinical significance (VUS) was conducted using a multiple evidence based approach according to the ENIGMA guidelines (Appendix H). These multi-evidence guidelines represented a universal and standardised approach to assess the pathogenicity of a variant. Once the population data was analysed, the variants were grouped according to mutation or variant type, namely nonsense, missense or synonymous variants. Nonsense variants lead to terminating codons which in most cases are pathogenic. Missense variants were examined with computational functional analysis, using Polymorphism Phenotyping version 2 (PolyPhen-2, http://genetics.bwh.harvard.edu/pph2/), Sorting Intolerant From Tolerant (SIFT, http://sift.jcvi.org/) and Mutation Taster (http://www.mutationtaster.org/) to determine the variants' conservation across species and to assess possible protein structure disruption. Variants that delivered inconclusive results across these databases were then analysed for splice effect predictions. Human Splicing Finder (http://www.umd.be/HSF3/HSF.html) was used to determine splice defects. Finally, all variants that showed possible clinical impact (Class 4 to 5) were then searched across the literature for published functional assay confirmation to confirm or substantiate the classification class.

Chapter 4

Results and discussion

The previous chapter, methodology, precisely stated in detail, the materials and methods applied in the study to obtain data. It described the ethical considerations and the process of patient selection. It further outlined the laboratory experiments which included DNA extraction and dilution. It further more described the mutation screening methods, which was based on qPCR based HRMA and DNA Sanger sequencing, as well as the guidelines and internet based websites used for the interpretation of the data acquired.

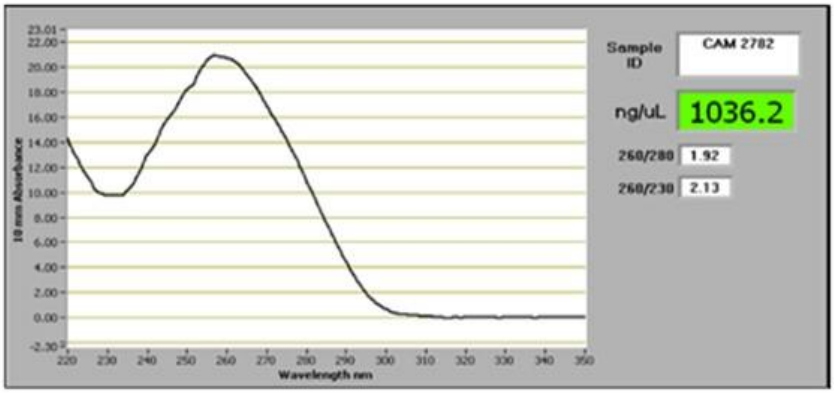
The purpose of this chapter is to then present, interpret and explain the results obtained in order to answer the research question stated in chapter 2. As a way of justifying the findings presented in this chapter, the study used a positivist paradigm approach, which states that valid/true knowledge was only derived from sensory experience and logical interpretation of data as opposed to mere intuition (Taylor & Medina, 2013). Based on this theory, only results that could be scientifically verified was included.

The interpretation of the observed data was based on the results of the *in silico* analysis that predicted the effect of DNA variants on the biochemical alterations of respective aa regarding protein folding and function. These findings were compared with those existing in the literature in order to formulate a hypothesis that either justified or dismissed the current knowledge.

4.1 Familial breast- and/or ovarian ca patients

The selection criteria allowed for inclusion of 86 BC/OVC patients representing the four main ethnic groups within SA (Table 3.1 – 3.4). The non-Afrikaner Caucasian patients were grouped as White and represented a fifth group (Table 3.5). The term “White” therefore refers to non-Afrikaner Caucasian women, for the purposes of this study. All these participants had sufficient stock DNA of a good quality and quantity in order to perform the screen (Figure 4.1 A – B).

A



B

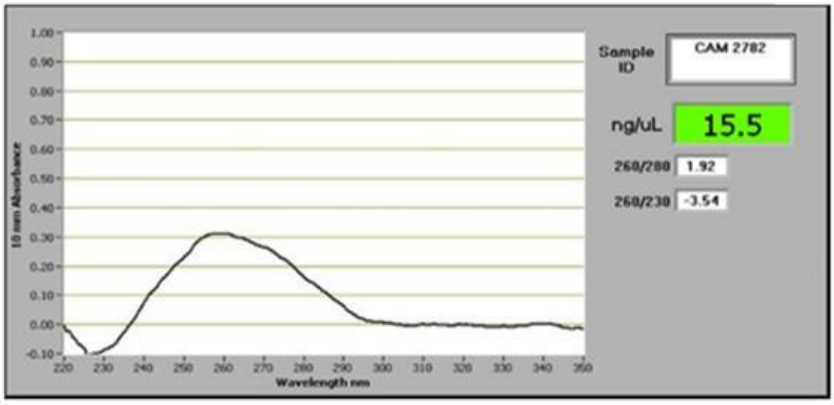


Figure 4.1 Inspection of the quantity and quality of the DNA for CAM2792 as determined by the NanoDrop®ND-100 Spectrophotometer v3.01. **A.** DNA concentration and quality values of the stock DNA at $A_{260/280}$ nm and $A_{260/230}$ nm. **B.** DNA concentration and quality values of the DNA dilution required for qPCR based HRMA, indicating the values at $A_{260/280}$ nm and $A_{260/230}$ nm.

4.1.1 Various ca types in families

Participants affected with BC and/or OVC were enrolled into the study based on a positive family history of breast- and other ca types. An example of an extensive family history of breast- and other ca types are indicated in the pedigree presented in Figure 4.2.

Despite the majority being affected with BC and OVC, various other ca types were observed (Figure 4.3). The five most common ca types present in these families were breast- (19%), prostate- (13%), stomach- (11%), ovarian- (9%) and colon ca (9%). These ca types were observed across all ethnic groups (CANSA, 2012). Four of these are represented in the top five most common ca for both men and women in SA, and account for the majority of ca cases (CANSA, 2012). For SA women, the most common ca are breast, cervical, ca of Unknown primary (CUP), colorectal and uterus ca. For men, the most common ca types include prostate-, colorectal-, lung ca, CUP and Kaposi Sarcoma (CANSA, 2012).

Although various ca types were observed amongst the families, BC remained the most common (Figure 4.4). In fact, ~ 19% of patients had a history of BC only, with 84 of the 86 affected patients presenting with at least one family member diagnosed with the disease. All the families studied, showed a clear indication of a familial ca syndrome, with an aggregation of various ca types amidst all the BC cases.

The Coloured population presented with the widest ca range of all (Figure 4.4). Amongst the 29 families studied, a total of 16 different ca types were observed, with especially prostate-, colon-, throat-, ovarian- and stomach ca repeatedly noted accompanying BC. Although these are not necessarily the most common ca types recorded for the SA Coloured population, they represent the most commonly diagnosed ca types in families of all SA ethnic groups (CANSA, 2012). This might be attributed to the origin and heritage of this unique SA population, as they genetically overlap between the indigenous African (Blacks), European and Indian populations that emigrated to SA during the last 300 years (Quintana-Murci *et al.,* 2010). Based on Figure 4.4, it is speculated that their susceptibility to genetically predisposed ca might be broader than that of the genetically more conserved ethnic groups.

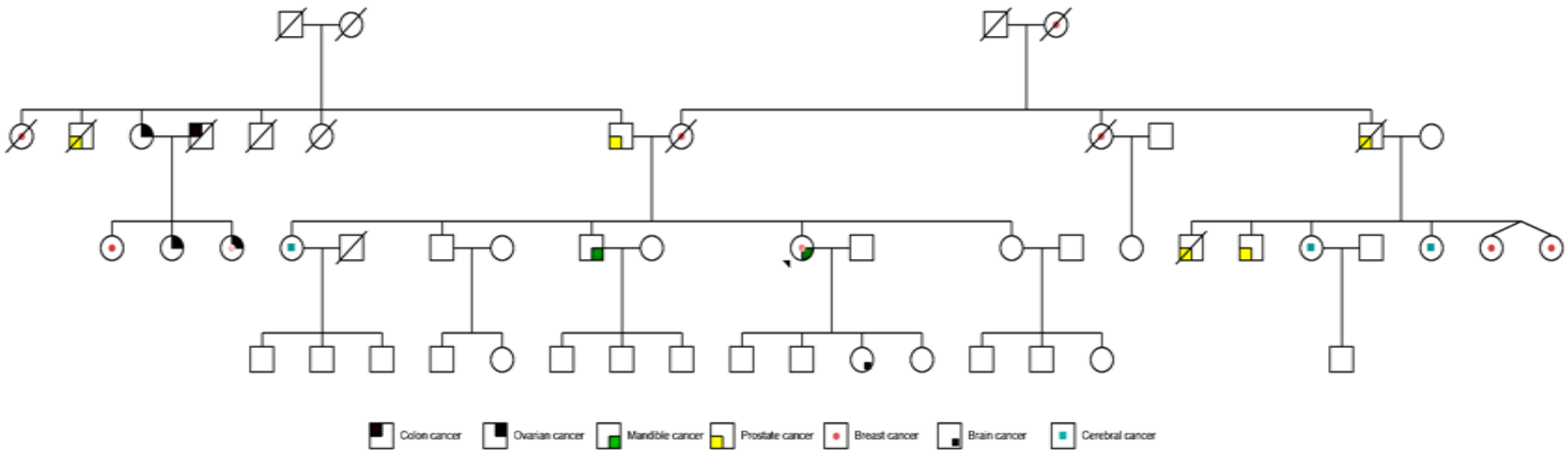


Figure 4.2 Example of a family pedigree that met the selection criteria. Presented is the pedigree for CAM2782, indicating a familial clustering of various ca types on both maternal and paternal sides. The index patient currently age 50, was diagnosed with both breast- and mandible ca. A total of 22 family members (including the index and an unrelated spouse) has been diagnosed with various ca types. The description for each ca, as well as the position of the index patient are indicated (pedigree constructed using https://pedigree.progenygenetics.com/, accessed on 12 November 2017).

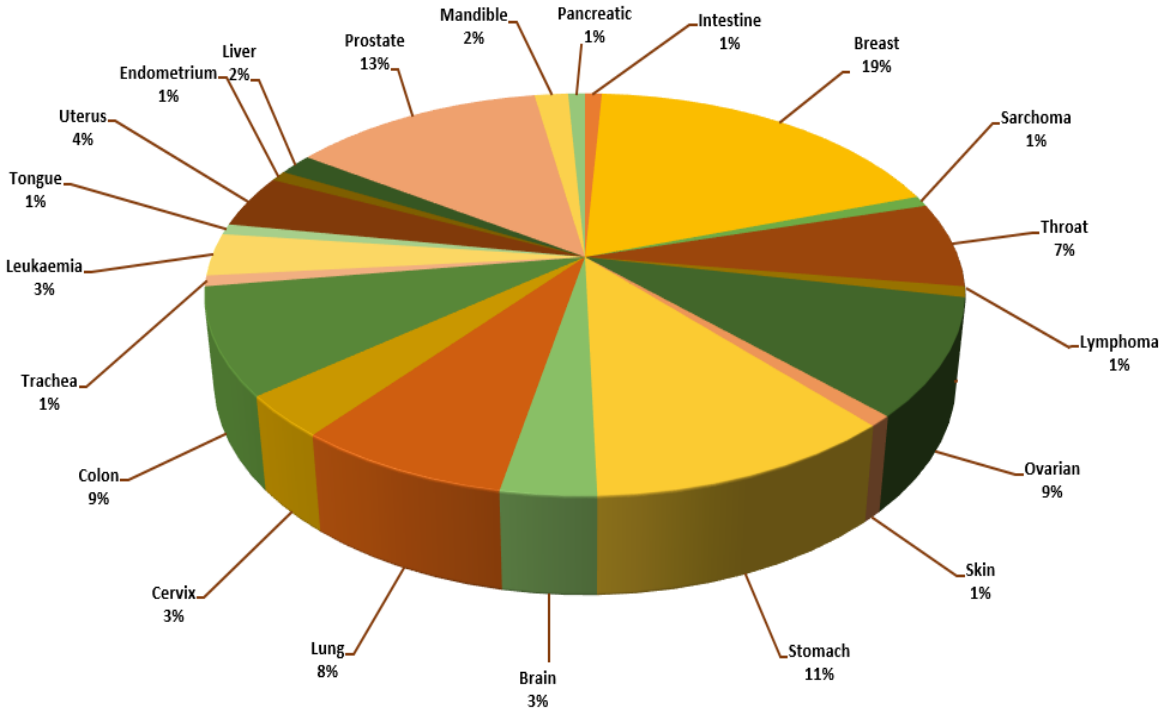


Figure 4.3 Representation of the different ca types recorded in the family pedigrees of the 86 participants. The prevalence of each ca type is indicated as a percentage.

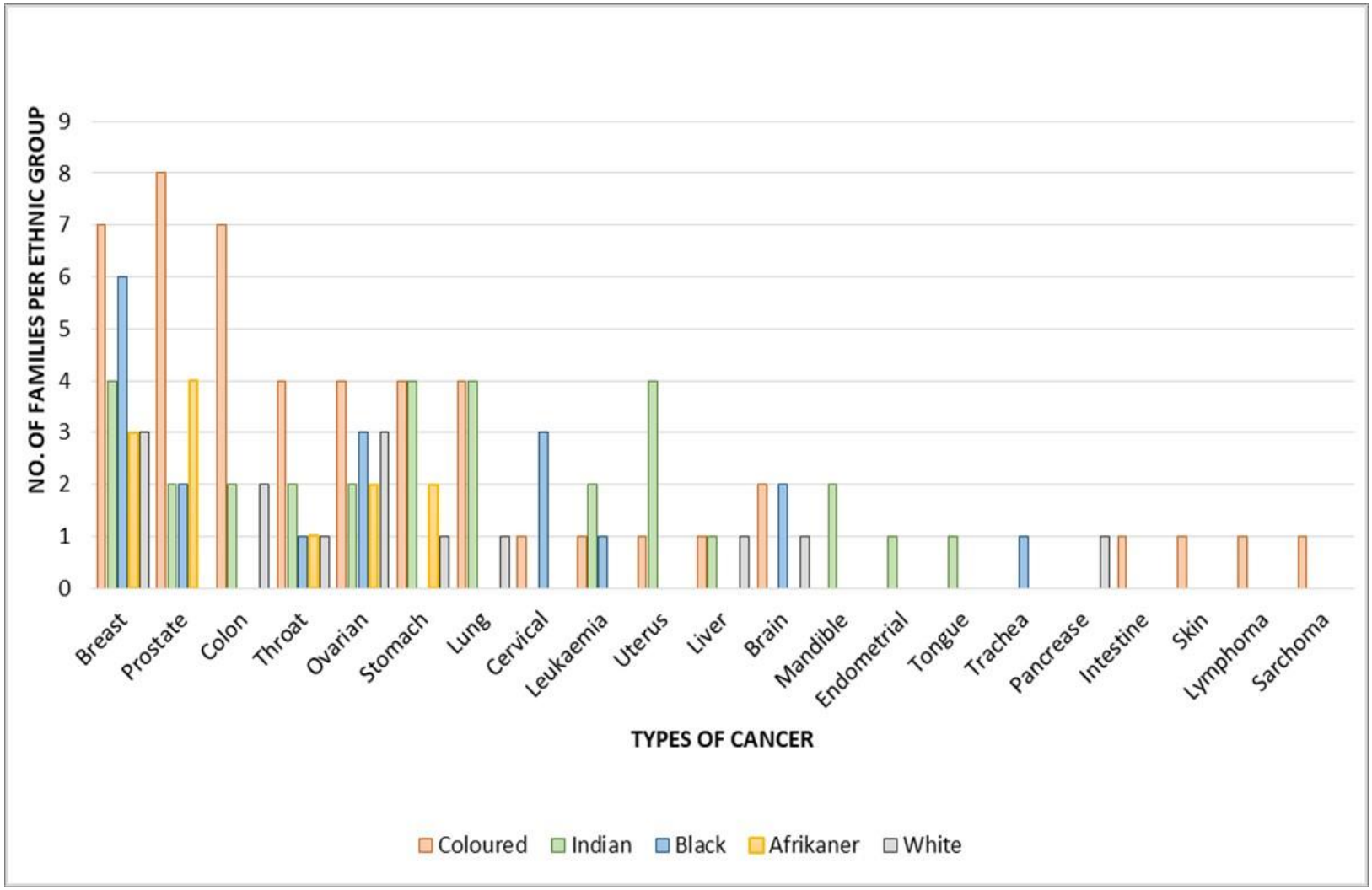


Figure 4.4 Prevalence of various ca types present in the families involved, recorded for each SA ethnic group.

Similarly, the SA Indian population also displayed a broad range of ca types, with 13 different ca types recorded in a total of 22 families. Although the range was also broad, it differed with regards to that observed for the Coloured population (Figure 4.4). This population group represents the largest settlement of Indians outside mainland India. This SA group is also genetically unique (based on SNP data obtained during *BRCA1/2* screening), as many of the SNPs and mutations detected, are specific to this group of individuals and have not previously been described for the original population of India (Roychoudhury *et al*., 2001; Mastana, 2014; Combrick *et al.,* 2016).

The Afrikaner families exhibited the smallest range of ca types. In a total of 10 families, only six different ca types were observed. These included breast, prostate, throat, ovarian, stomach and lung ca. These ca types represented the typical common ca types observed in the cohort. The small range of ca types observed does not necessarily imply that the Afrikaner ca profile is markedly different from the rest. This might not be a true reflection, as a low number of Afrikaner families were included in this study (10 of 86).

4.1.2 Affected family members

Literature has stated that the risk or likelihood of developing ca increases in accordance with the number of affected family members in the family. For example, for *BRCA1/2* it was found that for patients with four or more BC or OVC affected family members, the odds of detecting a pathogenic *BRCA* mutation increased to ~ 85% (Foulkes *et al.,* 2007). When considering all 86 index patients, the average number of family members diagnosed with breast- and other ca was 4.6 although the numbers varied (Figure 4.5 A – E).

The group with the most extensive family histories was the Afrikaners with an average of 5.7 affected members per family (Figure 4.5 D). In this group, the highest number of affected members was 10, with the lowest being three members per family (Figure 4.5 D). Succeeding the Afrikaner population, was the Coloured, Whites and Indians with 5, 4.5 and 4.4 family members respectively (Figure 4.5 A, B & E). The Black population had a more moderate family history with an average of 3.7 affected members per family (Figure 4.5 C). Along with genetic constitution,

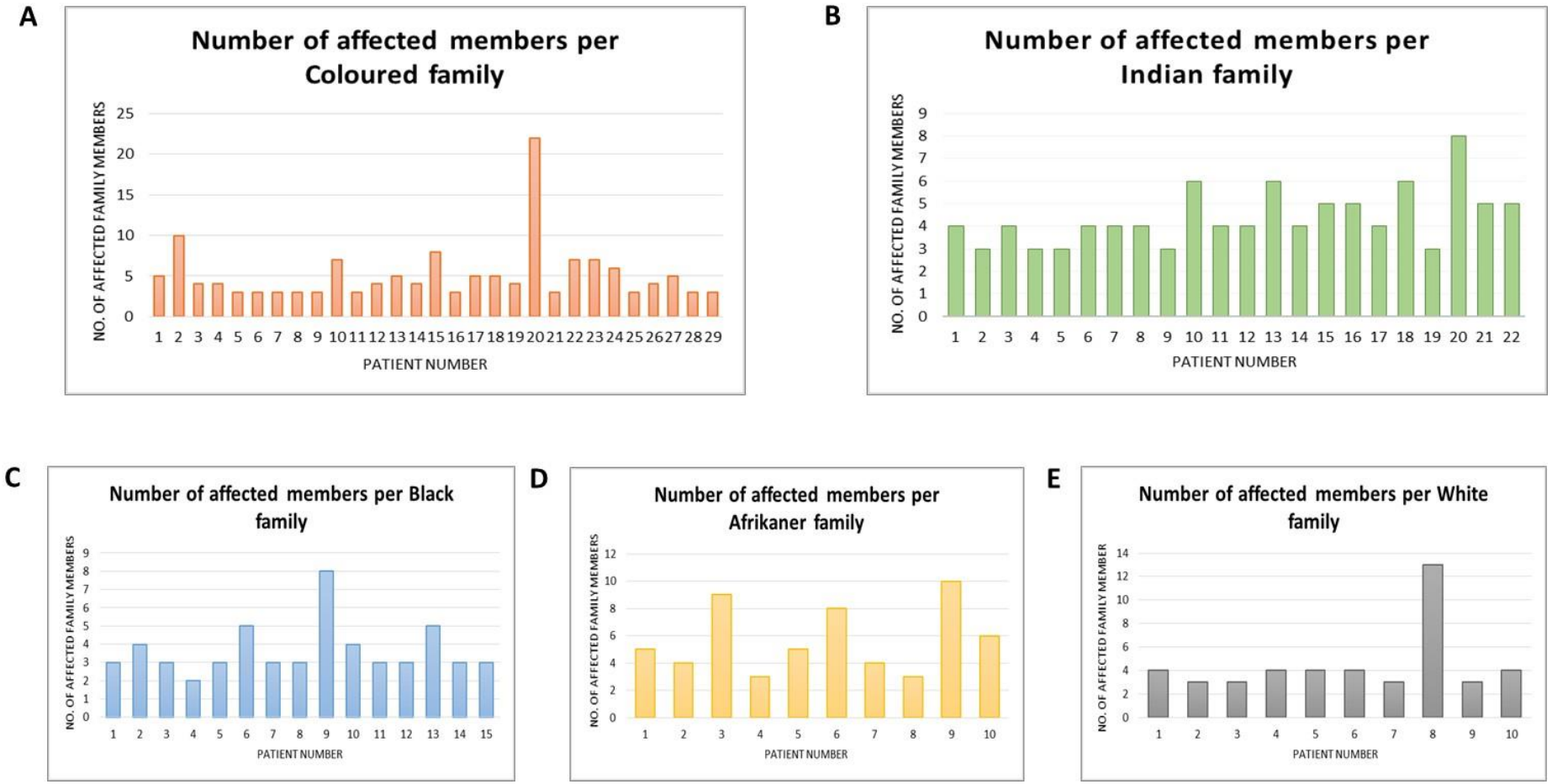


Figure 4.5 Family history of patients per ethnic group. **A.** Families of Coloured patients, **B.** Families of Indian patients, **C.** Families of Black patients, **D.** Families of Afrikaner patients and **E**. Families of White patients. The Afrikaner population had the most extensive history while Black families had the least number of members diagnosed with ca.

these high numbers are possible due to a more westernised lifestyle and cultural practices that have been maintained throughout generations. The lower average in the Black population could be due to a few reasons, which include lifestyle choices such as having many children, breastfeeding for longer periods and healthier dietary choices (non-Westernise) eating more vegetables and less meat. Another important factor that could contribute to this relative low number is the lack of knowledge regarding the medical history of families (Colditz *et al.,* 2006; CANSA, 2012). It is not custom in the Black population to discuss ca and other serious illnesses with the extended family due to religious and traditional beliefs.

The data illustrated in Figure 4.5 confirm to some extent the existing hypothesis that the overall life-time risk of developing BC in SA is higher for Whites than it is for Blacks. In 2004, the risk of developing BC was 1:13 for the Afrikaner and 1:81 for Black SA females. Currently, the risk is 1:15 for Afrikaner compared to 1:53 for Black women (Reeves *et al.,* 2004; Francies *et al.,* 2015). Despite the lower risk for Black females, it is rapidly increasing (2004 versus 2015). This is most likely due to the adoption of a more Western lifestyle (van der Merwe *et al.*, 2012).

4.1.3 Age of diagnosis

The risk of developing BC has been proven to increase with age (Frank, 2004; Erkko *et al.,* 2008; Antoniou *et al.,* 2014). This is evident in countries all over the world including SA. It is also known that White women are more likely to develop BC than African American, Hispanic and Asian women (Chupek *et al*., 2015). However, African and African American women are more likely to develop more aggressive, more advanced-stage BC that is diagnosed at a young age (<45) compared to other population groups (CANSA, 2012; Chupek *et al*., 2014; Keenan *et al*., 2015).

In the current study, the average age at diagnosis for patients was 46 years with the majority for the groups ranging between 46 and 49 (Figure 4.6). White patients generally had the most advanced age at diagnosis, namely 49.1 years whereas Black females had the lowest (dx = 39). According to CANSA (2012), ~ 50% of White females are diagnosed with BC between the ages of 50 and 70 years while the same percentage was diagnosed between 40 and 60 years of age in the Black population. Up to now, it is unclear what is causing this difference particularly

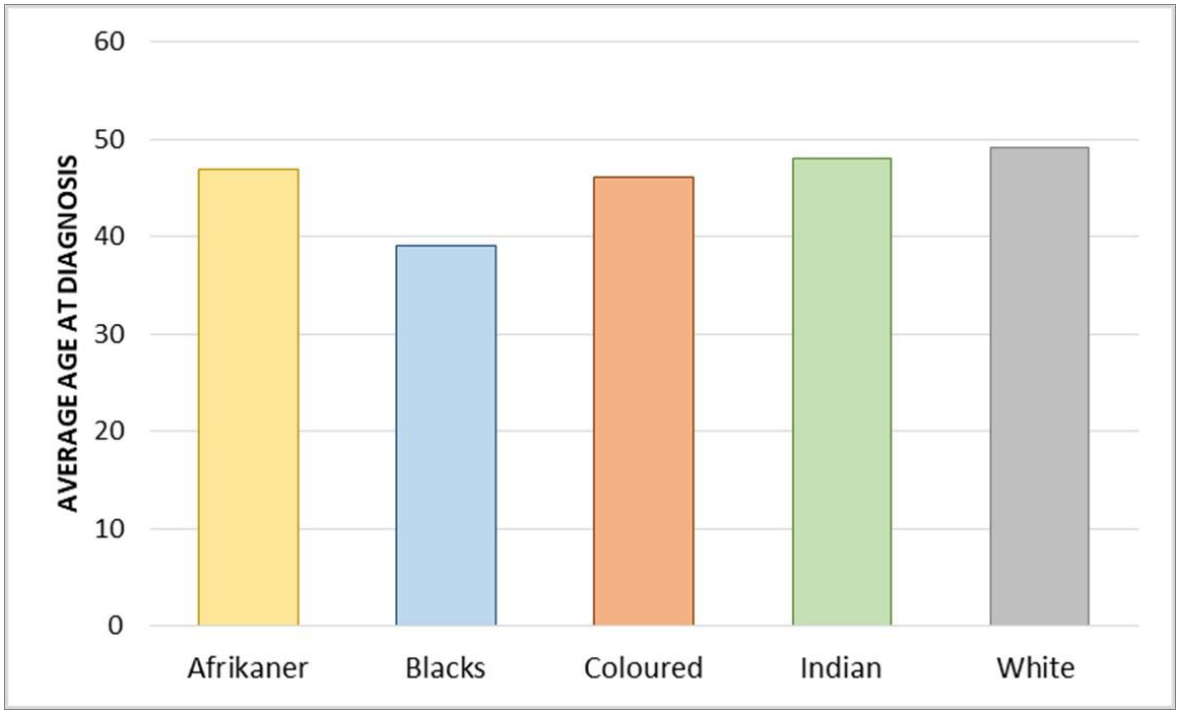


Figure 4.6 Average age at diagnosis illustrated per ethnic group for the total 86 index patients affected with breast- or OVC.

in *BRCA* negative patients, as age at diagnosis is often correlated to mutations in either *BRCA1* or *BRCA2* (Verhoog *et al.,* 1998; Xia *et al.,* 2007; Foulkes, 2008; Roy *et al.,* 2012).

4.2 *PALB2* primer optimisation

For the majority of the primer sets, the calculated Ta proved to be best as PCR resulted in a strong single product that corresponded with the expected fragment size (Table 4.1). For the remainder of the primer sets, more specific amplification results were obtained at higher Ta's, once visualized on an agarose gel (Figure 4.7).

Once the most stringent Ta was established for each primer set using gradient PCR, the Ta was used within an initial qPCR reaction to determine its specificity and performance. The qPCR regime as published within the package insert of the High Resolution Melting Dye designed for the LC480 II from Roche (Roche Diagnostics) proved to be optimal for the majority of primer sets.

To ensure optimal Ta, various factors were taken into account. This included inspection of the shape of the sigmoidal amplification curve, looking specifically at the Cp, the exponential phase and the final plateau phase. The Cp value represented the point where amplification first clearly rose above the noise thresholds (off the baseline) to a statistically significant degree. This value should ideally range between 20 – 30 cycles for qPCR based HRMA (van der Stoep *et al.*, 2008).

Once the Cp was crossed, the slope of the curve needed to be steep to represent the strong exponential amplification of the selected region. This occurred when the amplicon-derived fluorescent signal was reliably detected by the instrument. Once the limited resources provided (such as the primers, nucleotides and enzyme activity) were utilized, amplification decreased and the amplification curve tailed off and plateaued.

Additionally, a melt curve analysis or Tm calling was performed post qPCR to determine the specificity of each Ta. This was performed on the results that were acquired during the qPCR. The gradual increase in temperature resulted in the denaturation of the amplicon at a given temperature. This resulted in a decrease in fluorescence which was then converted into melt curve data to generate the melting

Table 4.1 Optimized annealing temperatures for qPCR based HRMA of *PALB2*. Indicated are primer names and the optimized Ta of each primer pair.

Exon	**Ta (°C)**
Exon 2	54
Exon 3	56
Exon 4_1	60
Exon 4_2	62
Exon 4_3	54
Exon 4_4	58
Exon 4_5	54
Exon 4_6	60
Exon 4_7	54
Exon 4_8	54
Exon 4_9	54
Exon 4_10	60
Exon 4_11	54
Exon 4_12	54
Exon 5_1	60
Exon 5_2	64
Exon 5_3	54
Exon 5_4	58
Exon 5_5	56
Exon 5_6	62
Exon 5_7	54
Exon 6	58
Exon 7	54
Exon 8	56
Exon 9	54
Exon 10	60
Exon 11	60
Exon 12	54
Exon 13_1	64
Exon 13_2	58

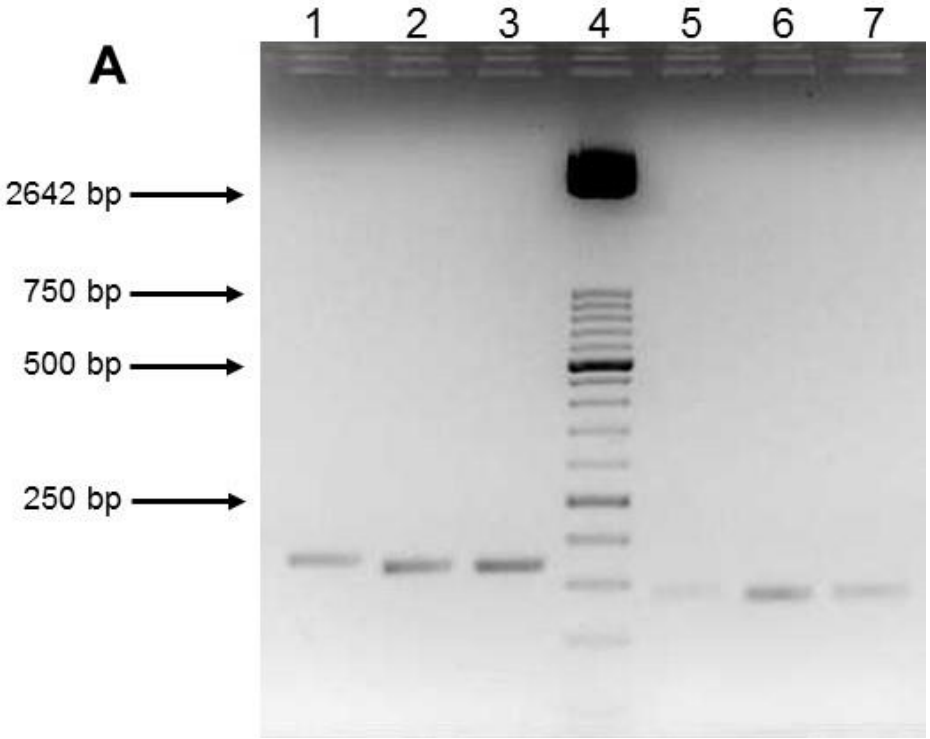


Figure 4.7 Agarose visualization of optimized PCR products for various *PALB2* primer sets. Loaded in Lane 1 – *PALB2* exon 4_12 (180 bp, Ta of 56°C); Lane 2 – *PALB2* exon 4_7 (170 bp, Ta of 56°C); Lane 3 – *PALB2* exon 4_8 (170 bp; Ta of 56°C); Lane 4 – 50 bp DNA molecular marker; Lane 5 – *PALB2* exon 4_1 (145 bp; Ta of 60°C); Lane 6 – *PALB2* exon 5_3 (138 bp; Ta of 54°C); Lane 7 – *PALB2* exon 4_3 (142 bp; Ta of 54°C).

peaks. The presence of more than one peak indicated either the presence of a second non-specific amplicon or primer dimers.

All the primer sets produced amplification curves exhibiting all three phases of amplification using the initial optimized PCR Ta. The samples representing a specific exon grouped and had a similar Cp (Fig. 4.8 A). All the primer sets produced a Cp of between 25 and 30 cycles, which according to literature was ideal (Erali *et al*., 2008; Vossen *et al.*, 2009). A total number of 45 cycles were sufficient for all qPCRs to plateau (Figure 4.8 A).

The slopes of the sigmoidal amplification curves of the 30 sets were strong, although the fluorescence units differed between the fragments at the plateau phase. The number of fluorescence units varied between 180 and 360. As the relative fluorescence units were a reflection of the quantity of the amplicon, it was imperative that an attempt was made to increase the fluorescence units without jeopardizing the sensitivity of the reaction.

In the majority of cases, evidence of non-specific binding was only observed once the Tm calling was performed, as more than one melting peak was present (Figure 4.8 C). In some instances such as *PALB2* exon 13_1, the presence of a secondary amplicon was already evident in the normalized melt curve (Figure 4.8 B). In order to increase the stringency of the amplification, the Ta of the qPCR was gradually increased from 58°C to 64°C (Figure 4.8). With this increase, the secondary peak became smaller, indicating more specific amplification. At 64°C, only a single amplicon was present (Figure 4.8 C). This process was followed until nearly all qPCR reactions produced only a single specific amplicon for the various regions of interest. The optimized Ta for the qPCR is listed in Table 4.1.

For a small number of primer sets, a secondary peak remained despite all attempts at further optimisation (increase in Ta and the addition of $MgCl_2$) (Fig. 4.9 A -C). From literature, it was gathered that these secondary peaks could be due to the inherent nucleotide sequence of that fragment (Krypuy *et al.,* 2006; Erali *et al*., 2008; Vossen *et al.,* 2009). These areas tend to exhibit multiple melting domains. To determine whether this primer set was fully optimized, the fragment was Sanger sequenced (Figure 4.9 D). The sequence obtained was clean with high Phred scores and barely any background noise. This indicated the presence of a single amplicon and absence of non-specific binding of the primers. The sequence further

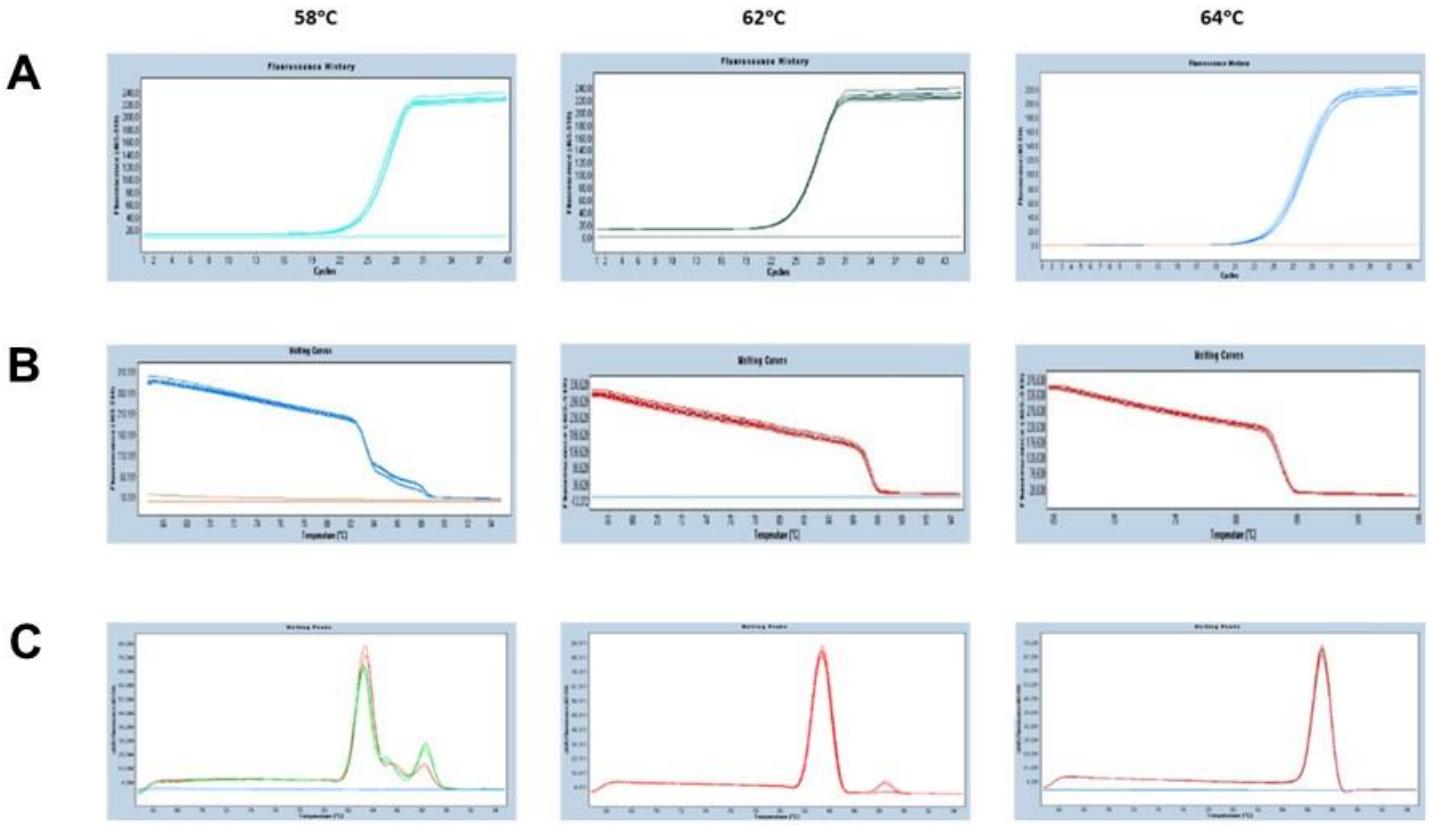


Figure 4.8 Optimisation of the Ta for *PALB2* exon 13_1 to be used within qPCR based HRMA. **A. q**PCR amplification curves for reactions performed at 58°C, 62°C and 64°C. **B.** Normalized melting curves obtained during qPCR amplification performed at 58°C, 62°C and 64°C. **C.** Graphs obtained for post qPCR Tm calling for amplification performed at 58°C, 62°C and 64°C.

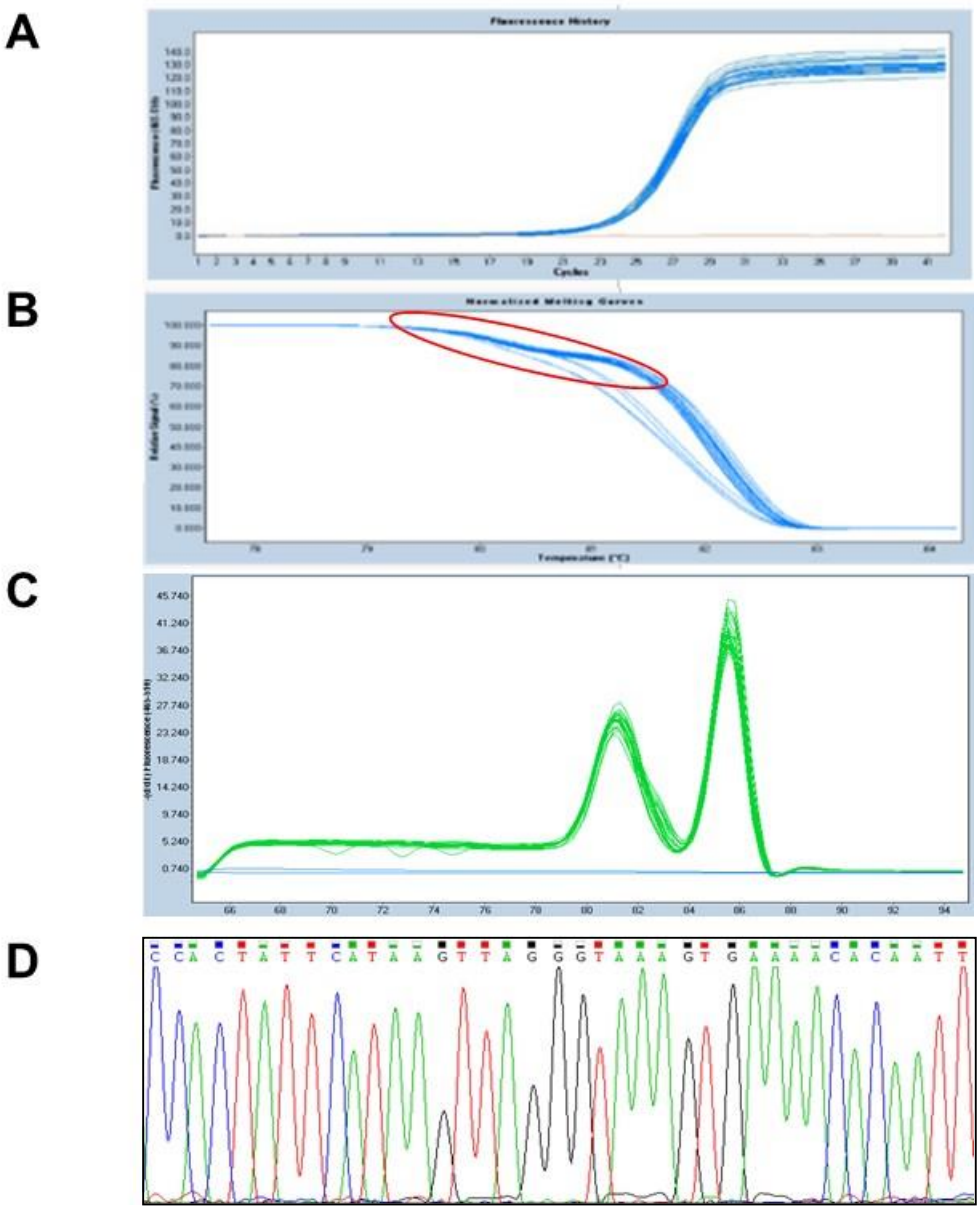


Figure 4.9. Evidence of more than one melting domains within *PALB2* exon 13_2. **A**. The amplification curve. **B**. Normalization of the melting curve, indicating two melting domains, highlighted by the red circle. **C**. Tm calling indicating two distinct melting domains. **D**. DNA sequence for *PALB2* exon 13_2, indicating the presence of a single amplification product with high Phred scores.

more aligned with the reference sequence for *PALB2* exon 13_2, indicating that although two distinct melting peaks were detected using Tm calling (Figure 4.9 C), the amplification was highly specific and optimal.

Once the optimal Ta of the qPCR was established, HRMA was performed using the data obtained from the qPCR. Gene scanning for HRMA made use of the Gene Scanning software v1.5 specific for the Light Cycler® 480II. The analysis was performed according to the Operator's Manual (Roche Diagnostics, Mannheim, Germany). The data obtained during HRMA was visualized as various graphs, which included the qPCR amplification curve, a negative calling, normalization of the amplification curves (by placement of the gliders), a temperature shift and a difference plot (Figure 4.10 A – D).

The normalisation, temperature shift, as well as the difference plot were all derived from the negative calling of the HRMA (Figure 4.10 B). The negative curve was captured during qPCR with the acquisitions taken from 60°C to 96°C for a 5 min period at a rate of 25 acquisitions per 1°C. The criteria for normalisation composed of moving the gliders to the area just before and after the melt, and adjusting them in such a way that the two bars were 1°C apart on both sides (Figure 4.10 C). The temperature shift was adjusted to exclude the last 3 – 5% of the melt curve data (Figure 4.10 D). The difference plot was performed using the default sensitivity settings and algorithms suggested by the manual of the Gene Scanning software (Figure 4.10 F). If a relative signal difference of < -2.5 or > 2.5 from the baseline occurred, the sample was considered different from the rest based on the difference in fluorescence intensity.

4.3 *PALB2* mutation screening results

4.3.1 Classification of *PALB2* mutations

Mutations were described according to the recommended nomenclature system described by the Human Genome Variation Society (HGVS; ref. 42) with nucleotides numbered from the A of the ATG translation initiation codon of the National Centre for Biotechnology Information (NCBI) reference sequence

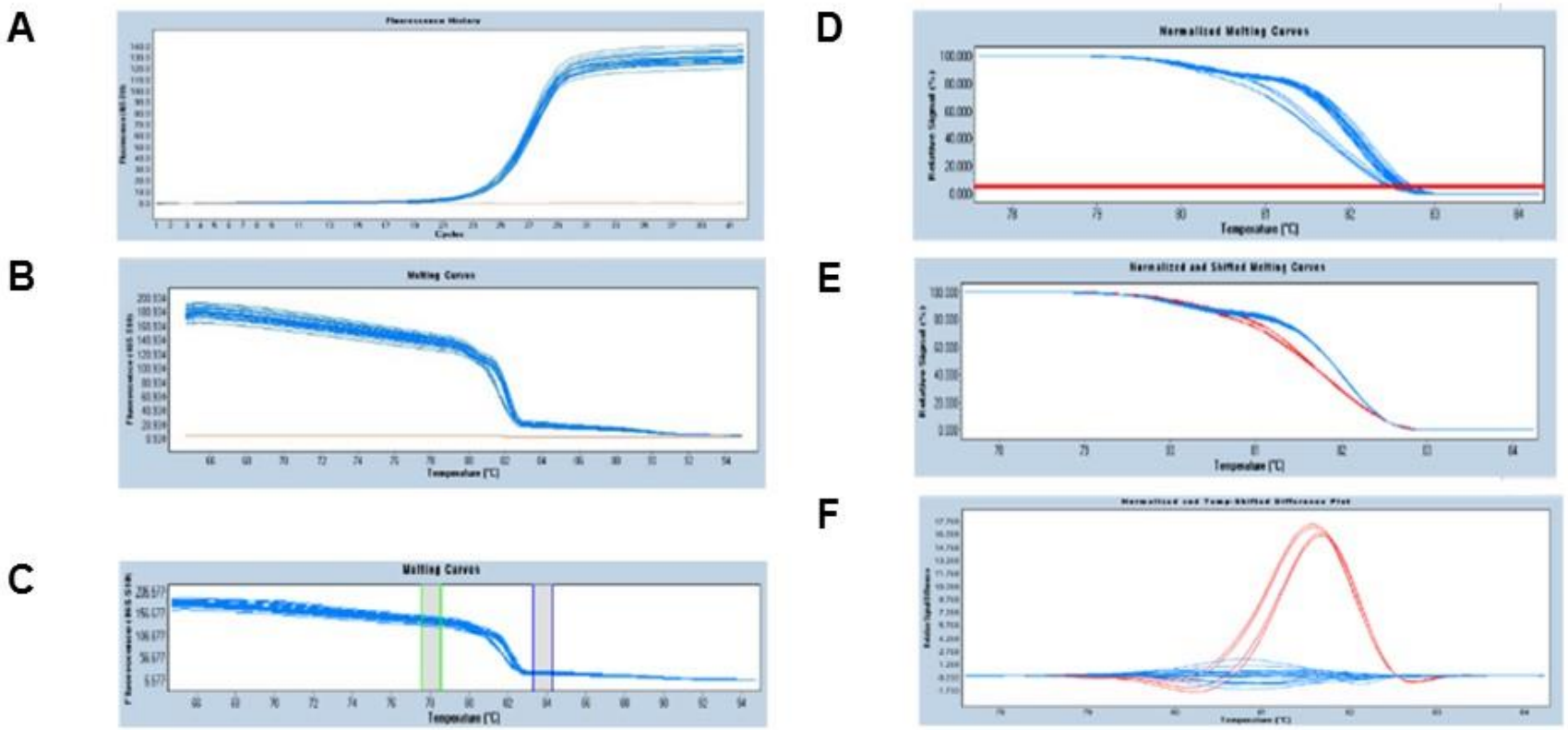


Figure 4.10 Graphs utilized for the visualization of HRMA results for *PALB2*. **A**. The qPCR amplification curve. **B**. The negative calling. **C**. Normalisation of the amplification curves, indicating the placement of the gliders before and after the melt. **D**. The temperature shift. **E**. The normalized and shifted melting curve. **F.** The HRMA difference plot for the experiment, with the variant (red curve) grouping separate from the bulk of the samples (representing the wild types) on the baseline, with a relative signal difference of 17.0 units.

NM_024675.3). Genomic and protein sequences for *PALB2* were obtained from NG_007406.1, and NP_078951.2.

Variant classification commenced with determining whether the variant had been detected before. *PALB2* variants were defined as novel when they were not listed in any databases such as ClinVar, dbSNP, 1000 Genomes (http://www.ncbi.nlm.gov/tools/1000genomes/) and HGMD (http://www.hgmd.cf.ac.uk/ac/). ClinVar includes listings by several diagnostic companies that classify mutations on the basis of variant types, *in silico* functionality predictions, population frequencies [Minor allele frequencies (MAF) scores] and familial pathogenicity segregation. The MAF gave an indication of the frequency of the variant in the various genomes. The entries into ClinVar are regularly reviewed based on the number and dates of the submissions. If the variant has been described before, the rs number was recorded and used for future reference.

The three most common *in silico* tools that were used to predict the impact of the various single base changes on protein function included SIFT, PolyPhen-2 and Mutation Taster. For the majority of cases, the ENIGMA classification was used to determine the clinical impact, if any, of the variant on the disease, substantiated by peer review publications where possible (Appendix H).

Only rare variants (MAF <0.01) qualified to be VUS. For this purpose, allele frequencies for the closest relative population were obtained from the 1000 Genomes database, if available. This database was used to determine whether the variant was restricted to a single continental gene pool or present in multiple global populations. If the variant has been reported for various global populations at a frequency >0.01 or 1%, it was considered benign. If the variant has been observed in a single population at an allelic frequency of <0.01 or <1%, it was considered to be likely benign. Variants that had the potential to possibly alter the silencing and enhancing elements at DNA level, was classified as VUS until more information could be obtained. All VUS variants were compared to the patient's clinical data and relevant literature. If no evidential literature was obtained, *in silico* analysis was performed to make a final functionality prediction (Oosthuizen, 2016).

4.3.2 Mutation screening results for *PALB2*

HRMA revealed the presence of various DNA changes within the cohort of 86 BC and OVC patients. All the changes represented single base variations. No small deletions or insertions were detected, which would have resulted in a shift in the reading frame. All of the single nucleotide variations (SNV) were located within the coding regions of the gene, with the majority being observed in exon 4 (Tables 4.2 & 4.3). This did not necessarily indicate a mutational hot spot, but was more attributed to the size of the specific exon, as exons 4 and 5 represent more than 50% of the entire coding region (Figure 4.11).

The classification of these SNVs were problematic, as to date, no *PALB2* missense mutation, apart from some clearly truncating mutations, has been classified as definitely pathogenic in large population studies (Tischkowitz *et al.*, 2012; Kluska *et al.*, 2017). The only large case control performed to date was by Southey and co-workers (2016) and Thompson *et al.* (2015), in which they proved strong evidence of association with BC risk for *PALB* c.3113G>A, p.Trp1038* (rs180177132, OR of 4.21 (95% CI 1.84 to 9.60, $p=6.9x10^{-8}$) and other truncating variants. This association was applicable mostly to European and Australian women. The specific variant was not detected within the SA cohort. Based on this, variants detected within this study were inspected for potential clinical impact by scrutinizing the available literature, performing *in silico* analyses and inspecting each variant's location within the gene. No final conclusions were drawn regarding pathogenicity as that still needs to be proven using segregation analysis within these families and future case-control studies. Only Class 3 to Class 5 variants will be discussed in detail.

4.3.2.1 Truncating mutations

4.3.2.1.1 *PALB2* c.424A>T, p.Lys142* (rs587782005)

PALB2 c.424A>T, p.Lys142* (rs587782005) represented a nonsense mutation that originated through a single base substitution of an adenine (A) by a thymine (T) at

Table 4.2 *PALB2* variants identified in SA population. Indicated are the exon or region where situated, the designation of the variant together with the associated protein change, the rs reference number, the minor allele frequency as detected within the Exome Aggregation Consortium (ExAC) or global databases, clinical impact as indicated by ClinVar and Ensemble, values for the various bioinformatics tools (SIFT, PolyPhen-2, Mutation Taster and Human Splicing Finder) and the final IARC classification. Class 5 mutations are indicated in red, with likely pathogenic (Class 4) indicated in pink, variants of unknown clinical significance (Class 3) in blue, likely benign (Class 2) in grey and benign polymorphisms (Class 1) in green. Novel variants are indicated in yellow. For not applicable data, the abbreviation na was used. The electropherograms utilised for the variant calling are indicated in Appendix I (raw data).

Exon	Variant ID	Clinical significance							IARC classification
		MAF Global/ExAC	ClinVar	Ensemble	SIFT Benign – 1.0 Pathogenic – 0.0	PolyPhen-2 Benign – 1.0 Pathogenic – 0.0	Mutation Taster	Human Splicing Finder	
2	c.84C>T p.Tyr28Tyr rs761533286	0.00002 Mostly African American	Benign/Likely benign 17/04/2017 5 submissions	Likely benign	na	na	Disease causing	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 2
4	c.424A>T p.Lys142Ter rs587782005	0.00001	Pathogenic 27/07/2017 3 submissions	Pathogenic	na	na	Disease causing	Na	Class 5
4	c.508A>T, p. Arg170Ter	novel	na	na	na	na	Disease causing	Na	Class 5
4	c.629C>T p.Pro210Leu rs57605939	0.02 African SNP	Benign 21/08/2017 12 submissions	Benign	1.0	0.021	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (75 - 100% variation)	Class 1
4	c.680C>A, p.Ala227Asp	novel	na	na	0.023	0.847	Polymorphism	No impact on splicing	Class 2
4	c.738T>G, p.Thr246Thr	novel	na	na	na	na	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 1
4	c.925A>G p.Ile309Val rs3809683	0.012	Benign - Likely benign 14/08/2017 9 submissions	Benign - Likely benign	1.0	0.00	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 1

Table 4.1 Continues

Exon	Variant ID	Clinical significance							IARC classification
		MAF Global/ExAC	ClinVar	Ensemble	SIFT Benign – 1.0 Pathogenic – 0.0	PolyPhen-2 Benign – 1.0 Pathogenic – 0.0	Mutation Taster	Human Splicing Finder	
4	c.1470C>T p.Pro490Pro rs45612837	0.00029	Benign/Likely benign 14/11/2017 8 submissions	Likely benign	na	na	Polymorphism	No impact on splicing	Class 2
4	c.1508C>A p.Ala503Glu	novel	na	na	0.04	0.904	Polymorphism	No impact on splicing	Class 3
4	c.1572A>G Ser524Ser rs45472400	0.002 European	Benign/Likely benign 14/11/2017 12 submissions	Benign/Likely benign	na	na	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 2
5	c.2014G>C p.Glu672Gln Rs45532440	0.01	Benign 21/08/2017 15 submissions	Benign	0.1	0.04	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 1
5	c.2393C>T p.Pro798Leu	novel	na	na	0.03	0.06	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 3
5	c.2411C>A p.Ser804Tyr	novel	na	na	0.01	0.891	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 3
7	c.2743G>A p.Ala915Thr rs374736398	0.0001	Likely benign/VUS 23/10/2017 3 submissions	Likely benign/ VUS	1.0	0.0	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 2
8	c.2794G>A p.Val932Met Rs45624036	0.001	Likely benign 15/08/2017 13 submissions	Benign	0.0	0.993	Disease causing	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 4
9	c.2968G>A p.Glu990Lys	novel	na	na	0.01	0.979	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 3
12	c.3300T>G p.Thr1100Thr rs45516100	0.0176	Benign/Likely benign 21/08/2017 9 submissions	Benign/Likely benign	na	na	Polymorphism	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 1
13	c.3434G>T p.Gly1145Val	novel	na	na	0.01	0.999	Disease causing	Alteration of an exonic ESE site Potential alteration of splicing (51 – 75% variation)	Class 4

Table 4.3 *PALB2* variants identified in SA population. Indicated are the exon number, the base and the associated amino acid change, the unique identifier or rs number for each SNP, the number of patients for whom it has been detected, their ethnicity and age at diagnosis.

Exon	Base change	AA change	rs number	Patient(s)	Ethnicity	Age
2	c.84C>T	p.Tyr28Tyr	rs761533286	2815	Black	27
4.3	c.424A>T	p.Lys142Ter	rs587782005	2338 2735	Indian Afrikaner	35 59
4.4	c.508A>T	p.Arg170Ter	novel	2411 2755	Coloured Coloured	38 45
4.4	c.629C>T	p.Pro210Leu	rs57605939	2755	Coloured	45
4.4	c.680C>A	p.Ala227Asp	novel	2411 2987	Coloured Coloured	38 45
4.5	c.738T>G	Thr246Thr	novel	702 1715 1818 2319 2335 2481 2496 2513 2742 2869 2906 2998 3077 3010 3244	White Coloured Coloured Coloured Indian Indian Afrikaner Indian White Black Coloured Afrikaner Coloured White Black	24 52 42 29 35 37 27 32 69 32 26 50 51 60 29
4.7	c.925A>G	p.Ile309Val	rs3809683	2861	Indian	28
4.8	c.1010T>C	p.Leu337Ser	rs45494092	2496	Afrikaner	27
4.10	c.1395G>A	p.Met465Ile	rs746805049	2703	Coloured	62
4.11	c.1470C>T	p.Pro490Pro	rs45612837	2960	Indian	68
4.11	c.1508C>A	p.Ala503Glu	novel	2339	Indian	38
4.12	c.1572A>G	pSer524Ser	rs45472400	702 2919	White White	24 46
5.4	c.2014G>C	p.Glu672Gln	rs45532440	2755	Coloured	45
5.7	c.2393C>T	p.Pro798Leu	novel	2820 2821	Indian Indian	67 48

Table 4.2 Continues

Exon	Base change	AA change	rs number	Patient(s)	Ethnicity	Age
5.7	c.2411C>A	p.Ser804Tyr	novel	2820	Indian	67
				2821	Indian	48
7	c.2743G>A	p.Ala915Thr	rs374736398	2826	Indian	67
				2854	White	49
				2987	Coloured	45
8	c.2794G>A	p.V932M	rs45624036	2336	Indian	46
				2580	White	30
				2854	White	49
9	c.2968G>A	p.Glu990Lys	novel	2335	Indian	35
12	c.3300T>G	p.Thr1100Thr	rs45516100	482	Coloured	32
				702	White	24
				1860	Coloured	58
				1998	Coloured	60
				2321	Coloured	56
				2513	Indian	32
				2722	White	55
				2753	Black	58
				2780	Indian	66
				2820	Indian	67
				2906	Coloured	26
				2933	Afrikaner	43
				3055	Afrikaner	66
				3244	Black	29
13.2	c.3434G>T	p.Gly1145Val	novel	2578	Coloured	42

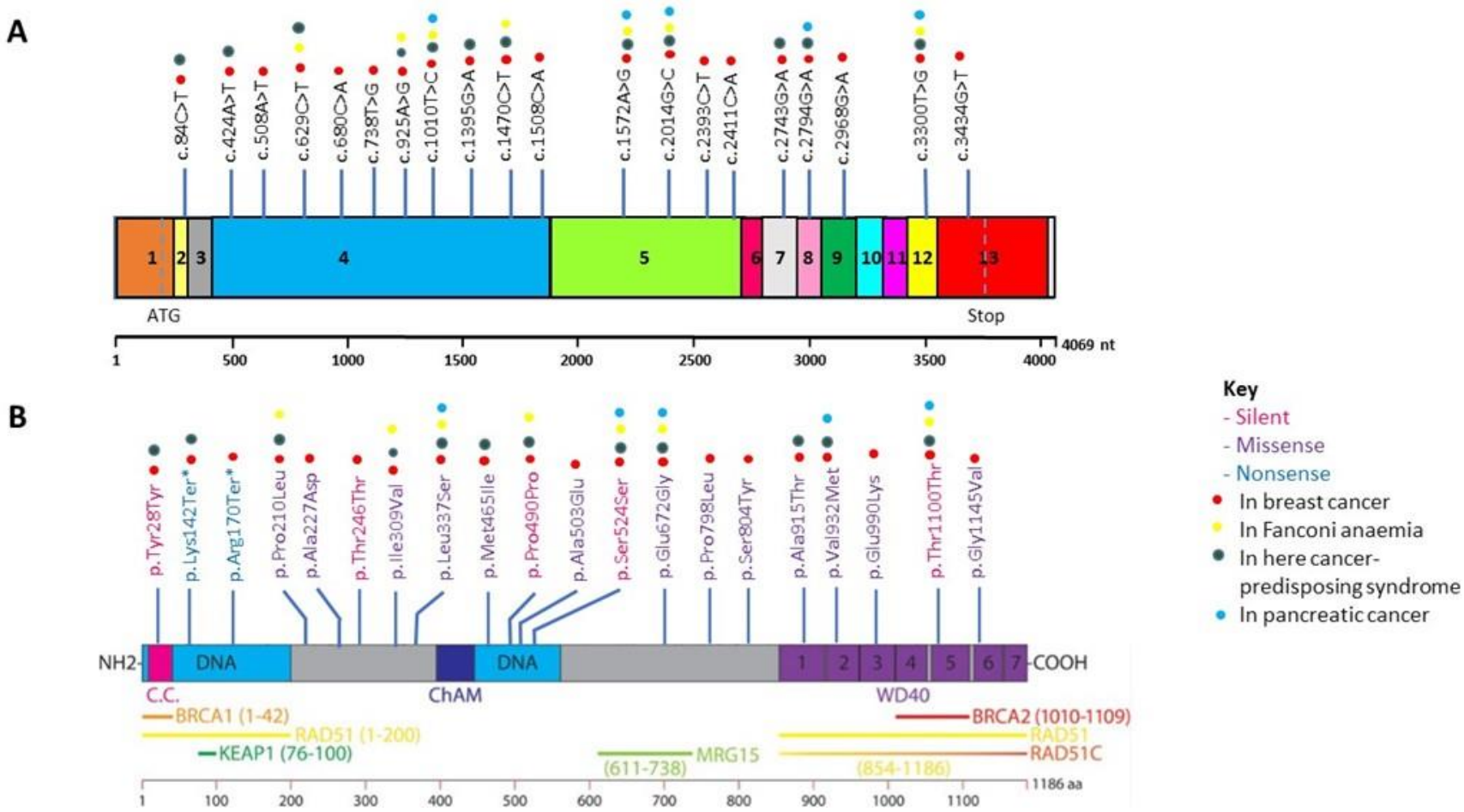


Figure 4.11 A. Schematic representation of *PALB2* with all the single nucleotide variants detected during this study superimposed on the exonic structure (A). **B**. The functional and structural domains of the protein are also indicated (adapted from Antoniou *et al.*, 2014).

DNA position 421 (chromosome position 16:23636122) in exon 4 (Fig. 4.12 A – C). The substitution resulted in an aa changing from Lysine (AAG, K) into a stop codon (TAG, X). This sequence change resulted in the creation of a premature translational stop signal at codon 142 of the protein.

This mutation was detected for two patients, namely CAM2338 and CAM2735 (Figure 4.12), and was therefore present in 2.3% (2/86) of the cohort. CAM2338 represented an Indian woman diagnosed with BC at the age of 35 years (born 1974). She developed a second primary BC at the age of 38. The second tumour tested positive for the expression of estrogen- (ER) and progesterone receptors (PR), but lacked the expression of HER2/NEU receptors (HER2). The patient had three paternal aunts affected with BC or OVC respectively (Figure 4.13). The aunts affected with OVC were diagnosed in their 40's. As the affected family members formed part of the previous generation, the index patient was the only remaining ca affected at the time of blood sampling, therefore segregation analysis could not be performed.

The second mutation carrier (CAM2735) was an Afrikaner patient. This patient was diagnosed with OVC at the age of 59 (born 1953). This patient had a sister (dx 46) and a mother affected with BC (dx 62) (Figure 4.14). None of the patient's three children have been diagnosed with any form of ca at the time of testing. Unfortunately both the patient and the affected sister have passed away since genetic testing commenced, thereby preventing segregation analysis within the family.

This mutation is rare as ClinVar only indicates three submissions, all by well-known international clinical pathology laboratories (Ambry Genetics, GeneDx and Invitae), between the years 2016 – 2017. The mutation has been evaluated by the expert panel and classified as pathogenic (Class 5) on July 27, 2017 (accessed on 16 November 2017). However, the mutation has not been described previously in the literature.

The presence of the terminating mutation will result in a shortened PALB2 protein compared to the wild type. This will be due to either being prematurely truncated or through nonsense-mediated mRNA decay (Hug *et al.*, 2016). This piece of protein will be lacking the important C-terminal domain (with WD-40 repeat-like segments, Figure 2.2), which facilitates various protein–protein interactions including the BRCA2/PALB2 and the RAD51/PALB2 complexes (Park *et al.,* 2014). The absence of this interaction has been associated with genomic instability (Hayakawa *et al*., 2010; Reddy *et al*., 2016) and is postulated to carry a 2 – 4 fold

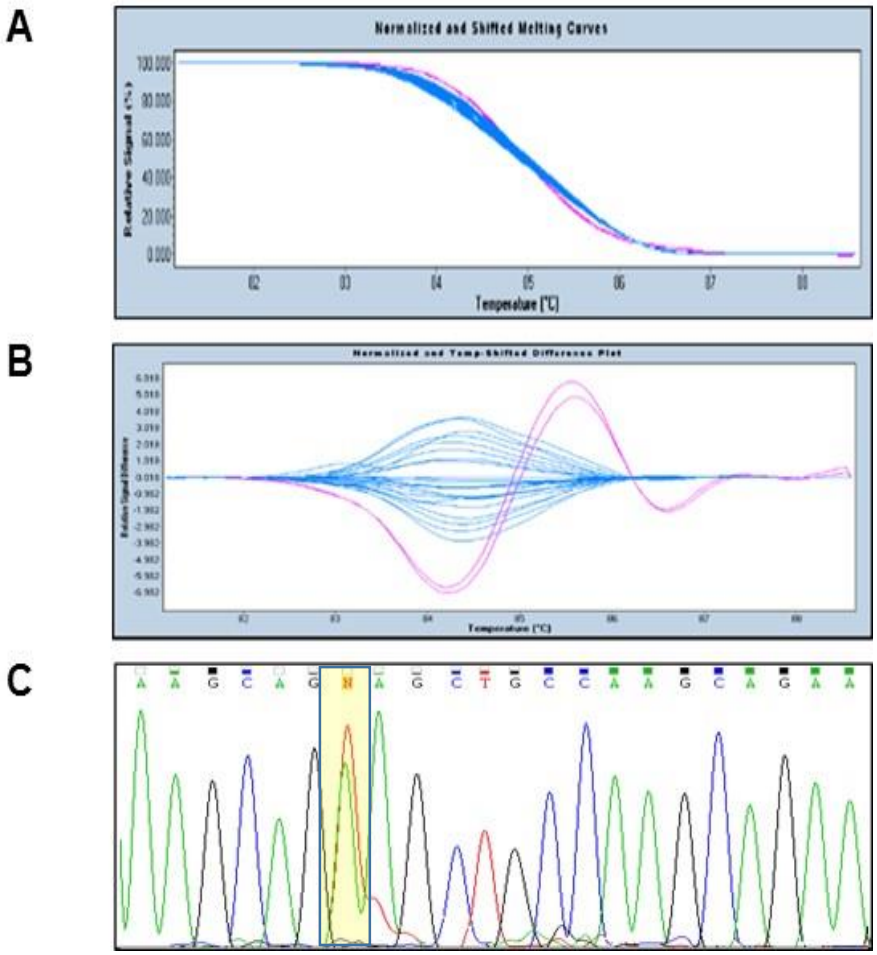


Figure 4.12 HRMA results for *PALB2* exon 4 for 23 familial BC patients, each performed in duplicate, of which a single patient (CAM 2338) carried the *PALB2* c.424A>T,p.Lys142Ter (rs587782005) mutation. **A**. The normalized and temperature shifted melt curves based on the qPCR HRMA data. **B.** The difference plot generated from the melt, with the variant (pink curves) grouping separate from the bulk of the samples (representing the wild types) on the baseline, with a relative signal difference of ~ 6.0 units. **C.** Sequencing electropherogram representing the *PALB2* c.424A>T, p.Lys142Ter mutation, highlighted by the yellow box.

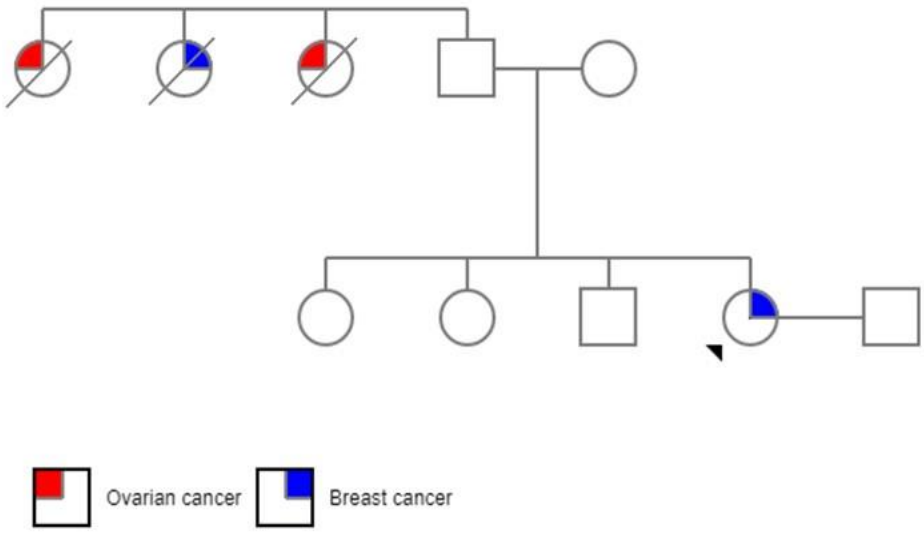


Figure 4.13 Family pedigree for CAM2338, indicating a positive paternal family history of BC and OVC in this Indian family. The index patient (who tested negative for disease-causing mutations in the *BRCA* genes) is indicated by the arrow head. This BC patient is affected with bilateral BC (dx 35 and 38) and carries the *PALB2* c.424A>T, p.Lys142Ter (rs587782005) mutation. The three paternal aunts were affected with either BC or OVC, but have since passed away. The symbol descriptions are indicated.

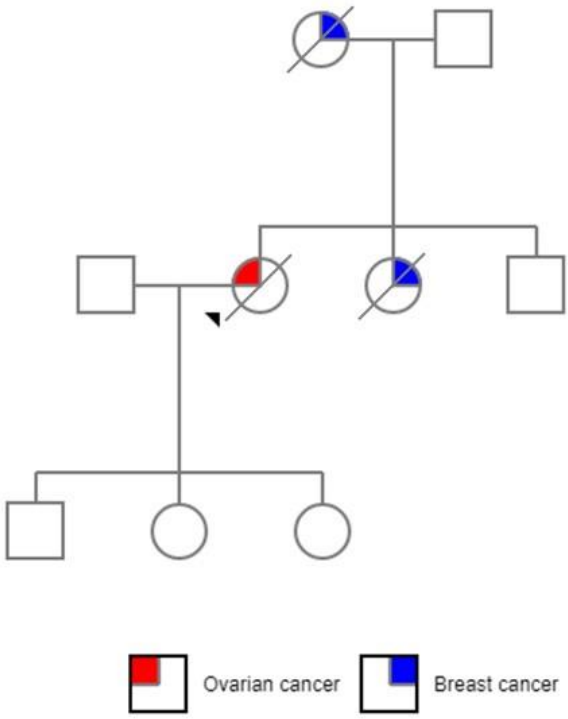


Figure 4.14 Family pedigree for CAM2735, indicating a positive maternal family history of BC and OVC in this Afrikaner family. The index patient (who tested negative for disease-causing mutations in the *BRCA* genes) is indicated by the arrow head. This BC patient is affected with OVC (dx 59) and carries the pathogenic *PALB2* c.424A>T, p.Lys142Ter (rs587782005) mutation. The mother and sister were diagnosed with BC at the ages of 62 and 46 respectively, and have since passed away. The symbol descriptions are indicated.

increase in risk to develop familial BC depending on the family history (SNPedia – www.snpedia.com/index.php/Rs587782005).

It is interesting to note that the two SA patients were affected with different types of ca, namely CAM2338 with bilateral BC and CAM2735 with OVC, although both pedigrees exhibit a family history of both ca types (Figures 4.13 & 4.14). Although the risk of BC for female *PALB2* mutation carriers with a family history of the disease was 8 – 9 times as high among those younger than 40 years of age when compared with the general population (such as CAM2338 diagnosed at ages 36 and 38), the risk did not distinguish between unilateral and bilateral disease (Antoniou *et al.*, 2014).

The only study to seek an association between deleterious *PALB2* mutations and bilateral disease, was Metcalfe and co-workers in Canada (Metcalfe *et al.*, 2017). They determined that the frequency of deleterious *PALB2* mutations were higher in women with bilateral BC compared to unilateral cases (2.4% compared to 0.6%, p = 0.01). Their study included 2 225 women with BC and a total of 429 cases with OVC. Their results however, did not indicate an association between *PALB2* mutations and OVC, as none of the 429 cases carried a deleterious *PALB2* mutation. Their observation was also substantiated by Ramus *et al.* (2015) for OVC patients from the United States of America.

The lower percentage of OVC cases carrying deleterious mutations within *PALB2* is supported in the literature by Kotsopoulos *et al.* (2016a). They evaluated the prevalence of *PALB2* germline mutations in 1421 epithelial OVC patients and also screened 4300 European controls. They reported a mere three OVC patients (0.21%) who carried a deleterious mutation, compared to 0.05% in the controls. All three the women passed away within two years of diagnosis. The average age at onset was relatively late, namely 59 years (range of 55 – 62), similar to the age of diagnosis for CAM2735. Although the incidence of deleterious mutations were lower in OVC cases, they still reported an association with a four-fold increased risk for the disease (OR = 4.55; 95% CI 0.76 – 27.24). The association however, was not significant (p = 0.10), possibly due to the small size of the cohort (Kotsopoulos *et al.*, 2016b). In a study by Antoniou *et al.* (2014), a relative risk of 2.31 (95% CI 0.77 – 6.97) for OVC was reported, but only a small number of patients were studied.

Larger studies of OVC patients unselected for a positive family history are needed to refine the prevalence and penetrance of deleterious *PALB2* mutations in these cases. To determine whether the *PALB2* c.424A>T, p.Lys142Ter (rs587782005) mutation carries similar risks to BC and OVC as described, the prevalence of this mutation needs to be thoroughly evaluated within the SA population to confirm its impact.

4.3.2.1.2 *PALB2* c.508A>T, p.Arg170Ter* (novel)

This mutation found in exon 4 represented a single base substitution of an A by a T at position 508 (Figure 4.15 A – C). The nucleotide change resulted in an aa change from Arginine (AGA, A) at codon 170 to a premature termination codon (TGA, X).

This nonsense mutation was detected for two BC patients, namely CAM2411 and CAM2755. This mutation was therefore detected for 2.3% (2/86) of the cohort. CAM2411 represented a Coloured woman diagnosed with BC at the age of 39 years (born 1972). Details regarding her tumour were not indicated. The index' mother was affected with BC (dx unknown), with two siblings affected with liver and stomach ca respectively (Figure 4.16). The second patient (born 1968) also self-identified as Coloured was diagnosed with BC at the age of 45. This patient had a moderate family history of BC, as two paternal aunts were diagnosed with the disease during their 40's, with a third paternal aunt diagnosed with colon ca at a later stage in life (dx 70's) (Figure 4.17).

This mutation is novel as it has not been detected before, therefore the variant has no identification number. The closest variant detected in this region is the *PALB2* c.509_510delGA mutation detected in populations within Central and Eastern Europe (Noskowicz *et al.*, 2014). This mutation represents a recurrent mutation in especially Germany and Poland (Casadei *et al.*, 2011; Dansonka-Mieszkowska *et al.*, 2010; Bogdanova *et al.*, 2011). The investigators determined that this truncating mutation is prevalent in about 1 in 400 BC patients from Belarus, Russia and Germany. The mutation was furthermore associated with ER+ disease (Noskowicz *et al.*, 2014).

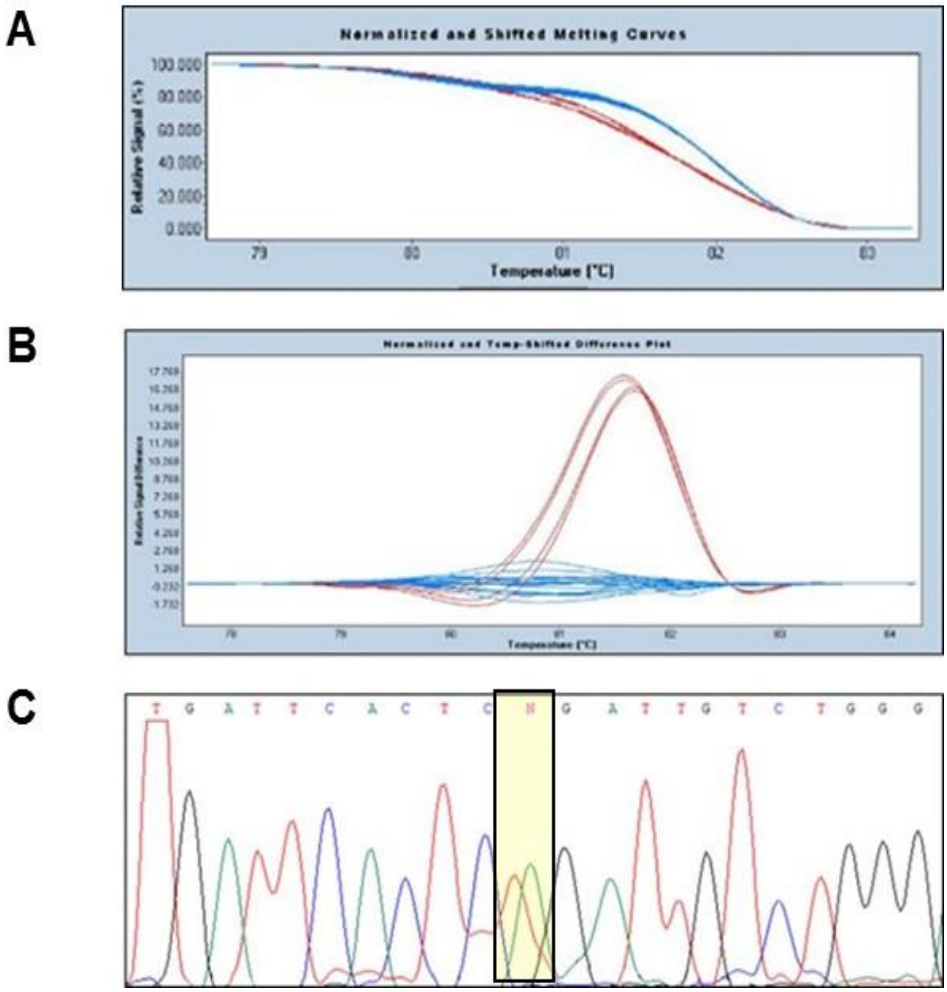


Figure 4.15 HRMA results for *PALB2* exon 4 for 23 familial BC patients, each performed in duplicate, of which two patients (CAM 2411 and CAM2755) carried the *PALB2* c.508A>T, p.Arg170Ter* novel mutation. **A**. The normalized and temperature shifted melt curves based on the qPCR HRMA data. **B.** The difference plot generated from the melt, with the variant (red curves) grouping separate from the bulk of the samples (representing the wild types) on the baseline, with a relative signal difference of ~ 18.0 units. **C.** Sequencing electropherogram representing the *PALB2* c.508A>T, p.Arg170Ter* mutation, highlighted by the yellow box.

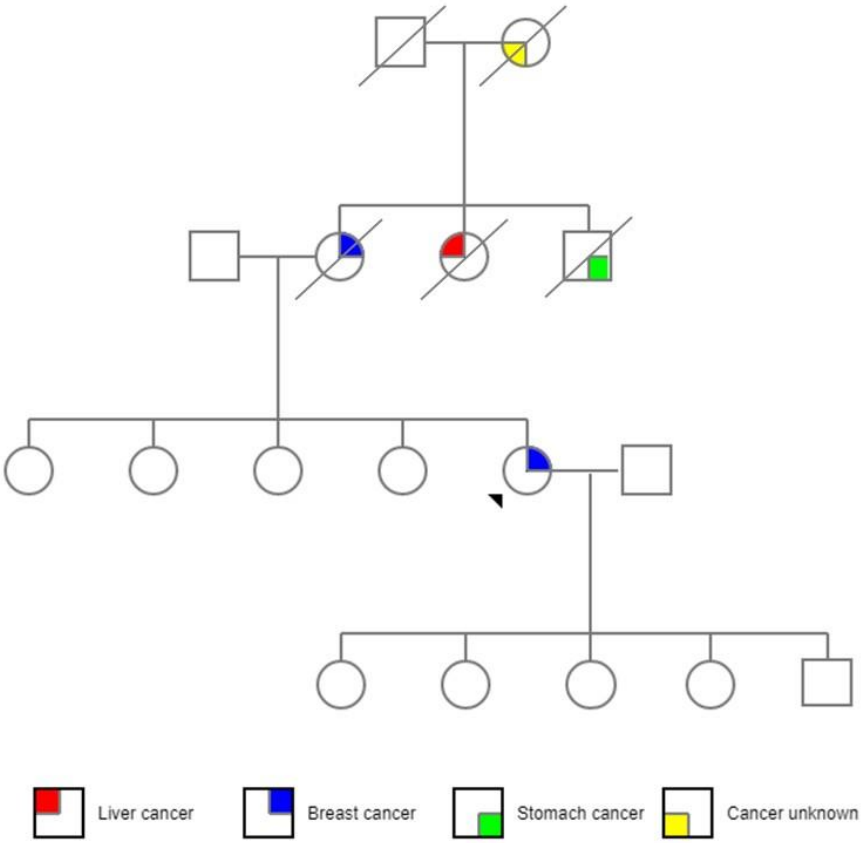


Figure 4.16 Family pedigree for CAM2411, indicating a positive maternal family history of breast- and other ca types in this Coloured family. The index patient (who tested negative for disease-causing mutations in the *BRCA* genes) is indicated by the arrow head. This patient is affected with bilateral BC (dx 38 & 43) and carries the truncating *PALB2* c.508A>T, p.Arg170Ter* novel mutation. The mother, maternal aunt and uncle were diagnosed with breast, liver and stomach ca respectively. All three family members have passed away. The symbol descriptions are indicated.

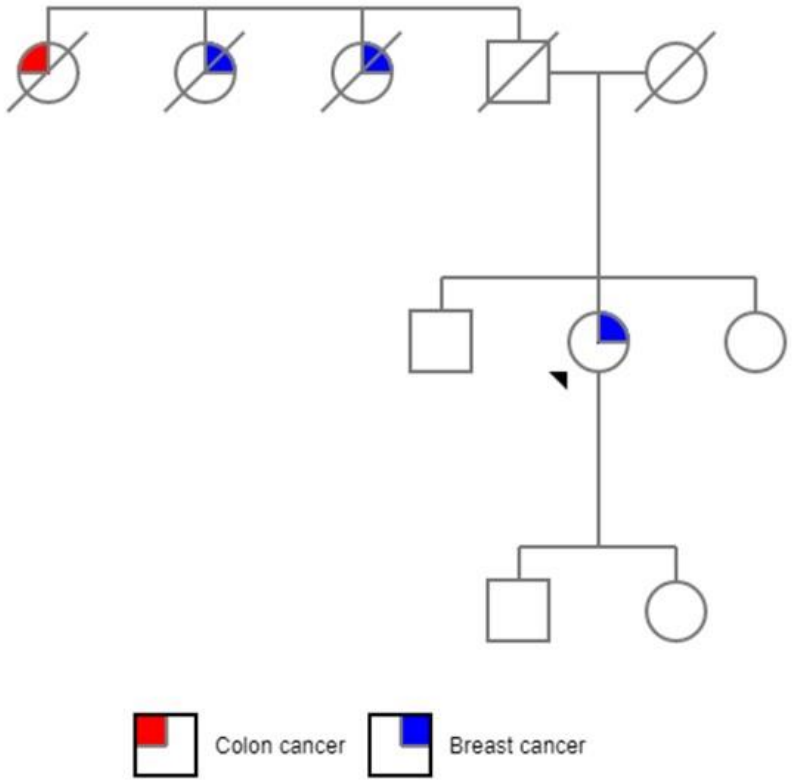


Figure 4.17 Family pedigree for CAM2755, indicating a positive paternal family history of mostly BC. A single female was also diagnosed with colon ca. The index patient (who tested negative for disease-causing mutations in the *BRCA* genes) is indicated by the arrow head. This patient is affected with unilateral BC (dx 38) and carries the truncating *PALB2* c.508A>T, p.Arg170Ter* novel mutation. Two paternal aunts were also diagnosed with BC, with a third aunt affected with colon ca. All three these family members have passed away. The symbol descriptions are indicated.

The SA novel *PALB2* c.508A>T, p.Arg170Ter* is situated within the amino-terminal region (aa 1 – 394) of the protein that comprises the functional binding domains for KEAP1, BRCA1 and RAD51 (Fig. 2.2 A). As the function of this section of the protein is to mediate these protein – protein interactions to ensure proper localisation of BRCA2 (Zhang *et al.,* 2009b; Park *et al.*, 2014), a premature truncation of the protein will be detrimental to DNA repair. Based on this evidence, this mutation is classified as a Class 4 to Class 5 mutation. Although both the mutation carriers were affected with BC at an early age (dx 38 and 45), unfortunately no clinical data regarding their tumour status were known.

4.3.2.2 Missense mutations

4.3.2.2.1 *PALB2* c.680C>A, p.Ala227Asp (novel)

A single base change [Cytosine (C) into A] occurred at 680, within exon 4 (Figure 4.18). The variant resulted in an aa change from Alanine (GCC, A) to Aspartic acid (GAC, D) at codon 227. The variant is novel, as it has not been recorded before, therefore it has no identification number (Table 4.1). The variant was detected for two patients (2.3%), namely CAM2411 and CAM2987. These Coloured women were diagnosed with BC at the ages of 38 and 45 years respectively (Figure 4.18). Both patients had a positive family history of breast- and various other ca types. The pedigree for CAM2411 is indicated in Figure 4.16. This patient also carried the *PALB2* c.508A>T, p.Arg170Ter* mutation.

In silico prediction was contradictory, as SIFT and PolyPhen-2 concurred, with values of 0.023 and 0.847 respectively, indicating a damaging effect (Figure 4.19 A – C). In contrast, Mutation Taster indicated the variant to represent benign variation (Table 4.2). The splicing predictor (Human Splicing Finder) also did not indicate any impact on splicing (Table 4.2).

The prediction results for PolyPhen-2 are illustrated in Figure 4.19. The graph for HumDiv is illustrated in Figure 4.19 A, whereas that of HumVar is presented in Figure 4.19 B. The existence of the two datasets came about when it was used to train and test the PolyPhen-2 prediction models. HumDiv was compiled from all damaging alleles with known effects on the molecular function causing human Men-

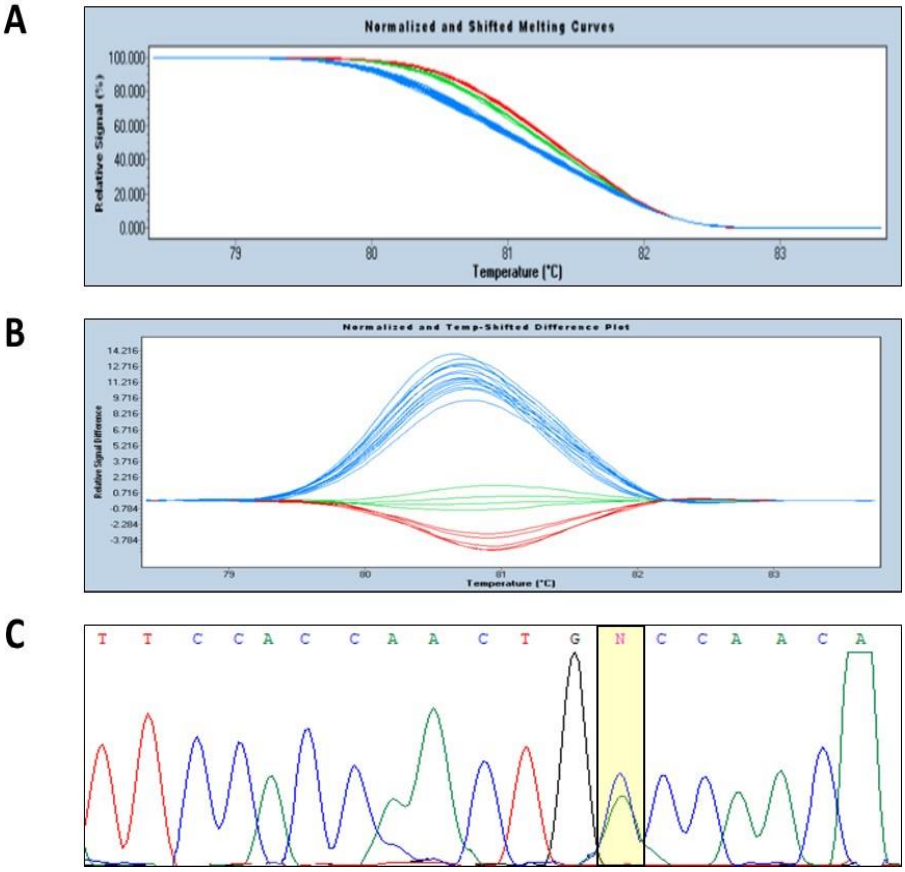


Figure 4.18 HRMA results for *PALB2* exon 4 for 23 familial BC patients, each performed in duplicate, of which two patients (CAM 2411 and CAM2987) carried the *PALB2* c.680C>A, p.Ala227Asp novel mutation. **A**. The normalized and temperature shifted melt curves based on the qPCR HRMA data. **B.** The difference plot generated from the melt, with the variant (red curves) grouping separate from the bulk of the samples [representing either the wild type (green curves) or a common benign polymorphism (*PALB2* c.738T>G, p.Thr246Thr – blue curves)], with a relative signal difference of ~ 4.0 units. **C.** Sequencing electropherogram representing the *PALB2* c.680C>A, p.Ala227Asp novel mutation, highlighted by the yellow box.

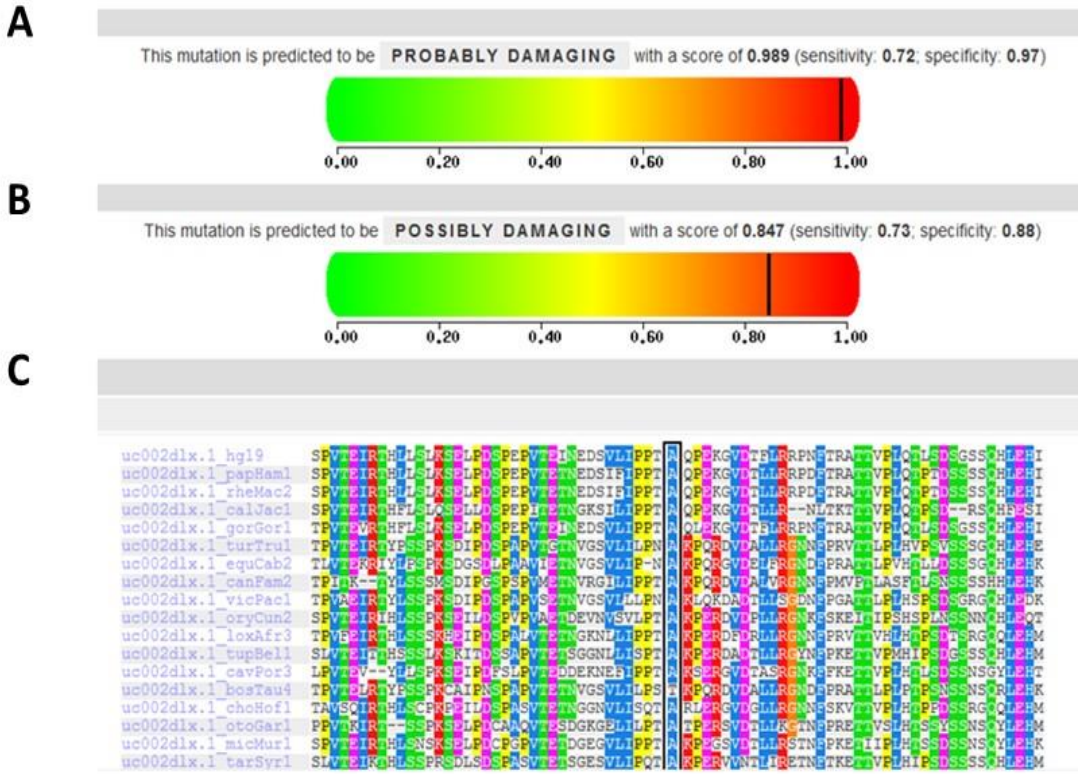


Figure 4.19 *In silico* results for the *PALB2* c.680C>A, p.Ala227Asp novel mutation using PolyPhen-2. **A.** Prediction results according to HumDiv. **B**. Prediction graph results according to HumVar. **C**. Multiple alignment across species of a section of 75 aa surrounding the position of the variant (http://genetics.bwh.harvard.edu/cgi-bin/ggi/ggi2.cgi, accessed on 27 December, 2017).

delian disease. These alleles were combined with differences between the human proteins and their closely related mammalian homologs, assumed to be non-damaging, illustrated in the multiple alignment (Figure 4.19 C). The second set of data (HumVar-trained PolyPhen-2 model) consisted of all human disease-causing mutations extracted from UniProtKB, together with common human SNPs that have a MAF > 1%, without annotated involvement in disease (non-damaging). Although both models can be used, HumVar is recommended by the website for the diagnostics of Mendelian diseases as it required distinguishing mutations with drastic effects from all the remaining human variation. For this reason, the PolyPhen-2 value indicated by HumVar was used for the purposes of this study, which was 0.847 (Figure 4.19 B, Table 4.2).

The classification of this variant was complicated due to the conflicting data obtained from the *in silico* prediction. From the multiple alignment, it is clear that this region within the protein is relatively conserved, for the surrounding aa sequence (four aa before and eight aa after) did not vary for the first five species listed (Figure 4.19 C). As the location of the variant at aa 227 fell just outside of the binding area of RAD51 (Figures 2.2 & 2.4) within the N-terminal portion of the protein (aa 1 – 200) and the fact that the variant was detected for CAM2411 (who also carried the truncating *PALB2* c.508A>T, p.Arg170Ter* novel mutation, Figure 4.15), a classification of likely benign (Class 2) was awarded. Bi-allelic pathogenic mutations would have resulted in FA.

4.3.2.2.2 *PALB2* c.1395G>A, p.Met465Ile (rs746805049)

A single base change [Guanine (G) into A] occurred at 1395, within exon 4 (Table 4.2, electropherogram presented in Appendix I). The variant resulted in an aa change from Methionine (ATG, M) to Isoleucine (ATA, I) at codon 465. The variant (rs746805049) has been detected before, although it has a very low MAF (0.0001 – Table 4.2). ClinVar indicates the variant as a VUS, as it has been submitted only once before (14/11/2017). The variant was detected for a single patient (1.2%) in this study, namely CAM2703 (Table. 4.2). This patient represented a Coloured woman diagnosed with BC at the age of 62 years. *In silico* prediction was contradictory, as SIFT indicated a value of 0.18 (ranging towards potentially deleterious), PolyPhen-2 (value of 0.011 – benign), with Mutation Taster also

indicating it to be polymorphic (Table 4.2). The variant was not situated in a critical functional domain of the protein (Fig. 2.4), although Human Splicing Finder indicated a possibility for the alteration of an exonic splicing enhancer (ESE) site (Table 4.2). As the variant was detected in only 1.2% of the cohort in a heterozygous state, the classification of ClinVar as a VUS (Class 3) was accepted.

4.3.2.2.3 *PALB2* c.1508C>A, p.Ala503Glu (novel)

This single base change C>A occurred at 1508, within exon 4 (Table 4.2, electropherogram presented in Appendix I). The variant resulted in an aa change from Alanine (GCA) to Glutamic acid (GAA, E) at codon 503. The variant is novel (Table 4.2) and was detected for a single patient (1.2%), namely CAM2339 (Table. 4.2). This patient represented an Indian woman diagnosed with BC at the age of 38 years, who had a moderate family history of breast- and stomach ca (Table 3.1).

In silico prediction was contradictory (Table 4.2). SIFT and PolyPhen-2 indicated values of 0.04 (tending towards potentially deleterious) and 0.904 respectively due to the change in the aa properties (from A to E) and the fact that it is located in a relatively conserved area of the gene across species. Alanine represents a neutral hydrophobic aa, whereas E is acidic and charged. Mutation Taster predicted the variant to be polymorphic (Table 4.2). Due to the contraction in prediction by the various programmes and the fact that the variant is novel, a classification of VUS (Class 3) is proposed according to the ENIGMA guidelines presented in Appendix H, until further evidence becomes available.

4.3.2.2.4 *PALB2* c.2393C>T, p.Pro798Leu (novel)

This SNP (C>T at c.2393) was located in exon 5 (Table 4.2, electropherogram presented in Appendix I). The presence of the variant resulted in an aa change from Proline (CCT, P) to Leucine (CTT, L) at codon 798. This SNP represented the fifth novel variant detected in the cohort (Table 4.2) and was present in two patients (2.3%), namely CAM2820 and CAM2821 (Table. 4.2). Both these patients were self-declared as Indian and were affected with BC at the ages of 67 and 48 respectively (Table 4.2). These patients had an extensive family history of mainly BC (Table 3.1).

For this variant, SIFT indicated a value of 0.03 (possibly damaging) due to the change in the aa properties and structure from P to L. The 5-membered aa residue was transformed to an acyclic aa, which would have resulted in conformational changes in the immediate vicinity of the substituted site. This could lead to a significant alteration in the thermodynamic stability of the mutant.

Such a change was investigated by Kundu *et al.* (2012). They studied the influence of a non-synonymous change similar to *PALB2* c.2393C>T, p.Pro798Leu for neuroblastoma. The SNP investigated was rs1052574 present in the Caspase-9 gene (*CASP9*), where the L at position 197 in the native protein was replaced by a P residue. The investigators used computational tools which predicted the effect of this SNP. They stated that the polymorphic variant resulted in an aa substitution from L to P, therefore the residue changed from an acyclic aa to a 5-membered aa (just the opposite to *PALB2* c.2393C>T, p.Pro798Leu) which resided in the buried area of the protein with a high level of conservation. With their bio-informatic analyses, they concluded that the aa substitution from L to P showed a transition from helix to coil in the mutant protein, which altered the properties of the aa. This alteration resulted in the reduction of the stability of the protein, thereby affecting the function of the CASP9 protein. The authors concluded that the presence of this variant led to the deregulation of apoptosis and neuroblastoma development (Kundu *et al.*, 2012).

The prediction by PolyPhen-2 however, was 0.102 (Table 4.2). This value was based on the moderate degree of conservation across species at the specific region (Figure 4.20), as this region was not situated in a critical functional domain of the protein (Figure 2.2). Mutation Taster also predicted the variant to be polymorphic, with Human Splicing Finder indicating a small possibility for the alteration of an ESE site (Table 4.2). This variant was therefore preliminarily classified as a VUS (Class 3).

4.3.2.2.5 *PALB2* c.2411C>A, p.Ser804Tyr (novel)

This missense mutation (C>A) in exon 5 at c.2411 resulted in an aa change from Serine (TCT, S) to Tyrosine (TAT, Y) at codon 804 (Fig. 4.21 A – C). Although the

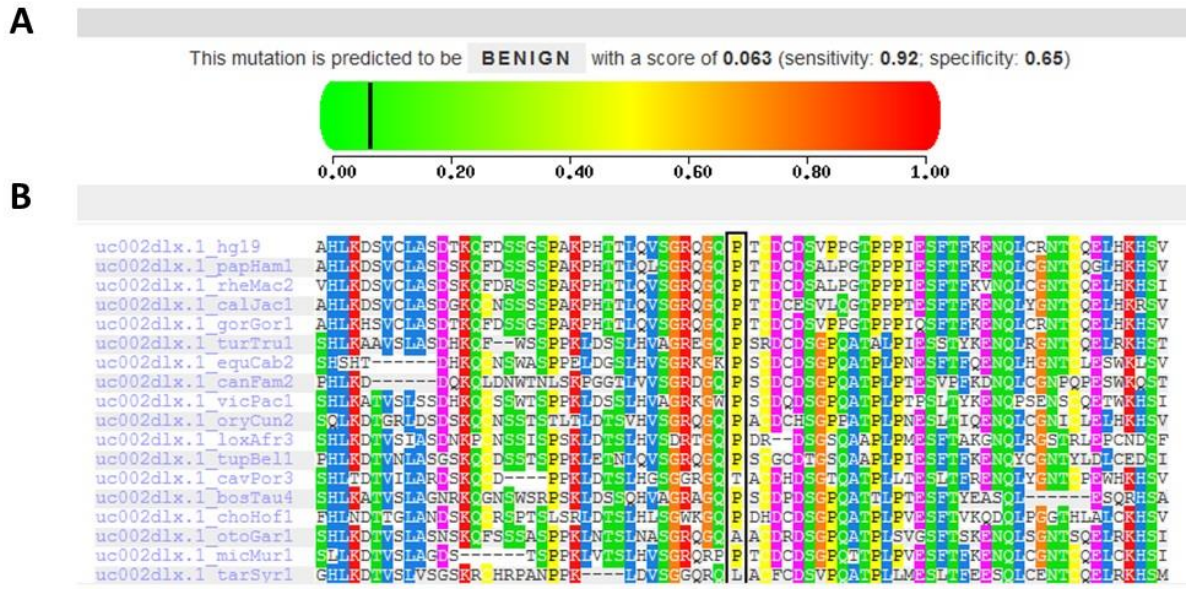


Figure 4.20 PolyPhen-2 prediction results for the *PALB2* c.2393C>T, p.Pro798Leu novel mutation. **A.** Prediction results according to HumVar. **B**. Multiple alignment across species of a section of 75 aa surrounding the position of the variant (http://genetics.bwh.harvard.edu/cgi-bin/ggi/ggi2.cgi, accessed on 27 December, 2017).

specific change had not been recorded before (therefore novel), two alternatives involving the same base have been reported to ClinVar, namely *PALB2* c.2411C>G, p.Ser804Cys and *PALB2* c.2411C>T, p.Ser804Phe (both rs149836639 – 16:23641064 on CHCh37). Both these variants have only been recorded once before, and therefore has an extremely low MAF. Due to the low MAF, both these variants were classified as a VUS Class 3 variant by ClinVar and Ensemble.

The mutation (*PALB2* c.2411C>G, p.Ser804Tyr) was detected for two Indian BC patients (CAM2820 & CAM2821), diagnosed at ages 67 and 48 respectively (Table 4.2). These patients had an extensive family history of mainly BC, although various ca types were also recorded for each of the families. The pedigrees for these patients are presented in Figure 4.22 A & B. This mutation seems to co-occur with the other novel variant in exon 5, namely *PALB2* c.2393C>T, p.Pro798Leu. as CAM2820 and CAM2821 were the only BC patients to exhibit both variants (Table 4.2). This potential co-occurrence needs to be investigated in future studies to determine whether the two variants are co-occurring by chance or consistently, and whether the two SNPs are in *cis* or *trans* on the chromosome.

In silico prediction using the various tools concurred, as both SIFT and PolyPhen-2 indicated potential damage (values of 0.01 and 0.891 respectively) due to the change in the aa properties and structure (S to Y) in a relatively conserved region of the protein (Figure 4.23). The missense mutation occurred in the C-terminus of the protein (aa 791 – 1186), although it was not located in the critical WD40 region (binding region of five different proteins, Figure 2.2). On the contrary, Mutation Taster predicted the variant to be polymorphic, together with Human Splicing Finder indicating a small possibility for the alteration of an ESE site (Table 4.2). Based on the *in silico* prediction and the fact that all possible alternatives have been recorded before for this base (although only once), the variant was preliminarily classified as a VUS (Class 3) until further evidence can be obtained.

4.3.2.2.6 *PALB2* c.2794G>A, p.Val932Met (rs34624036)

This mutation represented a single base substitution (G>A) at nucleotide position 2794 within exon 8 (Fig. 4.24 A – C). The presence of this variant resulted in an aa change from Valine (GTG, V) into M. The resulting aa at codon 932 of the protein

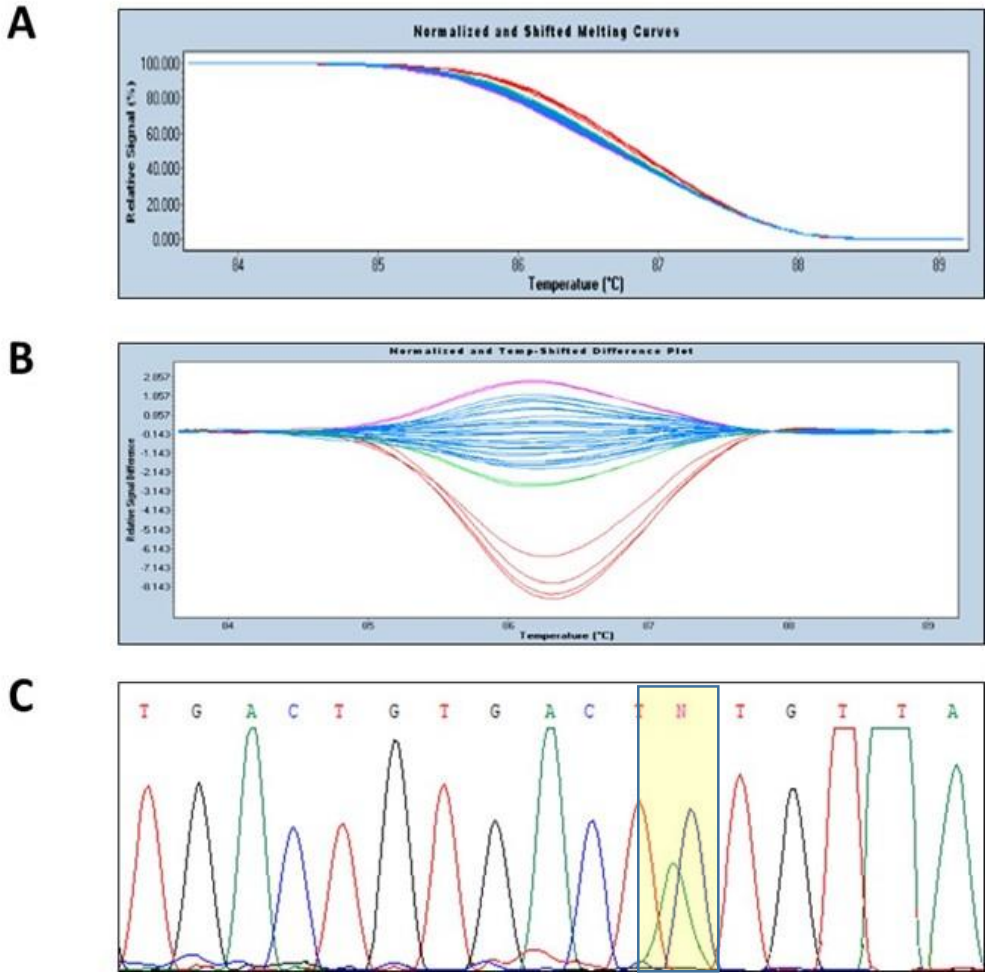


Figure 4.21 HRMA results for *PALB2* exon 5 for 23 familial BC patients, each performed in duplicate, of which two patients (CAM 2820 and CAM2821) carried the *PALB2* c.2411C>A, p.Ser804Tyr novel mutation. **A**. The normalized and temperature shifted melt curves based on the qPCR HRMA data. **B.** The difference plot generated from the melt, with the variant (red curves) grouping separate from the bulk of the samples representing the wild type (pink, green and blue curves), with a relative signal difference of ~ 8.0 units. **C.** Sequencing electropherogram representing the *PALB2* c.2411C>A, p.Ser804Tyr novel mutation, highlighted by the yellow box.

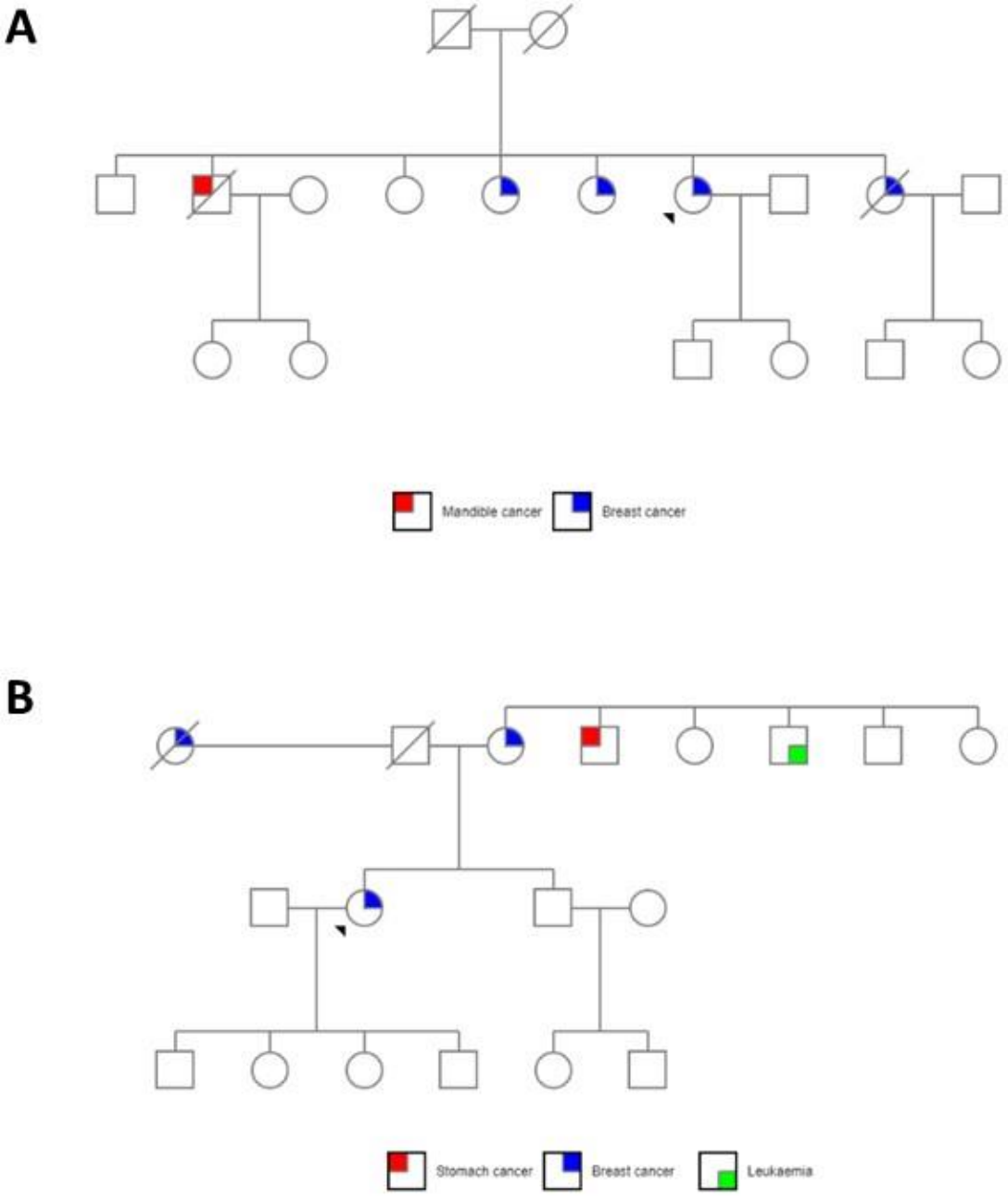


Figure 4.22 Family pedigrees for two familial BC patients identified with the *PALB2* c.2411C>A, p.Ser804Tyr novel mutation. **A**. Pedigree for CAM2820, exhibiting three sisters affected with BC and a brother with ca of the mandible. **B**. Pedigree for CAM2821, indicating a positive family history of breast-, stomach ca and leukaemia. The index patient is indicated by the arrow head. The symbol descriptions are indicated.

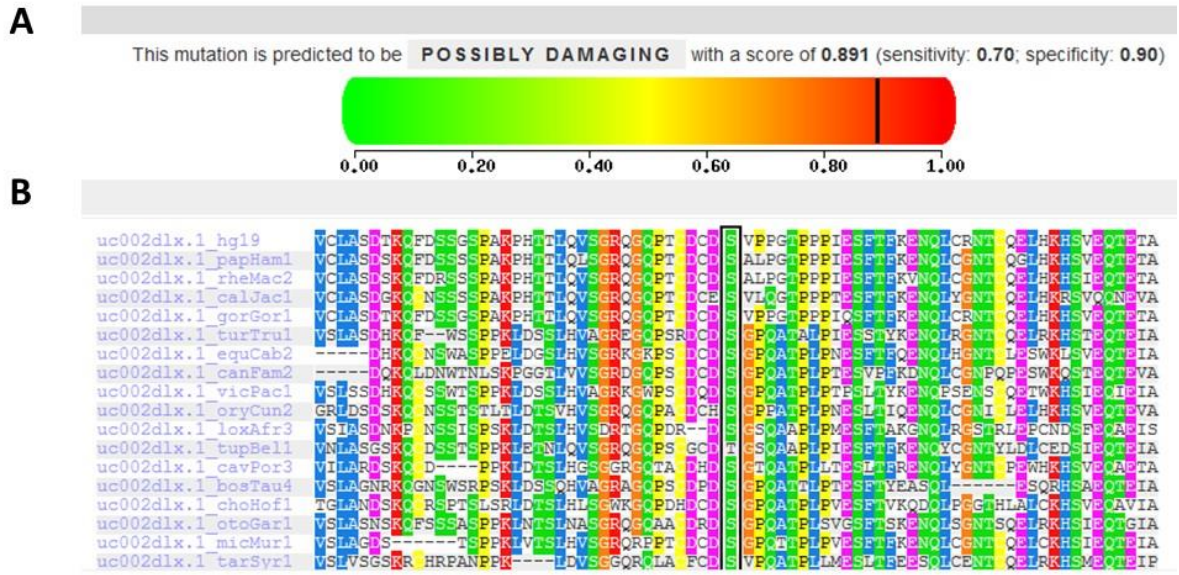


Figure 4.23 PolyPhen-2 prediction results for the *PALB2* c.2411C>A, p. Ser804Tyr novel mutation. **A.** Prediction results according to HumVar. **B**. Multiple alignment across species of a section of 75 aa surrounding the position of the variant (http://genetics.bwh.harvard.edu/cgi-bin/ggi/ggi2.cgi, accessed on 27 December, 2017).

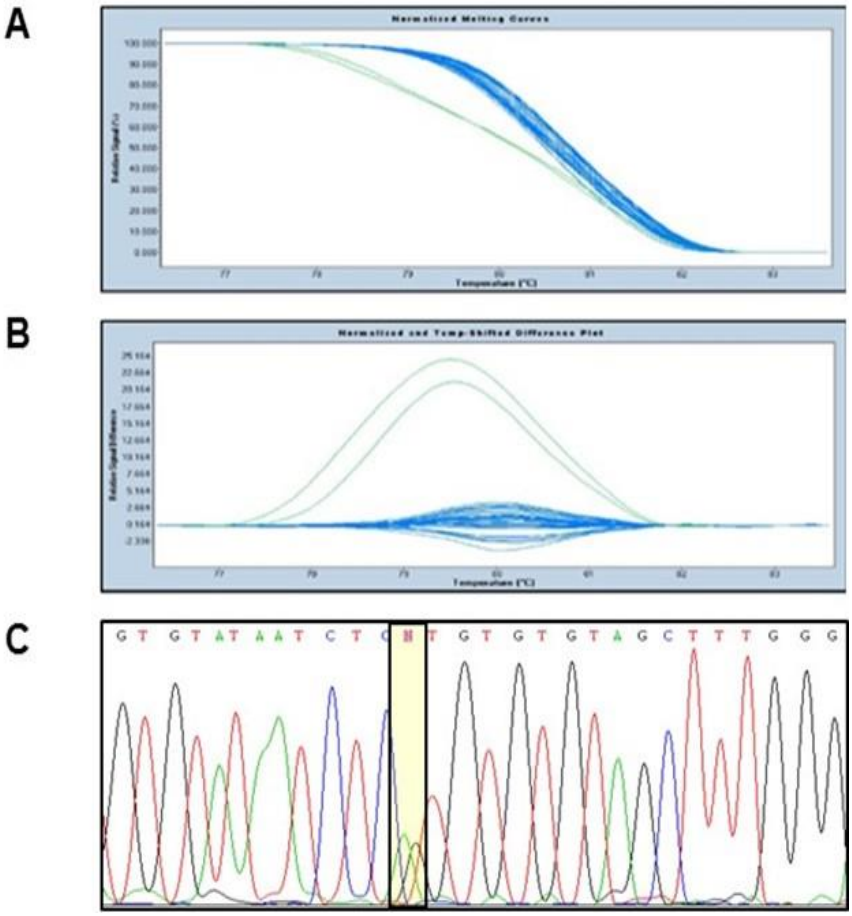


Figure 4.24 HRMA results for *PALB2* exon 8 for 23 familial BC patients, each performed in duplicate, of which a single patient (CAM2854) carried the *PALB2* c.2794G>A, p.Val932Met (rs34624036) mutation. **A**. The normalized and temperature shifted melt curves based on the qPCR HRMA data. **B.** The difference plot generated from the melt, with the variant (green curves) grouping separate from the bulk of the samples (representing the wild types) on the baseline, with a relative signal difference of ~ 25.0 units. **C.** Sequencing electropherogram representing the *PALB2* c. c.2794G>A, p.Val932Met mutation, highlighted by the yellow box.

had similar physio-chemical properties to that of the wild-type, for they are both medium sized aa and hydrophobic.

This variant was detected in three familial BC patients, namely CAM2336, CAM2580 and CAM2854 (Figures 4.25 A – B & 4.26, Table 4.2). It was present in 3.5% (3/86) of the cohort. CAM2336 (born in 1966) represented an Indian woman diagnosed with BC at the age of 46 years (Figure 4.25 A). The tumour that was removed tested ER+, PR+ and HER2-. Her sister was affected with bilateral disease, diagnosed at the ages of 35 and 45. She died at a young age, soon after developing her second BC. The older brother was diagnosed with prostate ca (unconfirmed at the time), as he passed away in a car accident before commencing treatment at the age of 50.

CAM2580 presented with bilateral BC at the extremely young age of 30. At the time of initial genetic testing for mutations in *BRCA1* and *BRCA2*, this patient was already 63 years of age (born in 1948) (Figure 4.25 B). The patient's sister also presented with BC at a young age (dx 38) and was also still alive at the time of testing. She herself had not been screened for mutations within *BRCA1* and *BRCA2* at the time. Both these siblings have therefore been surviving their ca diagnosis for more than 30 years. No ca has previously been recorded for either of their parents (Figure 4.25 B).

The third mutation carrier (CAM2854) represented the White (non-Afrikaner) population and presented with BC at the age of 49 (born 1965) (Figure 4.26). Her sister was also diagnosed with the disease and passed away during the year 2014 (age at diagnosis not indicated). The maternal family history entailed an aunt and grandmother affected with OVC, age at diagnosis not indicated (Figure 4.26).

This mutation is situated within the highly conserved WD-40 repeat-like segments located in the C-terminal domain (Figures 2.2 & 4.27 A – B). The results of the bio-informatic predictions are presented in Table 4.2. The presence of the variant was predicted to be deleterious by both SIFT (value of 0), Mutation Taster and indicated as probably damaging by PolyPhen2 (value of 0.993).

This *in silico* analysis corresponded to that described for this mutation by Wong-Brown *et al.* (2014) and Thompson *et al.* (2015). Thompson *et al*. (2015) also indicated bioinformatic results for Condel (https://omictools.com/) to be deleterious. Splicing analysis using Human Splicing Finder indicated a possibility for the creation

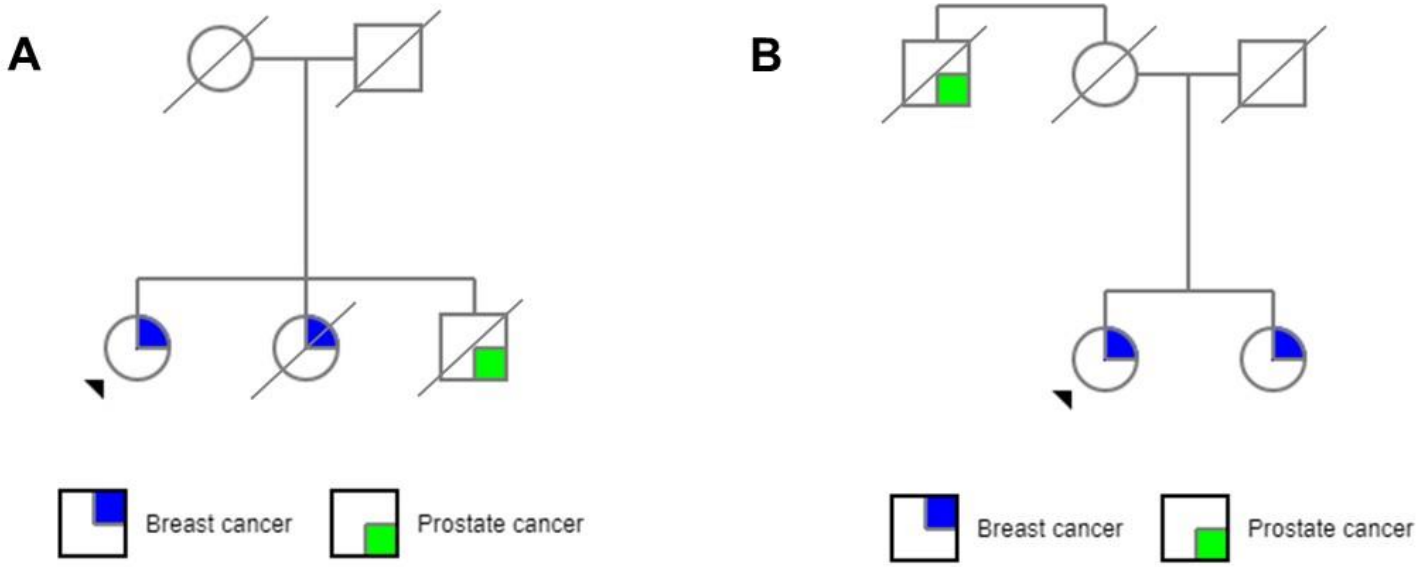


Figure 4.25 Family pedigrees for two familial BC patients identified with the *PALB2* c.2794G>A, p.Val932Met (rs34624036) mutation. **A**. Pedigree for CAM2336, exhibiting three siblings affected with breast or prostate ca in this Indian family. **B**. Pedigree for CAM2580, indicating a positive maternal family history of BC and prostate ca in this White family. The index patient is indicated by the arrow head. The symbol descriptions are indicated.

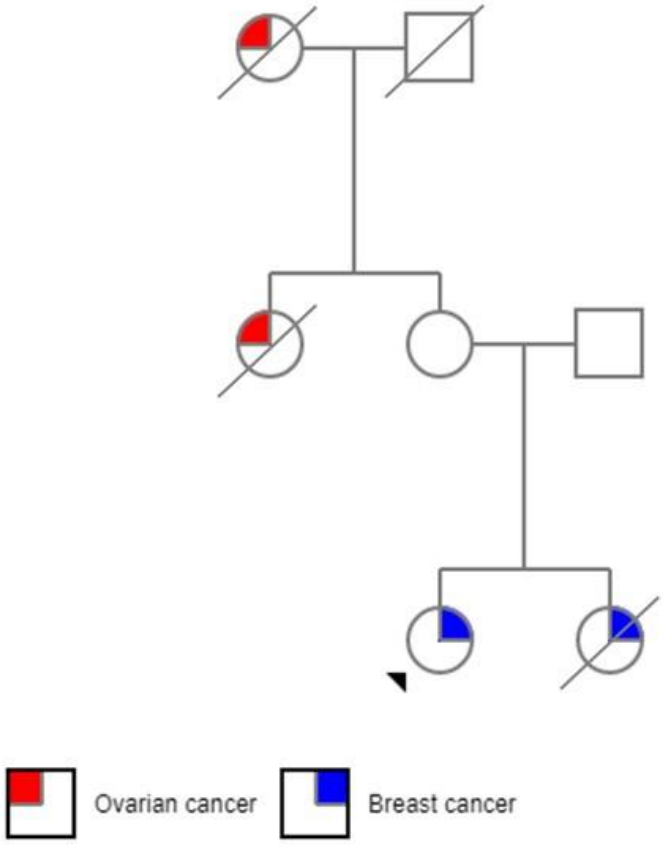


Figure 4.26 Family pedigree for CAM2854, indicating a positive maternal family history of BC and OVC in this White family. The index patient is indicated by the arrow head. This BC patient was diagnosed with unilateral BC (at the age of 49) and carried the *PALB2* c.2794G>A, p.Val932Met (rs34624036) mutation. Her sister was also diagnosed with BC, the age at onset is not indicated. She has since passed away. The maternal aunt and grandmother were both affected with OVC. The symbol descriptions are indicated.

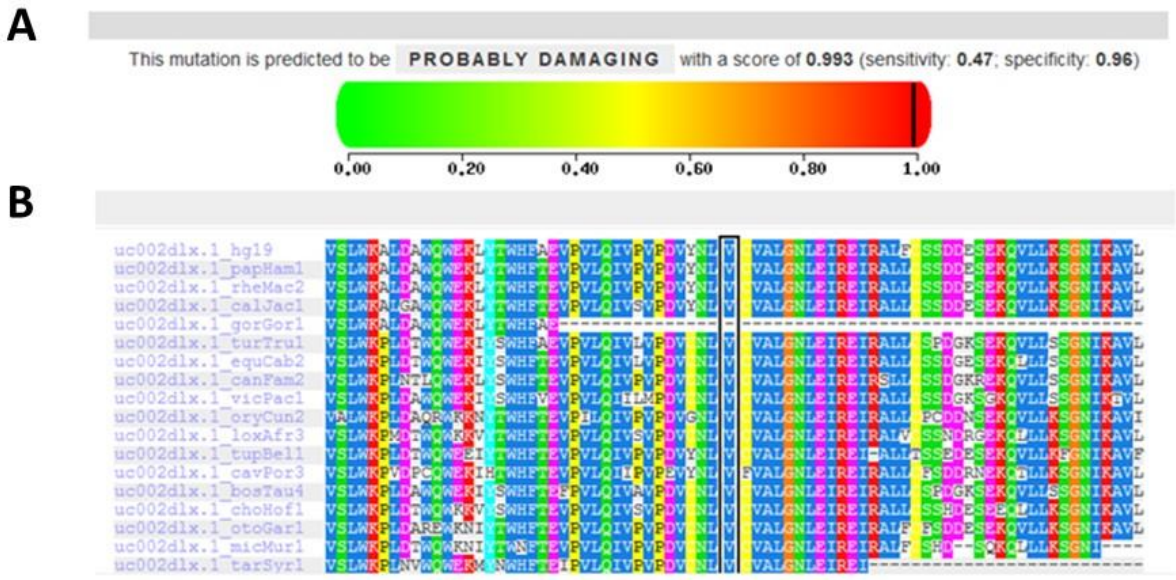


Figure 4.27 PolyPhen-2 prediction results for *PALB2* c.2794G>A, p.Val932Met. **A.** Prediction results according to HumVar. **B**. Multiple alignment across species of a section of 75 aa surrounding the position of the variant (http://genetics.bwh.harvard.edu/cgi-bin/ggi/ggi2.cgi, accessed on 27 December, 2017).

of an altered exonic splice site (51% – 75%), thereby potentially alteration of splicing. This result also corresponded with the findings of Wong-Brown *et al*. (2014), as they made use of four splicing programmes, namely Splicing Sequences Finder – SSF, Max-EntScan, NNSplice and HumanSplicingFinder. Although a destruction of a cryptic splice site was indicated by only two of the four programmes, all four indicated an increase in the score of a cryptic site for this variant (Wong-Brown *et al.*, 2014). This variant was located 46 bp from the 5' splice site.

Cryptic splice sites are present at the mRNA level and represents an mRNA sequence that has the potential for interacting with the spliceosome. These splice sites are only used when a natural splice site was disrupted by a mutation. The presence of mutations such as *PALB2* c. c.2794G>A, p.Val932Met in the underlying DNA can potentially activate a cryptic splice site in part of the transcript that usually is not spliced.

This missense mutation was present in 3.48% of the cohort (3/86). All three the patients were affected with BC at various stages in life. The global MAF for this variant is 0.0010, which indicates a rare occurrence throughout all populations (<0.01%). From the data presented for individual populations, it is clear that the variant has never been described for any Asian nor African population (https://www.ncbi.nlm.nih.gov/). It has mainly been detected within the European population, with a MAF of 0.005. This concurs with the results obtained for this study, as two of the three mutation carriers reported a White non-Afrikaner heritage, possibly reflecting an European connection. The detection of the variant within an Indian patient is new, as there is currently no MAF value for any of the Asian populations. This variant has furthermore never been observed in a homozygous state, which is normally considered natural variation. This was also the case for this study (Figure 4.10 C). Based on the bioinformatic prediction tools, the low global MAF (< 0.01), together with the variant being present globally in a heterozygous state only, this variant was provisionally classified to be likely pathogenic (Class 4). An IARC classification of Class 4 indicates that the variant, based on the evidence obtained, has a likelihood of pathogenicity of 95% and above. This class was differentiated from a VUS (variant of unknown clinical significance) based on the fact that there was significant but not irrefutable evidence for pathogenicity, as evident from the multiple agreement of *in silico* predictions and the low MAF.

This mutation has been submitted 13 times to ClinVar by various pathology laboratories, all reporting it to range from benign (Class 1) to likely benign (Class 2). The last date of interpretation of this variant is listed as 17 February 2017. For the majority of submissions, the criteria provided for classification was not indicated. The majority of submitters also did not state whether the patient was affected with ca, and if so, with which type of ca. All the submissions classified the variant based on literature reports, which included multiple peer review publications, ranging from 2004 – 2016. It is important to remember that ClinVar is a freely accessible, public archive for reports on the relationships between human variation and phenotypes. These submissions often are presented with or without supporting evidence, as in this case. The submissions are from both research and diagnostic laboratories, and are often not regularly curated. With diagnostic laboratories using a variety of methods and pipelines, the calculation models might result in a different classification. The classification of a variant in ClinVar should therefore be seen as a guide and not the ultimate for variant classification. It should be interpreted together with all other scientific findings. As their disclaimer states, they do not independently verify the submitted information as all variants are analysed using a global pipeline.

One of the criteria needed for a potential IARC classification of Class 4 (as proposed by Persaran *et al*., 2016) includes significant disease association in appropriately sized case-controlled studies. Such a study was performed by Thompson *et al.* (2015). The authors screened a total of 1996 Australian BC index cases and 1998 non-cancer controls for germline mutations in the coding regions of the gene. The *PALB2* c.2794G>A, p.Val932Met variant was detected and present in a total of 23 familial BC cases (23/1996, 1.15%), but was also detected within the same number of non-cancer control individuals (23/1998). Not only was there no difference between the occurrence of the variant between the BC cases and the controls, the percentage of occurrence in the Australian cohort was > 0.01 (MAF of the 1000 genome database was 0.0009). This relatively high incidence in both patients and controls outweighed the bioinformatic characterisation, which might have resulted in the classification of this variant by ClinVar and various investigators as benign (Class 1) to likely benign (Class 2) (Bodian *et al.*, 2014; Richards *et al.,* 2015; Thompson *et al.,* 2015).

Interesting to note, is that this variant has been reported for a wide range of populations, ranging from Poland (Kluska *et al.*, 2017), Spain (Blanco *et al.*, 2013) to Australia (Wong-Brown *et al.*, 2014; Thompson *et al.*, 2015). The majority of these populations are all located within Europe, which explains the European MAF of 0.005. This makes the classification and the interpretation of the clinical relevance of variants such as this difficult, as the results obtained comes from studies on highly selected populations. The variant has not been detected for populations such as the Chinese, African and African-American population.

The classification by ClinVar and other researchers should not be seen as the final conclusion regarding the clinical effect of this variant. Bodian and co-workers (2014) examined germline variation in five cancer-susceptibility genes in six healthy, ancestrally diverse individuals representing the general population. They determined that allele frequencies vary between ancestral groups. They concluded that for some variants, the minor allele frequency in one population could be the major allele frequency in another. Based on the various arguments discussed, the missense variant detected in exon 8 of *PALB2* was considered likely-pathogenic (Class 4) until a formal case-control study has been performed and the prevalence of the variant in the general population has been determine in an SA context. Only then, will its role be clearly defined.

4.3.2.2.7 *PALB2* c.2968G>A, p.Glu990Lys (novel)

Another novel missense mutation was detected within exon 9 (Table 4.2). This variant represented a G>A change at nucleotide 2968 (electropherogram displayed in Appendix I). The conversion resulted in an aa change, from Glutamic acid (GAA, E) to Lysine (AAA, K) at codon 990. The SNP was detected for a single Indian patient, namely CAM2335 (Table 4.2). The patient was diagnosed with BC at an early age (dx of 35) and had a strong family history of the disease (2 family members affected with BC excluding the index, and a single male affected with throat ca, Table 3.1). The age at diagnosis of the two family members affected with BC, was not known.

Glutamic acid is a non-essential aa and represent an acidic aa, whereas K is basic. Based on the change in the aa properties at codon 990, SIFT predicted the variant to be deleterious (0.01). This prediction was strengthened by the results for PolyPhen-2 (0.979), which also indicated the variant to be deleterious (Table 4.2).

This variant occurred in a highly conserved and critical region of the protein (Figure 4.28 A – B), namely the carboxy-terminal region which comprises of the WD-40 repeats (ranging from aa 791 – 1186, Figure 2.2). These beta-transducin repeats are involved in the localisation of BRCA2 as well as the association of BRCA2 with chromatin and various nuclear structures (Xia *et al.*, 2006) during its response to DNA damage. The association of PALB2 with BRCA2 is vital for the prevention of tumour formation (Park *et al.*, 2013). It is evident from the illustration in Figure 4.28 B that very little variation exist within this region across the 18 species highlighted. The occurrence of this variant at codon 990 will therefore most probably affect the surrounding residues at the mutation site, and could have an impact on the association of BRCA2 with PALB2.

Two of the *in silico* predictions indicated less damaging effects, namely Mutation Taster and Human Splicing Finder (Table 4.2). Based on the location of the variant in a critical functional domain and the low MAF of the variant in this study (1/86), it is proposed that this variant be considered a VUS (Class 3) due to the discrepancy amongst the prediction programmes. Without the comparison of 3D modelling of the wild type with that of the variant, no conclusion can be reached regarding the interior interaction of the new aa within the protein. It is proposed that this variant be screened for in more familial BC patients that tested negative for pathogenic *BRCA* mutations, in order to determine its prevalence in the SA population.

4.3.2.2.8 *PALB2* c.3434G>T, p.Gly1145Val (novel)

The final missense mutation to be discussed was detected in exon 13 (Table 4.2). This variant represented a G>T change at nucleotide 3434 (Figure 4.29 A – C). The conversion resulted in an aa change, from Glycine (GGT, G) to Valine (GTT, V) at codon 1145. The SNP was detected for a single Coloured BC patient, namely CAM2578 (Table 4.2). The patient was diagnosed with BC at an early age (dx of 42) and had a paternal family history of the disease (2 aunts affected with BC, Figure 4.30).

The presence of the variant resulted in an aa change, from Glycine (GGT, G) to V at codon 1145. Glycine represents a neutral and polar aa. It is evolutionarily conserved and represents the smallest aa of the 20. For this reason, it has the ability to fit into tight spaces of molecules where no other aa could possibly fit.

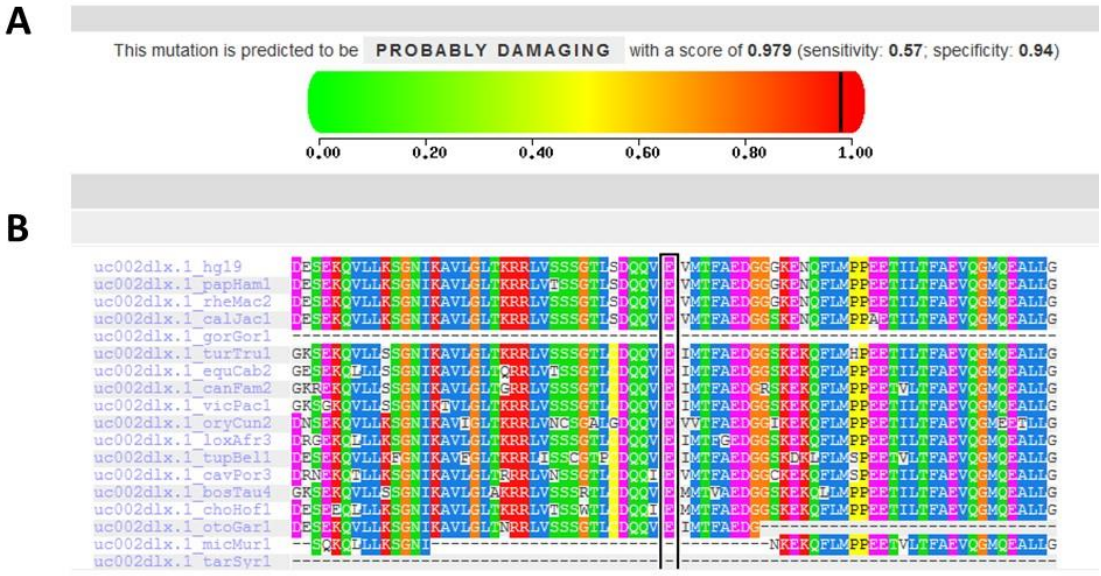


Figure 4.28 PolyPhen-2 prediction results for *PALB2* c.2968G>A, p.Glu990Lys. **A.** Prediction results according to HumVar. **B**. Multiple alignment across species of a section of 75 aa surrounding the position of the variant (http://genetics.bwh.harvard.edu/cgi-bin/ggi/ggi2.cgi, accessed on 27 December, 2017).

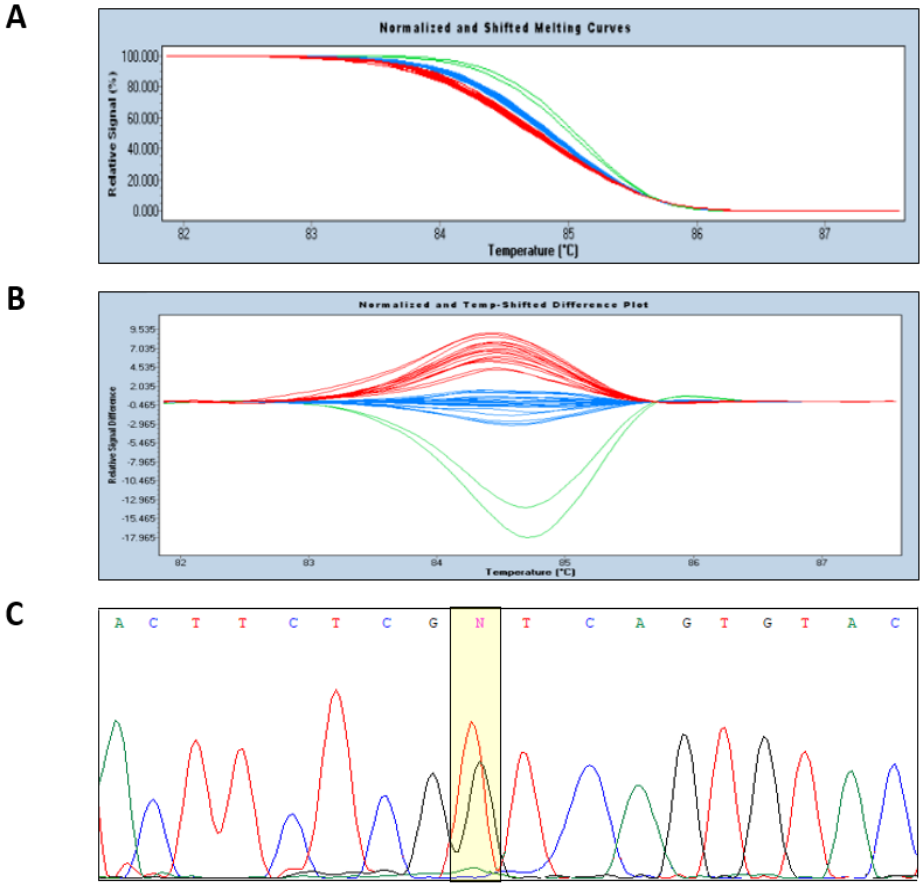


Figure 4.29 HRMA results for *PALB2* exon 13 for 23 familial BC patients, each performed in duplicate, of which a single patient (CAM2578) carried the *PALB2* c.3434G>T, p.Gly1145Val novel mutation. **A**. The normalized and temperature shifted melt curves based on the qPCR HRMA data. **B.** The difference plot generated from the melt, with the variant (green curves) grouping separate from the bulk of the samples (representing the wild type in blue and a common variant *PALB2* c.3300T>G, p.Thr1100Thr in red) on the baseline, with a relative signal difference of ~ 18.0 units. **C.** Sequencing electropherogram representing the *PALB2* c.3434G>T, p.Gly1145Val novel mutation, highlighted by the yellow box.

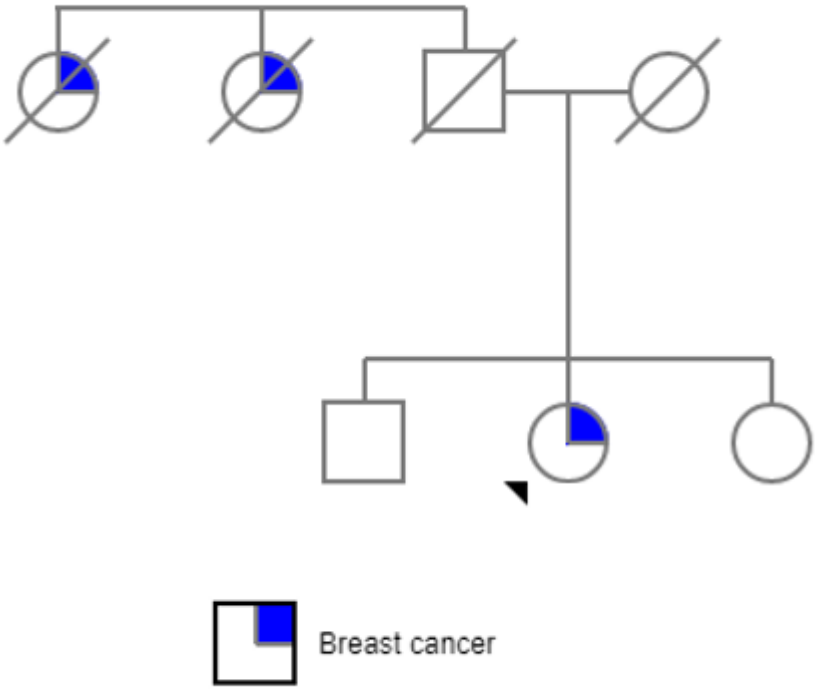


Figure 4.30 Family pedigree for CAM2578, indicating a positive paternal family history of BC in this Coloured family. The index patient is indicated by the arrow head. This BC patient was diagnosed with unilateral BC (at the age of 42) and carried the *PALB2* c.3434G>T, p.Gly1145Val novel mutation. The symbol description is indicated.

In contrast, V has an aliphatic, isopropyl side chain which makes it neutral and hydrophobic. This aa is also nonpolar. Based on these differences in structure and composition, SIFT predicted the variant to be deleterious (0.01, Table 4.2). This prediction was supported by the results for PolyPhen-2 (0.999), which also indicated the presence of the variant to be deleterious (Table 4.2, Fig. 4.31). This variant is also situated in the highly conserved WD-40 repeats (ranging from aa 791 – 1186, Figures 2.2 & 4.31), which facilitates the binding of BRCA2 in order for its localisation to the point of DNA damage. It is hypothesized that the substitution of the small G residue with the larger V with a side chain would impact on the surrounding vicinity of these aa's. As the binding of the various proteins to the WD-40 region are critical for DNA repair, the transformation and lack of space could have an impact on these binding sites, therefore impact DNA repair.

All four *in silico* prediction models are in agreement regarding the classification of this missense mutation. They all predicted the variant to be deleterious, although for Human Splicing Finder it was to a lesser extent (Table 4.2). Based on the location of the variant and the fact that this novel variant was only detected for one BC patient in this study (1/86), it is proposed that segregation analysis be performed in the extended family in order to determine whether the variant might segregate with the disease. If no affected family members are available to participate in such a study, a control study should be performed to determine the presence and prevalence of this variant in the general SA Coloured population before any predictions regarding risk for family members can be formulated.

4.4 Conclusion

In total, the study revealed two protein truncating and 18 missense variants (Table 4.2) in the 86 BC/OVC cases studied. Of these, two protein truncating mutations were classified as pathogenic (Class 5 – four patients), together with two putative likely pathogenic (Class IV) missense mutations (four patients). All the patients carrying a protein truncating mutation had an extensive family history of BC and other ca types, whereas in the families of the four patients with a putative Class 4 missense mutation, BC was the dominant ca type.

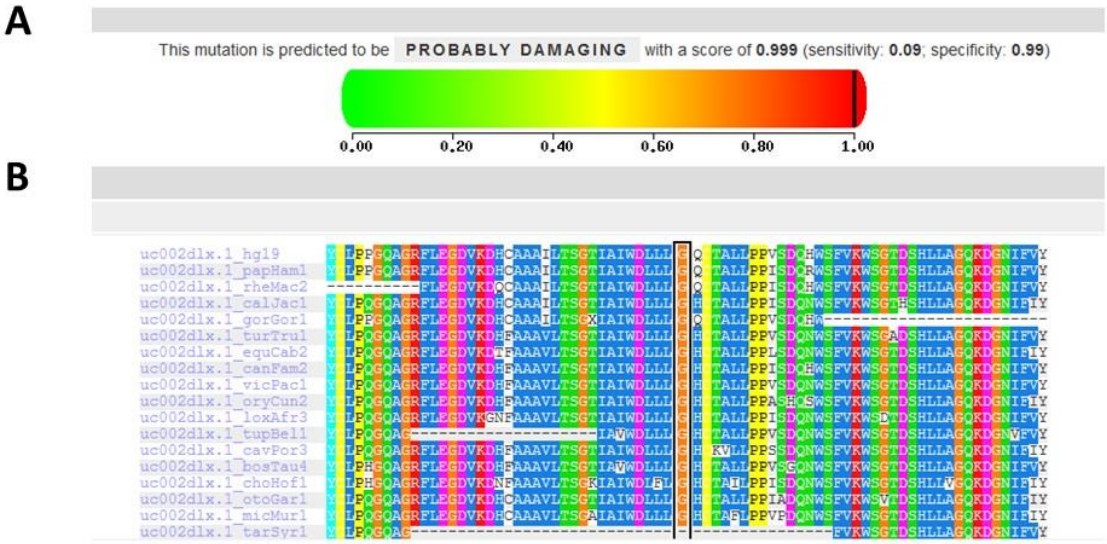


Figure 4.31 PolyPhen-2 prediction results for *PALB2* c.3434G>T, p.Gly1145Val novel mutation. **A.** Prediction results according to HumVar. **B**. Multiple alignment across species of a section of 75 aa surrounding the position of the variant (http://genetics.bwh.harvard.edu/cgi-bin/ggi/ggi2.cgi, accessed on 27 December, 2017).

The majority of variants was observed for the Coloured and SA Indian populations (Table 4.2), whereas for the other three groups, it was limited. This could be due to the unique admixture of these population groups as SA is characterized by extensive population diversity. The country harbours groups originating from African (79%), Asian (2.5%) and European (9.6%) populations (de Wit *et al.*, 2010), due to a multi-faceted colonization history based on the country's location with respect to major historical trade routes. The genetic contribution of these previously continentally seperated population groups led to the establishment of the highly admixed SA Coloured and Indian populations. Although the SA Indian population has been isolated from mainland Indian to some extent for more than three centuries, the original Indian population represents a substantial fraction of global diversity and has been shaped by multiple waves of migrations and admixture events (Basu *et al*., 2003; Abdulla *et al.*, 2009; Reich *et al.*, 2009). This initial admixture would have been brought to SA and also be present in the SA grouping representing India.

On the other hand, the higher prevalence in the Coloured and SA Indian subsets could also have been due to the higher number of Coloured and Indian patients included in the cohort. A total of 29 SA Coloured and 22 Indian BC/OVC patients were included, compared to 15 Black, 10 Afrikaner and 10 White families. Although 15 Black patients were studied, the number of variants observed for this group was very low (Table 4.2). The results obtained from this study indicate that the prevalence of *PALB2* mutations/variants might be different for each of the SA population groups, depending on the genetic heritage of these patients. The number of BC/OVC patients representing each of these ethnic groups needs to be increased in order to determine the prevalence of deleterious mutations. Once the range has been determined, the general population has to be screened for these variants in order to determine the possible relative risk involved with each variant for the development of breast- and other ca types.

Chapter 5

Conclusion and recommendations

In 2007, various familial BC investigators such as Rahman and colleagues stated that with the identification of *PALB2* as an additional BC predisposition gene, a clearer picture of the genetic architecture of BC susceptibility was emerging. At the time, investigators acknowledged that although the familial BC genes *BRCA1/2* would most probably represent the only major high-penetrance genes, the other genes could perhaps explain some of the remaining BC burden in high risk families negative for pathogenic *BRCA* mutations.

Although the estimated cumulative risk for the development of female BC in *PALB2* pathogenic mutation carriers is proposed to be 14% at age 50 years and 35% by the age 70 years (Antonio *et al.*, 2014), it was significantly influenced by birth cohort and other familial factors. These factors resulted in an absolute BC risk for mutation carriers with two or more first-degree relatives (such as the patients included in this study) of 58% for BC at 50 years of age. Although this increase in risk is lower than that for *BRCA1*, it is postulated to be similar to that of *BRCA2* for hereditary BC families (Antonio *et al.*, 2014).

This was the first SA study to investigate the prevalence of *PALB2* germline mutations in familial BC/OVC patients with a hereditary predisposition that tested negative for *BRCA1/2* pathogenic mutations. *PALB2* encodes a protein that together with *BRCA1*, *BRCA2* and *RAD51*, plays a critical role in maintaining genomic integrity by partaking in DNA repair. Loss of function *PALB2* mutations will indirectly affect the expression of BRCA2 which will result in genetic instability and tumorigenesis.

Germline inactivating mutations are relatively rare, varying from 0.1% to 3.9% depending on the population and the selection criteria used for the subset. The *PALB2* mutation rate for this study (4.6%, 4/86) is higher than those reported in the earlier SA studies (Sluiter, 2009; Francies *et al.*, 2015). This could be the due to the difference in the selection criteria, as both studies did not select for the presence of a positive family history of BC nor OVC. Both Sluiter (2009) and Francies *et al.* (2015) included cohorts of SA BC patients unselected for family history, diagnosed at a very early age with BC

or OVC, with the latter group adding triple-negative disease as an additional selection criterion. The strength of the current study lies in the fact that the focus was to include BC patients with a strong familial predisposition, as appropriate when investigating the impact of a familial BC susceptibility gene.

Another explanation for the difference in *PALB2* mutation rate could be that the majority of the participants in the Francies study was Black (78.7%), with very few Indian and Coloured representation (4.6% and 1.9% respectively). In the present study, not only did the Indian and Coloured patients represent the majority of patients, a very small number of variants were observed for the Black patients. The data presented preliminarily confirms the low prevalence of *PALB2* variants for this SA ethnic group. Francies *et al.* (2015) could have missed the detection of *PALB2* variants, as the Indian and Coloured groups were under represented in their study. Three of the four loss of function mutations were detected for patients representing these two SA ethnic groups.

Currently, no definitely pathogenic *PALB2* missense variants have been identified that play a significant role in BC predisposition, suggesting that they might be extremely rare (Tishckowitz *et al.*, 2012; Tavtigian *et al.*, 2014). Although the present study identified multiple missense variants, two (*PALB2* c.2794G>A, p.Val932Met and *PALB2* c.3434G>T, p.Gly1145Val) were provisionally proposed as potentially damaging based on the variant's location in the carboxy-terminal WD-40 domain (aa 836 – 1186). This domain has the characteristics of a seven-bladed β-propeller and provides a binding site for the N-terminus of BRCA2 (Oliver *et al.*, 2009). Interpretation of these missense variants regarding impact on clinical treatment remains challenging. There is currently a need to extend international efforts that classify rare missense variants observed for *BRCA1/2*, to include those reported for *PALB2.* This will assist in the clinical management of patients carrying these variants.

The classification of such a variant will most likely require the incorporation of many pieces of evidence of which one will be functional assays. Two missense mutations in this domain in close proximity to the location of the SA mutations (*PALB2* c.2816T>G, p.Lys939Trp and *PALB2* c.3428T>A, p.L1143P), have already been functionally tested and were proven to exhibit decreased capacity for DNA double-stranded break-induced homologous recombination (Park *et al.*, 2014). It is anticipated that groups such as the Functional Working Group of ENIGMA will develop *PALB2*

functional assays for clinical assessment in the future. There is optimism in the research community that a number of large initiatives will generate data quickly and on a large scale to make definite analyses related to variant classification and clinical outcomes a reality for *PALB2* in the near future (Southey *et al.*, 2016).

This study, together with Sluiter *et al.* (2009) confirmed the presence of pathogenic mutations in the SA population, particularly in the White, Coloured and Indian groups. It is therefore recommended that larger studies be conducted to determine the prevalence of these specific and other mutations in these cohorts. Attempts should be made to perform segregation analyses for the mutations detected, in order to determine the clinical impact of these *PALB2* mutations on related family members.

Chapter 6

References

6.1 Books and articles

Abbott DW, Freeman ML, Holt JT (1998). Double-strand break repair deficiency and radiation sensitivity in *BRCA2* mutant cancer cells. Journal of Natural Cancer Institute 90: 978 – 985.

Abdulla L, Taib I, Saller R (2009). Public perception of cancer risk using analytical hierarchy process. Journal of Applied Science 9 (12): 2319 – 2324.

Anjum S, Fourkala E, Zikan M, *et al.* (2014). A *BRCA1*-mutation associated DNA methylation signature in blood cells predicts sporadic breast cancer incidence and survival. Genome Medicine 6: 47.

Antoniou A, Pharoah PD, Narod S, *et al*. (2003). Average risks of breast and ovarian cancer associated with *BRCA1* or *BRCA2* mutations detected in case series unselected for family history: a combined analysis of 22 studies. American Journal of Human Genetics 72: 1117 – 1130.

Antoniou AC, Casadei S, Heikkinen T, *et al.* (2014). Breast-Cancer risk in families with mutations in *PALB2.* The New England Journal of Medicine 371: 6.

Antoniou AC, Sinilnikova OM, Simard S, *et al.* (2007). RAD51135GrC modifies breast cancer risk among *BRCA2* mutation carriers: results from a combined analysisof 19 studies. The American Journal of Human Genetics 81:1186 –1200.

Balia C, Sensi E, Lombardi G, *et al.* (2010). A novel inactivating mutation in an Italian breast cancer family. Springer Science 9: 531 – 536.

Basu A, Mukherjee N, Roy S, *et al.* (2003). Ethnic India: A genomic view, with special reference to peopling and structure. Genome Research 13: 2277 – 2290.

Blanco A, de la Hoya M, Balmana J, *et al.* (2013). Detection of a large rearrangement in *PALB2* in Spanish breast cancer families with male breast cancer. Breast Cancer Research and Treatment 132: 307 – 315.

Bodian DL, McCutcheon JN, Kothiyal P, *et al.* (2014). Germline Variation in Cancer-Susceptibility Genes in a healthy, Ancestrally Diverse Cohort: Implications for individual Genome Sequencing. Genome Sequencing. PLoS ONE 9 (4): e94554.

Bogdanova N, Sokolenko AP, Iyevleva AG *et al.* (2011). *PALB2* mutations in German and Russian patients with bilateral breast cancer. Breast Cancer Research and Treatment 126: 545 – 550.

Bray F, Jemal A, Grey N, *et al.* (2012). Global cancer transitions according to the Human Development Index (2008–2030): a population-based study. The Lancet Oncology 13: 790 – 801.

Buisson R & Masson JY (2012). PALB2 self-interaction controls homologous recombination. Nucleic Acids Research 40: 10312 – 10323.

Buisson R, Dion-Cote AM, Coulombe Y, *et al.* (2010). Cooperation of breast cancer proteins PALB2 and piccolo BRCA2 in stimulating homologous recombination. Nature Structural and Molecular Biology 17: 1247 – 1254.

Buisson R, Niraj J, Puah J, *et al.* (2014). Breast cancer proteins PALB2 and BRCA2 stimulate polymerase in recombination-associated DNA synthesis at blocked replication forks. Cell Reports 6: 553 – 564.

Calmon MF, Jeschke J, Zhang W, *et al.* (2015). Epigenetic silencing of neurofilament genes promotes an aggressive phenotype in breast cancer. Epigenetics 7: 622 – 632.

Cao A, Huang J, Hu Z, *et al.* (2008). The prevalence of *PALB2* germline mutations in *BRCA1/BRCA2* negative Chinese women with early onset breast cancer or affected relatives. Breast Cancer Research and Treatment 114: 457 – 462.

Carrozza MJ, Li B, Florens L, *et al.* (2005). Histone H3 methylation by Set2 directs deacetylation of coding regions by Rpd3S to suppress spurious intragenic transcription. Cell 123: 581 – 592.

Casadei S, Norquist BM, Walsh T, *et al.* (2011). Contribution to familial breast cancer of inherited mutations in the BRCA2-interacting protein PALB2. Cancer Research 71 (6): 2222 – 2229.

Castera L, Krieger S, Rousselin A, *et al.* (2014). Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture

targeting multiple candidate genes. European Journal of Human Genetics 22: 1305 – 1313.

Catucci, I, Peterlongo P, Ciceri S, *et al.* (2014). *PALB2* sequencing in Italian familial breast cancer cases reveals a high-risk mutation recurrent in the province of Bergamo. Genetics and Medicine 6: 9.

Churpek JE, Walsh T, Zheng Y, *et al.* (2015). Inherited predisposition to breast cancer among African-American women. Breast Cancer Research and Treatment 149: 31 – 39.

Claus EB, Schildkraut JM, Thompson WD, *et al.* (1996). The Genetic attributable risk of breast and ovarian cancer. American Cancer Society 77: 2318 – 2324.

Colditz GA, Sellers TA, Trapido E (2006). Epidemiology - identifying the causes and preventability of cancer? Nature Publishing Group volume 6.

Cook TD (2015). Quasi-Experimental Design. Wiley Encyclopaedia of Management. 11: 1 – 2.

Cui H, Cruz-Correa M, Giardiello FM, *et al.* (2003). Loss of *IGF2* Imprinting: A Potential Marker of Colorectal Cancer Risk. Science 5613: 1753 – 1755.

Cybulski C, Carrot-Zhang J, Kluźniak W, *et al.* (2015). Germline *RECQL* mutations are associated with breast cancer susceptibility. Nature Genetics 47: 643 – 646.

Damiola F, Schultz I, Barjhoux L, *et al.* (2015). Mutation analysis of *PALB2* gene in French breast cancer families. Breast Cancer Research and Treatment 154: 463 – 471.

Dansonka-Mieszkowska A, Kluska A, Moes J, *et al.* (2010). A novel germline *PALB2* deletion in Polish breast and ovarian cancer patients. BMC Medical Genetics 11: 20.

De Chalon SA, Teo Z, Park DJ, *et al.* (2010). Are *PALB2* mutations associated with increased risk of male breast cancer? Breast Cancer Research and Treatment 121: 253 – 255.

De Martel C, Ferlay J, Franceschi S, *et al.* (2012). Global burden of cancers attributable to infections in 2008: a review and synthetic analysis. Lancet Oncology 13: 607 – 615.

De Wit E, Delport W, Rugamika CE, *et al.* (2010). Genome-wide analysis of the structure of the South African Coloured Population in the Western Cape. Human Genetics 128: 145 – 153.

Ding YC, Steele L, Kuan CJ, *et al.* (2011). Mutations in *BRCA2* and *PALB2* in male breast cancer cases from the United States. Breast Cancer Research and Treatment 126: 771 – 778.

Dray E, Etchin J, Wiese C, *et al.* (2010). Enhancement of RAD51 recombinase activity by the tumour suppressor *PALB2*. Nature Structural and Molecular Biology 17: 1255 – 1259.

Easton DF, Pharoah DP, Antoniou AC (2015). Gene-panel Sequencing and the prediction of Breast-Cancer Risk. New England Journal of Medicine 372 (23): 2243 – 2257.

Erali M, Palais R, Wittwer C (2008). SNP genotyping by unlabelled probe melting analysis. Methods in Molecular Biology 429: 199 – 206.

Erkko H, Dowty GJ, Nikkila J, *et al.* (2008). Penetrance analysis of the *PALB2* c.1592delT founder mutation. Clinical Cancer Research 14: 4667 – 4671.

Erkko H, Xia B, Nikkilä J, *et al.* (2007). A recurrent mutation in *PALB2* in Finnish cancer families. Nature 446: 316 – 319.

Ernst T, Chase AJ, Score J, *et al.* (2010). Inactivating mutations of the histone methyltransferase gene *EZH2* in myeloid disorders. Nature Genetics 42 (8): 722 – 726.

Ewing B & Green P (1998). Base-calling of automated sequencer traces using Phred-II. Error probabilities. Genome Research 8: 186 – 94.

Ferlay J, Shin J, Bray F, *et al.* (2010). Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International Journal of Cancer 127: 2893 – 2917.

Fernandes PH, Saam J, Peterson J, *et al.* (2013). Comprehensive Sequencing of *PALB2* in Patients With Breast Cancer Suggests PALB2 Mutations Explain a Subset of Hereditary Breast Cancer. Cancer 120: 963 – 967.

Foulkes WD, Ghadirian P, Akbari MR, *et al.* (2007). Identification of a novel truncating *PALB2* mutation and analysis of its contribution to early-onset breast cancer in French-Canadian women. Breast Cancer Research 9: R83.

Foulkes WD, Smith IE, Reis-Filho JS (2010). Triple-negative breast cancer. The New England Journal of Medicine 363: 1938 – 1948.

Francies FZ, Wainstein T, De Leeneer K, *et al.* (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: 912 – 921.

Frank AS (2004). Genetic predisposition to cancer — insights from population genetics. Nature Reviews in Genetics 5: 764 – 772.

Frio RT, Haanpaa M, Pouchet C, *et al.* (2010). Mutation analysis of the gene encoding the PALB2-binding protein MRG15 in *BRCA1/2*-negative breast cancer families. Journal of Human Genetics 55: 842 – 843.

Garcia MJ, Ferna´ndez V, Osorio A, *et al.* (2009). Mutational analysis of FANCL, FANCM and the recently identified FANCI suggests that among the 13 known Fanconi Anemia genes, only FANCD1/BRCA2 plays a major role in high-risk breast cancer predisposition. Carcinogenesis 30: 1898 – 1902.

Ghosh S, Sur S, Yerram SR, *et al.* (2014). Hypersensitivities for acetaldehyde and other agents among cancer cells null for clinically relevant Fanconi anaemia genes. American Journal of Pathology 184: 260 – 270.

Gospodinov A & Herceg Z (2013). Shaping chromatin for repair. Mutation Research 752: 45 – 60.

Greenlee RT, Murray T, Bolden S, *et al.* (2000). Cancer Statistics, 2000. CA Cancer Journal of Clinicians 50: 7 – 33.

Gunnarsson H, Arason A, Gillanders EM, *et al.* (2008). Evidence against *PALB2* involvement in Icelandic breast cancer susceptibility. Journal of Negative Results in Biomedicine 7: 5.

Haanpaa M, Pylkäs K, Moilanen JS, Winqvist R (2013). Evaluation of the need for routine clinical testing of *PALB2* c.1592delT mutation in *BRCA* negative Northern Finnish breast cancer families. BioMedical Genetics 14: 82.

Hayakawa T, Zhang F, Hayakawa N, *et al.* (2010). MRG15 binds directly to PALB2 and stimulates homology-directed repair of chromosomal breaks. Journal of Cell Science 123: 1124 – 1130.

Heikkinen T, Karkkainen H, Aaltonen K, *et al.* (2009). The breast cancer susceptibility mutation *PALB2* 1592delT is associated with an aggressive tumour phenotype. Clinical Cancer Research 15: 3214 – 3222.

Hellebrand H, Sutter C, Honisch E, *et al.* (2011). Germline mutations in the PALB2 gene are population specific and occur with low frequencies in familial breast cancer. Human Mutations 32: 2176 – 2188.

Hug N, Longman D, Caceres F (2016). Mechanisms and regulation of the nonsense-mediated decay pathway. Nucleic Acid Research 44 (4): 1483 – 1495.

Jemal A, Siegel R, Xu J, *et al.* (2010). Cancer Statistics, 2010. CA Cancer Journal of clinicians 60: 277 – 300.

Jensen RB, Carreira A, Kowalczykowski SC (2010). Purified human BRCA2 stimulates RAD51-mediated recombination. Nature 467: 678 – 683.

Keenan T, Moy B, Mroz EA, *et al.* (2015). Comparison of the genomic landscape between primary breast cancer in African American versus White women and the Association of Racial Differences with Tumor Recurrence. Journal of Clinical Oncology 33: 3621 – 3627.

Kim JH, Choi DH, Cho DY, *et al.* (2010). *PALB2* mutations 1592delT and 229delT are not present in Korean breast cancer patients negative for *BRCA1* and *BRCA2* mutations. Breast Cancer Research and Treatment 122: 303 – 306.

Kircher M & Kelso J (2010). Methods, Models & Techniques: High-throughput DNA sequencing – concepts and limitations. Bioessays 32: 524 – 536.

Klug WS, Cummings MR, Spencer CA (2006). Concepts of Genetics. Pearson education Corporation ISBN 0-13-196894 – 7.

Kluska A, Balabas A, Piatkowska M, *et al.* (2017). *PALB2* mutations in *BRCA1/2*-mutation negative breast and ovarian cancer patients from Poland. BMC Medical Genomics 10: 14.

Kotsopoulos J, Rosen B, Fan I, *et al.* (2016a). Ten-year survival after epithelial ovarian cancer is not associated with *BRCA* mutation status. Gynaecology Oncology 140 (1): 42 – 47.

Kotsopoulos J, Sopik V, Rosen B, *et al.* (2016b). Frequency of germline PALB2 mutations among women with epithelial ovarian cancer. Familial Cancer 16: 29 – 34.

Krypuy M, Newnham GM, Thomas DM, *et al.* (2006). High resolution melting analysis for the rapid and sensitive detection of mutations in clinical samples: KRAS codon 12 and 13 mutations in non-small cell lung cancer. BMC Cancer 6: 295.

Kundu A, Bag S, Ramaiah S, *et al.* (2012). Leucine to proline substitution by SNP at position 197 in Caspase-9 gene expression leads to neuroblastoma: a bioinformatics analysis. Biotechnology 3: 225 – 234.

Leyton Y, Gonzalez-Hormazabal P, Blanco R, *et al.* (2015). Association of *PALB2* sequence variants with the risk of familial and early-onset breast cancer in a South-American population. BMC Cancer 15: 30.

Lichtenstein P, Holm Nv, Verkasalo Pk, *et al.* (2000). Environmental and heritable factors in the causation of cancer analyses of cohorts of twins from Sweden, Denmark, and Finland. New England Journal of Medicine 343: 78 − 85.

Liu J, Doty T, Gibson B, Heyer WD (2010). Human BRCA2 protein promotes RAD51 filament formation on RPA-covered single-stranded DNA. Nature Structural and Molecular Biology 17: 1260 – 1262.

Liu N, Schild D, Thelen MP, *et al.* (2002). Involvement of RAD51C in two distinct protein complexes of RAD51 paralogs in human cells. Nucleic Acid Research 30 (4): 1009 – 1015.

Lupas A (1996). Coiled coils: new structures and new functions. Trends in Biochemical Science 21: 375 – 382.

Ma J, Cai H, Wu T, *et al.* (2012). PALB2 interacts with KEAP1 to promote NRF2 nuclear accumulation and function, Molecular and Cellular Biology 32: 1506 – 1517.

Majumder P & Chatterjee D (2017). In silico study of increased susceptibility of breast cancer by mutant *PALB2* Gene 1. European Journal of Biotechnology and Bioscience 5 (4): 1 – 10.

Mamrak N, Shimamura A, Howlett N (2017). Recent discoveries in the molecular pathogenesis of the inherited bone marrow failure syndrome Fanconi anaemia. Blood reviews 31: 93 – 99.

Mastana SS, (2014). Unity in diversity: An overview of the genomic anthropology of India. Annals of Human Biology 41 (4): 287 – 299.

McInerney MN, Miller N, Rowan A, *et al.* (2010). Evaluation of variants in the *CHEK2, BRIP1* and *PALB2* genes in an Irish breast cancer cohort. Breast Cancer Research and Treatment 121: 203 – 210.

Meindl A, Hellebrand H, Wiek C, *et al.* (2010). Germline mutations in breast and ovarian cancer pedigrees establish RAD51C as a human cancer susceptibility gene. Nature Genetics 42: 410 – 414.

Metcalfe K, Dennis C, Poll A *et al.* (2017). Effect of decision aid for breast cancer prevention on decisional conflict in women with a BRCA1 or BRCA2 mutation:

a multisite, randomized, controlled trial. Genetics in Medicine volume 19: 330 – 336.

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

Mimitou EP & Symington LS (2009). DNA end resection: many nucleases make light work. DNA Repair 8: 983 – 995.

Moldovan GL & D'Andrea AD (2009). How the Fanconi anaemia pathway guards the genome. Annual Review of Genetics 43: 223 – 249.

Moynahan ME, Cui TY, Jasin M (2001a). Homology-directed DNA repair, mitomycin-c resistance, and chromosome stability is restored with correction of a Brca1mutation. Cancer Research 61: 4842 – 4850.

Moynahan ME, Pierce AJ, Jasin, M (2001b). BRCA2 is required for homology-directed repair of chromosomal breaks. Molecular Cell 7: 263 – 272.

Noskowicz M, Bogdanova N, Bermisheva M, *et al.* (2014). Prevalence of *PALB2* mutation c.509_510delGA in unselected breast cancer patients from Central and Eastern Europe. Familial Cancer 13: 137 – 142.

Oliver AW, Swift S, Lors CJ, *et al.* (2009). Structural basis for recruitment of BRCA2 by PALB2. EMBO reports 10: 990 – 996.

Papi L, Putignano AL, Congregati C, *et al.* (2009). A *PALB2* germline mutation associated with hereditary breast cancer in Italy. Familial Cancer 9 (2):181 – 185.

Park JY, Singh TR, Nassar N, *et al.* (2013). Breast cancer-associated missense mutants of the *PALB2* WD40 domain, which directly binds RAD51C, RAD51 and BRCA2, disrupt DNA repair. Oncogene 33: 4803 – 4812.

Park JY, Zhang F, Andreassen PR (2014). PALB2: The hub of a network of tumour suppressors involved in DNA damage response. Elsevier 1846: 263 – 275.

Pauty J, Rodrigue A, Couturier A, *et al.* (2014). Exploring the roles of PALB2 at the crossroads of DNA repair and cancer. Biochemistry Journal 460: 331 – 342.

Persaran T, Karam R, Huether R, *et al*. (2016). Beyond DNA: An Integrated and Functional Approach for Classifying Germline Variants in Breast Cancer Genes. International Journal of Breast Cancer ID 2469523: 1 – 10.

Pylkäs K, Erkko H, Nikkilä J, *et al.* (2008). Analysis of large deletions in *BRCA1*, *BRCA2* and *PALB2* genes in Finnish breast and ovarian cancer families. BioMedical Cancer 8: 146.

Quintana-Murci L, Harmant C, Quach H *& et al*. (2010). Strong Maternal Khoisan Contribution to the South African Coloured Population: A Case of Gender-Biased Admixture. The American Journal of Human Genetics 86: 611 – 620.

Rahman N, Seal S, Thompson D, *et al.* (2007). *PALB2,* which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nature Genetics 39 (2): 165 – 167.

Ramus S, Song H, Dicks E, *et al.* (2015). Germline Mutations in the *BRIP1, BARD1, PALB2*, and NBN Genes in Women with Ovarian Cancer. Journal of National Cancer Institute 107 (11): 214.

Rea S, Xouri G, Akhtar A (2007). Males absent on the first (MOF): from flies to humans. Oncogene 26: 5385 – 5394.

Reeves MD, Yawitch MT, Van Der Merwe NC, *et al.* (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: 677 – 682.

Reich D, Thangaraj K, Patterson N, *et al.* (2009). Reconstructing Indian population history. Nature 461: 489 – 494.

Reid S, Schindler D, Hanenberg H, *et al.* (2007). Bi-allelic mutations in *PALB2* cause Fanconi anaemia subtype FA-N and predispose to childhood cancer. Nature genetics 39: 162 – 164.

Richards S, Aziz N, Bale S, *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: 405 – 423.

Roy R, Chun J, Powell SN (2012). BRCA1 and BRCA2: different roles in a common pathway of genome protection. Nature Review of Cancer 12: 68 – 78.

Roychoudhury S, Sangita R, Analabha B, *et al.* (2001). Genomic structures and population histories of linguistically distinct tribal groups of India. Human Genetics 109: 339 – 350.

Ruskovaa L & Raclavskya V (2011). The potential of high resolution melting analysis (HRMA) to streamline, facilitate and enrich routine diagnostics in medical

microbiology. Biomed Pap Medical Faculty University of Palacky Olomouc Czech Republic 155 (3): 239 – 252.

Scully R & Xie A (2013). Double strand break repair functions of histone H2AX, Mutation Research 750: 5 – 14.

Siegel R, Ma J, Zou Z, Jemal A (2014). Cancer Statistics. CA: A Cancer Journal for Clinicians 64: 9 – 29.

Slater EP, Langer P, Niemczyk E, *et al.* (2010). *PALB2* mutations in European pancreatic cancer families. Clinical Genetics 78: 490 – 494.

Sluiter M, Mew S, van Rensburg EJ (2009). *PALB2* sequence variants in young South African breast cancer patients. Springer Science 8: 347 – 353.

Smeenk G & van Attikum H (2013). The chromatin response to DNA breaks: leaving a mark on genome integrity, Annual Reviews of Biochemistry 82: 55 – 80.

Southey M, Goldga D, Winqvist R *et al.* (2016). *PALB2, CHEK2* and *ATM* rare variants and cancer risk: data from COGS. Journal of Medical Genetics 10: 1 – 12.

Southey MC, Teo ZL, Dowty JG, *et al.* (2010). A *PALB2* mutation associated with high risk of breast cancer. Breast Cancer Research 12: R109.

Stirnimann CU, Petsalaki E, Russell RB, *et al.* (2010). WD40 proteins propel cellular networks. Trends in Biochemical Science 35: 565 – 574.

Stratton MR & Rahman N (2008). The emerging landscape of breast cancer susceptibility. Nature Genetics 40 (1): 17 – 22.

Suwaki N, Klare K, Tarsounas M (2011). RAD51 paralogs: roles in DNA damage signalling, recombination repair and tumourigenesis. Seminars in Cell and Developmental Biology 22: 898 – 905.

Sy SM, Huen MS, Chen J (2009a). MRG15 is a novel *PALB2*-interacting factor involved in homologous recombination, Journal of Biological Chemistry 284: 21127 – 21131.

Sy SM, Huen MS, Chen J (2009b). *PALB2* is an integral component of the *BRCA* complex required for homologous recombination repair. Proceedings of the National Academy of Sciences USA 106: 7155 – 7160.

Takata M, Sasaki MS, Tachiiri S, *et al.* (2001). Chromosome instability and defective recombination repair in knockout mutants of the five Rad51 paralogs. Molecular Cell Biology 21: 2858 – 2866.

Tavtigian SV & Chenevix-Trench G (2014). Growing recognition of the role of rare missense substitutions in breast cancer susceptibility. Biomark Medicine 8 (4): 589 – 603.

Teo ZL, Park DJ, Provenzano E, *et al.* (2013). Prevalence of PALB2 mutations in Australian multiple-case breast cancer families. Breast Cancer Research 15: R17.

Thompson AR, Gorringe KL, Rowley SM, *et al.* (2015). Prevalence of *PALB2* mutations in Australian familial breast cancer cases and controls. Breast Cancer Research 17: 111.

Thompson D & Easton D (2004). The genetic epidemiology of breast cancer genes. Journal of Mammary Gland Biology and Neoplasia 9: 221 – 36.

Thompson JF, Reifenberger JG, Giladi E, *et al.* (2012). Single-step capture and sequencing of natural DNA for detection of BRCA1 mutations. Genome Research 22 (2): 340 – 345.

Thorslund T, McIlwraith M J, Compton SA, *et al.* (2010). The breast cancer tumour suppressor BRCA2 promotes the specific targeting of RAD51 to single-stranded DNA. Nature Structural and Molecular Biology 17: 1263 – 1265.

Tischkowitz M, Capanu M, Sabbaghian N, *et al.* (2012). Rare germline mutations in *PALB2* and breast cancer risk: A population-based study. Human Mutation 33 (4): 674 – 680.

Tischkowitz M, Xia B (2010). *PALB2/FANCN*: recombining cancer and Fanconi anaemia. Cancer Research 70: 7353 – 7359.

Tischkowitz M, Xia B, Sabbaghian N, *et al.* (2007). Analysis of *PALB2/FANCN*-associated breast cancer families. The National Academy of Sciences of the USA 104 (16): 6788 – 6793.

Tischkowitz MD, Sabbaghian N, Hamel N, *et al.* (2009). Analysis of the gene coding for the BRCA2-interacting protein PALB2 in familial and sporadic pancreatic cancer. Gastroenterology 137: 1183 – 1186.

Van der Groep P, Bouter A, van der Zanden R, *et al.* (2006). Distinction between hereditary and sporadic breast cancer on the basis of clinicopathological data. Journal of Clinical Pathology 59: 611 – 617.

Van der Merwe NC, Hamel N, Schneider SR, *et al.* (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 Stoep N, van Paridon CDM, Janssens T, Krenkova P, *et al.* (2008). Diagnostic Guidelines for High-Resolution Melting Curve (HRM) Analysis: an interlaboratory validation of *BRCA1* mutation scanning using the 96-well LightScannerTM. Human mutation 30: 899 – 900.

Venkitaraman AR (2002). Cancer susceptibility and the functions of *BRCA1* and *BRCA2*. Cell 108: 171 − 182.

Verhoog LC, Brekelmans CM, Seynaeve C, *et al.* (1998). Survival and tumour characteristics of breast-cancer patients with germline mutations of *BRCA1*. The Lancet 351: 316 – 321.

Vietri MT, Caliendo G, Schiano C, *et al.* (2015). Analysis of *PALB2* in a cohort of Italian breast cancer patients: identification of a novel *PALB2* truncating mutation. Springer 14 (3): 341 – 348.

Vossen RH, Aten E, Roos A, *et al.* (2009). High-Resolution Melting Analysis (HRMA)—More Than Just Sequence Variant Screening. Human Mutation 6: 860 – 866.

Walker AR, Adam FI, Walker BF (2004). Breast cancer in black African women: a changing situation. The Journal of the Royal Society for the Promotion of Health 124 (2): 81 – 85.

Wang AT & Smogorzewska A (2015). Snapshot: Fanconi anaemia and associated proteins. Cell 160 (1): 354.

Wang W (2007)**.** Emergence of a DNA-damage response network consisting of Fanconi anaemia and BRCA proteins. Nature Reviews in Genetics 8: 735 – 748.

Wojcik P, Jasiowka M, Strycharz E, *et al.* (2016). Recurrent mutations of *BRCA1, BRCA2* and *PALB2* in the population of breast and ovarian cancer patients in Southern Poland. Hereditary Cancer in Clinical Practice 14: 5.

Wong RS (2011). Apoptosis in cancer: from pathogenesis to treatment. Journal of Experimental and Clinical Cancer Research 30: 87.

Wong-Brown MW, Avery-Kiejda KA, Bowden NA. Scott RJ (2013). Low prevalence of germline *PALB2* mutations in Australian triple-negative breast cancer. International Journal of Cancer 134: 301 – 305.

Wooster R, Bignell G, Lancaster J, *et al.* (1995). Identification of the breast cancer susceptibility gene BRCA2. Nature 378: 789 – 792.

Wooster R, Neuhausen SL, Mangion J*, et al.* (1994). Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12–13. Science 265: 2088 – 2090.

Wu J, Chen Y, Lu LY, *et al.* (2011). Chfr and RNF8 synergistically regulate ATM activation. Nature Structural and Molecular Biology 18: 761 – 768.

Xia B, Dorsman JC, Ameziane N, *et al.* (2007). Fanconi anaemia is associated with a defect in the *BRCA2* partner *PALB2*. Nature Genetics 39: 159 – 161.

Xia B, Sheng Q, Nakanishi K, *et al.* (2006). Control of BRCA2 cellular and clinical functions by a nuclear partner, PALB2. Molecular Cell 22 (6): 719 – 729.

Xie T, Graveline R, Kumar GS, *et al.* (2012). Structural basis for molecular interactions involving MRG domains: implications in chromatin biology. Structure 20: 151 – 160.

Zhang DD, Hannink M (2003). Distinct cysteine residues in Keap1 are required for Keap1-dependent ubiquitination of Nrf2 and for stabilization of Nrf2 by chemopreventive agents and oxidative stress, Molecular Cell Biology 23: 8137 – 8151.

Zhang F, Fan Q, Ren K, *et al.* (2009a). PALB2 functionally connects the breast cancer susceptibility proteins BRCA1and BRCA2. Molecular Cancer Research 7 (7): 1110 –1118.

Zhang F, Ma J, Wu J, Ye L, Cai H, *et al.* (2009b). PALB2 Links BRCA1 and BRCA2 in the DNA-Damage Response. Current Biology 19: 524 – 529.

Zheng Y, Zhang J, Niu Q, *et al.* (2012). Novel Germline *PALB2* Truncating Mutations in African American Breast Cancer Patients. Cancer 1: 2012.

Ziyaie D, Hupp RT, Thompson AT (2000). p53 and breast cancer. The breast 9: 239 – 246.

6.2 Electronic sources

BRCA1 gene: BRCA1, DNA repair associated [Online]. Available at https://ghr.nlm.nih.gov/gene/BRCA1#conditions (accessed 30 May 2016).

Burke KA, Berman S, Geyer FC, *et al.* (2017). Mutational landscape of breast cancers from *PALB2* germline mutation carriers [abstract]. Cancer Research 2017: 77, DOI 10.1158/1538-7445. Available at http://cancerres.aacrjournals.org/content/77/4_Supplement/P2-03-01.short. (Accessed 10 December 2017).

Combrink HMVE, Van der Merwe NC, Visser B (2016). Molecular screening of the South African Indian population for *BRCA1* and *BRCA2* using high resolution melting analysis. University of the Free State 11660.4563. Available at http://hdl.handle.net/11660/4563 (Accessed 20 November 2017).

ENIGMA guidelines http://enigmaconsortium.org/documents/publications/ENIGMA_Rules_2015-03-26.pdf (Accessed 20 June 2016).

Hamosh A (2016). Partner and localiser of BRCA2; PALB2 [Online]. Available at http://www.omim.org/entry/610355?search=palb2&highlight=palb2.

Oosthuizen J, Van der Merwe NC, Foulkes WD (2016). Molecular screening of Coloured South African breast cancer patients for the presence of *BRCA* mutations using high resolution melting analysis [Online]. University of Free State 11660.6426. Available at http://hdl.handle.net/11660/6426 (Accessed 20 June 2017).

Reddy N, Malipatil B, Kumar S (2016). A rare case of familial multiple subcutaneous lipomatosis with novel PALB2 mutation and increased predilection to cancers [Online]. Hematology and oncology stem cell theory. Available at http://dx.doi.org/10.1016/j.hemonc.2016.01.001 (Accessed 7 January 2016).

Taylor PC, Medina MND (2013). Educational research paradigms: From positivism to multiparadigmatic [Online]. Journal for Meaning Centered Education 1. Available at http://www.meaningcentered.org/journal/volume-01/educational-research-paradigms-from-positivism-tomultiparadigmatic/ (Accessed 14 February 2016).

Technical Bulletin, NanoDrop - http://www.nanodrop.com/Library/T042-NanoDrop-spectrophotometers-Nucleic-Acid-Purity-Ratios.pdf (Accessed 5 June 2015).

Vogt H (2017). Systems Medicine as a theoretical framework for primary care medicine: A critical analysis. Norwegian University of Science and technology. Available at https://brage.bibsys.no/xmlui/bitstream/handle/11250/2441891/Henrik%20Vogt_PhD.pdf?sequence=1 (accessed 12 February 2018).

6.3 Databases

1000 Genomes	http://www.ncbi.nlm.gov/tools/1000genomes/
1000 Genomes project database	https://www.ncbi.nlm.nih.gov/projects/SNP/
Align Grantham Variation Grantham Deviation	http://agvgd.hci.utah.edu/agvgd_input.php
CANSA	http://www.cansa.org.za/south-african-cancer-statistics/
Chromas version 2.31	www.technelysium.com.au
ClinVar	http://www.ncbi.nlm.nih.gov/clinvar/
Condel	(https://omictools.com/
dbSNP	http://www.ncbi.nlm.nih.gov/snp/
Exome Aggregation Consortium (ExAC)	http://exac.broadinstitute.org/
Expasy	http://au.expasy.org/tools/dna.html
HGMD	http://www.hgmd.cf.ac.uk/ac/
HGVS	http://www.hgvs.org/mutnomen/
Human Splicing Finder	http://www.umd.be/HSF3/HSF.html
Integrated DNA technologies	https://www.idtdna.com
LALIGN	www.ch.embnet.org/software/LALIGN
Leiden Open Variant Database	http://www.lovd.nl/3.0/home
Mutation Taster	http://www.mutationtaster.org/
Online Mendelian of Man	www.omim.org
Polymorphism Phenotyping version 2 (PolyPhen-2)	http://genetics.bwh.harvard.edu/pph2/
Progeny (pedigree constructed tool)	https://pedigree.progenygenetics.com/

SNPedia	www.snpedia.com/index.php/Rs587782005
South African Cancer Statistics	http://www.cansa.org.za/south-african-cancer-statistics/
South African genetics counselling association.	http://www.geneticcounselling.co.za.
Sorting Intolerant From Tolerant (SIFT)	http://sift.jcvi.org/

Appendix A

health

Department:
Health
PROVINCE OF KWAZULU-NATAL

INKOSI ALBERT LUTHULI CENTRAL HOSPITAL
Ethekwini Health District
Office of the Chief Executive Officer
800 Vusi Mzimela (Bellair) Road, Mayville, 4091
Private Bag X03, Mayville, 4058
Tel: 031 - 240 1034
Fax: 031 - 240 1005
Email: Gugu.Duma@ialch.co.za
www.kznhealth.gov.za

Enquiries: Dr S.T. Mtshali
Date: 15 August 2014

Ms Gugu Khumalo
c/o Department of Health KwaZulu-Natal

Email: gugu.khumalo@kznhealth.gov.za

RESEARCH PROPOSAL

A research proposal has been submitted to me by another centre requesting that data of IALCH patients be used for the study "MOLECULAR SCREENING OF THE SOUTH AFRICAN INDIAN POPULATION FOR BRCA1 AND BRCA2 USING HIGH RESOLUTION MELTING ANALYSIS".

The protocol has been forwarded to the Research Committee in Pietermaritzburg for clearance and the research has been approved.

I wish to inform you that I consent to the request for samples of IALCH patients to be used for the research in question.

Yours sincerely,

Dr S.T. Mtshali
Chief Executive Officer

uMnyango Wezempilo. Departement van Gesondheid

FIGHTING DISEASE, FIGHTING POVERTY, GIVING HOPE

Appendix B

Western Cape Government
Health

Tygerberg Hospital

REFERENCE: Research Projects
ENQUIRIES: Dr G G Marinus
TELEFONE: 021 938-6267

ETHICS NO: ECUF 108/2014

Molecular screening of South African coloured breast cancer patients for the presence of BRCA mutations using high resolution melting analysis

Dear Van der Merwe

PERMISSION TO CONDUCT YOUR RESEARCH AT TYGERBERG HOSPITAL

In accordance with the Provincial Research Policy and Tygerberg Hospital Notice No 40/2009, permission is hereby granted for you to conduct the above-mentioned research here at Tygerberg Hospital.

DR D ERASMUS
CHIEF EXECUTIVE OFFICER
Date: 22 August 2014

Administration Building, Francie van Zijl Avenue, Parow, 7500
tel: +27 21 938-6267 fax: +27 21 938-4890

Private Bag X3, Tygerberg, 7505
www.capegateway.gov.za

Appendix C

Guidelines from the SASHG Committee for publication purposes regarding:
Nomenclature for South African populations

Prepared by Lisa Roberts, Jacquie Greenberg and Michele Ramsay, with input from Himla Soodyall and the current SASHG committee - June 2013

When publishing genetic or biomedical research studies involving patients, controls or population groups in South Africa, it is often relevant and necessary to identify the groups by their ethnic affiliations. The categories are usually used as proxies for genetic ancestry, but the ways in which individuals "self-identify" are influenced by many factors, including geographic origin, language and culture. As genetic and biomedical researchers, we take cognisance of the fact that these are labels of convenience that do not necessarily accurately reflect ancestry, nor do they define homogenous groups.

We recognise that preferred identifiers for South African ethnic groups vary between individuals, groups, nations and organisations. In citing ethnic labels, we therefore recommend the use of the names or categories that the people call themselves. Today, South Africans broadly consider themselves White, Black (black African), Coloured or Indian. Importantly too, these are also the categories used in the South African population census (also at www.statssa.gov).

Due to the international opinion that "Coloured" is a derogatory term, it is necessary to clarify that in the South African context, populations who self-identify as Coloured have a complex history of ancestrally derived admixture. These groups, also frequently termed "Mixed Ancestry" are ancestrally derived from admixtures of one or more of the indigenous African populations (Khoe- and San-speaking or Bantu-speaking), immigrants from Western Europe, or slave labourers from West Africa, Indonesia, Madagascar, Java, India and Malaysia. The term "Coloured" in South Africa is therefore a name that encompasses a wide range of people who are unique to this country.

As a guide, the SASHG committee therefore recommends, for publication purposes for genetic and biomedical studies, the use of the following nomenclature for South African populations: White, Black African, Coloured and Indian. This terminology should be accompanied by a brief description of the geographic origin and inclusion criteria of the participants of the specific research study.

If necessary, depending on the journal and/or the reviewers' comments, it is acceptable to substitute the term "Mixed Ancestry" for "Coloured", accompanied by a description of the cohort(s) under study.

Black African communities in South Africa also identify themselves by language: the Zulu people or amaZulu (who speak isiZulu), Xhosa or amaXhosa (isiXhosa), Sotho (seSotho), etc. and these affiliations should be used if the data are available and relevant to the study. Similarly, since San hunter-gatherers and Khoe pastoralists identify themselves with their community names (e.g. !Xun, ‡Khomani, //Ganaa, Hai//om, etc.) these affiliations should be used when available and relevant, however if they are grouped together the term Khoe-San is recommended.

We acknowledge that in the future there may no longer be a need for such "labels" but at present, particularly for genetic research purposes, this information is pertinent and relevant.

Appendix D

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

NB: IT IS IMPORTANT TO INFORM THE STUDENT ABOUT POSTGRADUATE BURSARIES – ADVERTISED IN JANUARY OF EACH YEAR

SECTION B MUST BE COMPLETED BY THE CHAIR OF THE EVALUATION COMMITTEE

Initials and surname (student) MF Makhetha

Student number 2010079537

Chair of the Evaluation Committee: Prof PH Wessels

Members of the Evaluation Committee:	Present	Absent
Prof PH Wessels (Chairman)	X	
Prof C Viljoen	X	
Prof G Joubert	X	
Prof F Burt		X
Dr A de Kock	X	

Date of meeting: 30 October 2015

Proposed title: ***Screening for PALB2 mutations in South African women with BRCA negative familial breast cancer***

Title has been language edited by: Committee members

*(Marked in **BOLD**)*

Ethical aspects have been addressed:	**YES**	No
Facilities available:	**YES**	No
Funds available:	**YES**	No
Time schedule acceptable:	**YES**	No
The study is recommended:	**YES**	No

Remarks: A very well planned protocol for an exciting study with much potential to contribute to the body of knowledge in this field. It is the opinion of the committee that this protocol is ready to be sent to the Research Ethics Committee of the Faculty of Health Sciences.

SIGNATURE OF CHAIR (EVALUATION COMMITTEE)

17 November 2015
DATE

Appendix E

UNIVERSITY OF THE FREE STATE
UNIVERSITEIT VAN DIE VRYSTAAT
YUNIVESITHI YA FREISTATA

UFS·UV
HEALTH SCIENCES
GESONDHEIDSWETENSKAPPE

IRB nr 00006240
REC Reference nr 230408-011
IORG0005187
FWA00012784

26 February 2016

DR NC VAN DER MERWE
DIV HUMAN GENETICS
FACULTY OF HEALTH SCIENCES
UFS

Dear Dr Van Der Merwe

ECUFS NR 31/95 C (SUB STUDY)
DR NC VAN DER MERWE
MF MAKHETHA
DIVISION HUMAN GENETICS
PROJECT TITLE: SCREENING FOR *PALB2* MUTATIONS IN SA WOMEN WITH BRCA NEGATIVE FAMILIAL BREAST CANCER

1. You are hereby kindly informed that the Health Sciences Research Ethics Committee (HSREC) reviewed the above research project and it was presented at the meeting on 23 February 2016. Research may not be conducted before the following condition(s) has/have been met and the HSREC grants final approval for the project:

   1.1 The signed permission letters from NHLS and UKZN must be submitted before final approval will be granted.

   *Upon receipt of the above document(s), the HSREC will issue a final approval letter. Only thereafter may the study be conducted.

2. The Committee must be informed of any serious adverse event and/or termination of the study.

3. Any amendment, extension or other modifications to the protocol must be submitted to the HSREC for approval.

4. Kindly use the **HSREC NR** as reference in correspondence to HSREC Administration.

5. Thus, this letter only serves as **conditional** approval.

6. The HSREC functions in compliance with, but not limited to, the following documents and guidelines: The SA National Health Act. No. 61 of 2003; Ethics in Health Research: Principles, Structures and Processes (2015); SA GCP(2006); Declaration of Helsinki; The Belmont Report; The US Office of Human Research Protections 45 CFR 461 (for non-exempt research with human participants conducted or supported by the US Department of Health and Human Services-(HHS), 21 CFR 50, 21 CFR 56; CIOMS; ICH-GCP-E6 Sections 1-4; The International Conference on Harmonization and Technical Requirements for Registration of Pharmaceuticals for Human Use (ICH Tripartite), Guidelines of the SA Medicines Control Council as well as Laws and Regulations with regard to the Control of Medicines, Constitution of the Ethics Committee of the Faculty of Health Sciences.

Yours faithfully

.....
DR SM LE GRANGE
CHAIR: HEALTH SCIENCES RESEARCH ETHICS COMMITTEE

Health Sciences Research Ethics Committee
Office of the Dean: Health Sciences
T: +27 (0)51 401 7795/7794 | F: +27 (0)51 444 4359 | E: ethicsfhs@ufs.ac.za
Block D, Dean's Division, Room D104 | P.O. Box/Posbus 339 (Internal Post Box G40) | Bloemfontein 9300 | South Africa
www.ufs.ac.za

Appendix F

Additional DNA consent for Inclusion into a research project on familial breast cancer

CONSENT FOR ADDITIONAL DNA STUDIES

1. I, _____ , request that an attempt be ma using genetic material to assess the probability that: I (name:_____) might have inheritec disease-causing mutation in the gene for: **FAMILIAL BREAST CANCER.**

2. I understand that the genetic material for analysis is to be obtained from: blood cells.

3. I acknowledge that a portion of the sample will be stored indefinitely for later use for:

   (a) possible re-analysis for the same disorder

   (b) research purposes (for any disorder), subject to the approval of the Institution's Ethics Committee, provided that any information from such research will remain confidential and anonymous.

4. The results of the analysis carried out on this sample of stored biological material will be made known to me, via my doct

5. I authorise my doctor(s) to provide relevant clinical details in order for the research to be published.

6. I have been informed that:

   (a) **I will not receive any rumeneration for my involvement in this study.**

   (b) **I will not endure additional financial costs due to the tests being performed.**

   (c) The risks and benefits associated with genetic analysis have been explained to me.

   (b) The analysis procedure is specific to the genetic condition mentioned above and cannot determine the complete genetic makeup of an individual.

   (c) The genetics laboratory is under an obligation to respect and maintain medical confidentiality.

   (d) Genetic analysis may not be informative for some families or family members.

   (e) **where biological material is used for research purposes, there may be no direct benefit to me.**

7. I understand that I may withdraw or modify my consent for any aspect of the above at any time without this affecting my future medical care.

8. **ALL OF THE ABOVE HAS BEEN EXPLAINED TO ME IN A LANGUAGE THAT I UNDERSTAND AND MY QUESTIONS ANSWERE BY:** Dr Ines Buccimazza or Me Namitha Chabilal at Inkosi Albert Luthuli Hospital

9. **Address:**

Tel. no.:

Signature _____ Date: _____
of person giving consent Capacity _____

Patient signature: _____ **Witnessed consent:**_____

Appendix G

NATIONAL HEALTH LABORATORY SERVICE

Practice No. 5200296

Office of the Business Manager
UNIVERSITAS ACADEMIC LABORATORIES
PO BOX 339(G3)
C/O: CHEMICAL PATHOLOGY
1st FLOOR
BLOCK C
FACULTY OF HEALTH SCIENCES
UNIVERSITY OF FREE STATEB
BLOEMFONTEIN
9301

REQUEST FOR APPROVAL OF LABORATORY RESOURCES FOR ACADEMIC PURPOSES

Date: 20 April 2016

Requestor: Dr. N van der Merwe

Project Name: **"Screening for PALB2 mutations in South African women with BRCA negative familial breast cancer - MF Maketha."**

Dear Dr. van der Merwe

Your request for use of laboratory facilities / data is hereby granted under following conditions:

1) That University Ethical Committee approval is obtained
2) That approval for use of NHLS patient data is obtained
3) All existing laboratory data remain confidential to the patient and doctor(anonymity is maintained)
4) This Office must be notified before any publication of any results / findings is made.
5) NHLS is recognised in all publications
6) Only existing data may be used. Any billable tests to be done must be funded by research or other funds.
7) Equipment may be used only upon approval of relevant manager and supply of own consumables.

May your project be successful.

.....
Prof Henry Pieters
Business Manager
Universitas Academic Laboratories

Acting Chairperson: Dr Fazel Randera Interim CEO: Ms Joyce Mogale
Physical Address: 1 Modderfontein Road, Sandringham, Johannesburg, South Africa Postal Address: Private Bag X8, Sandringham, 2131, South Africa
Tel: +27 (0) 11 386 6000/ 0860 00 NHLS(6457) www.nhls.ac.za
Practice number: 5200296

Version 2.5 29 June 2017

Rationale for ENIGMA classification criteria

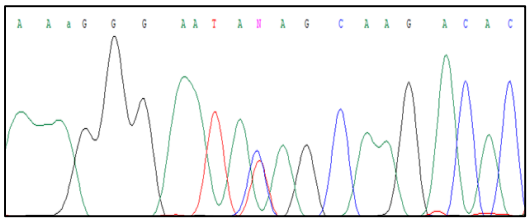
Class	Criterion	Rationale for Criterion	Rationale/Summary of evidence stated for classification in ClinVar/other.
Class 5: pathogenic	Posterior probability of pathogenicity >0.99 from multifactorial likelihood analysis.	IARC recommendation for Class 5 Pathogenic (Plon et al., 2008)	IARC class based on posterior probability from multifactorial likelihood analysis, thresholds for class as per Plon et al. 2008 (PMID: 18951446). Class 5 Pathogenic based on posterior probability = [insert posterior].
	Coding sequence variant encoding a premature termination codon i.e. nonsense/frameshift predicted to disrupt expression of clinically important functional domain(s)/residue(s).	Treated clinically as pathogenic	Variant allele predicted to encode a truncated non-functional protein.
	The variant allele produces only transcripts that lead to a premature stop codon, or in-frame deletion predicted to disrupt clinically important domains, as determined by RNA assays on patient germline tissue that assess allele-specific transcript expression.	Treated clinically as pathogenic	Allele-specific assay on patient-derived mRNA demonstrated that the variant allele produces only predicted non-functional transcripts. Variant allele produces [insert r.#_#del] transcript(s).
	Copy number deletion removing exon(s) spanning clinically important residue(s) ***or*** proven to result in a frameshift alteration predicted to interrupt expression of clinically important residue(s).	Treated clinically as pathogenic	Copy number deletion variant allele predicted to encode a non-functional protein.
	Copy number duplication proven to result in frameshift alteration predicted to interrupt expression of clinically important residue(s).	Treated clinically as pathogenic	Copy number duplication variant allele predicted to encode a non-functional protein.
Class 4: likely pathogenic	Posterior probability of pathogenicity 0.95-0.99 from multifactorial likelihood analysis.	IARC recommendation for Class 4 Likely Pathogenic (Plon et al., 2008)	IARC class based on posterior probability from multifactorial likelihood analysis, thresholds for class as per Plon et al. 2008 (PMID: 18951446). Class 4 Likely Pathogenic based on posterior probability = [insert posterior].
	In the absence of clinical evidence to assign an alternative classification, sequence variant that alters the *BRCA1* or *BRCA2* translation initiation methionine site.	Bioinformatic and laboratory evidence indicates that a variant disrupting translation from the native initiation methionine site will result in lead to transcript(s) that encoding protein with abrogated function. For *BRCA1*, the first in-frame methionine codon (p.M18) falls well within the RING domain, and the resulting N-truncated protein would delete several residues that are important for the BARD1 interaction (Starita et al., 2015). For *BRCA2*, *in vitro* evidence indicates that preferential transcription from several out of frame ATGs located in	*BRCA1*: A variant disrupting the native initiation start site is expected to lead to use of the first downstream in-frame methionine codon (p.M18), which lies well within the RING domain and will result in an N-truncated protein lacking a clinically important functional protein domain (PMID: 25823446). *BRCA2*: A variant disrupting the native initiation start site is expected to lead to transcription from out-of-frame methionine codons, and transcripts predicted to encode

		the mRNA upstream of the first in-frame methionine codon (p.M124) indicates that very little protein synthesis will originate from p.M124 (Parsons et al., 2015).	a truncated non-functional protein (PMID: 24985344).
	Variant at IVS±1 or IVS±2 or G>non-G at last base of exon when adjacent intronic sequence is not GTRRGT that is predicted to alter used of native donor/acceptor site AND is untested for splicing aberrations using RNA assays on patient blood that assess allele-specific transcript expression, AND is not predicted or known to alter production of (naturally occurring) in-frame RNA isoforms that may rescue gene functionality.	Disruption of highly conserved bases at acceptor and donor splice sites is extremely likely to result in a splicing aberration, with suggested prior probability 0.96 (Walker et al., 2013). Conservative classification is warranted since pathogenicity cannot be assumed for all mRNA profiles arising from a variant allele e.g. incomplete effect on splicing, or potential to lead to in-frame transcripts encoding functional protein	Consensus donor/acceptor site variant allele with high likelihood to result in splicing aberration with pathogenic consequences; G>non-G change at last base of the exon with high bioinformatic likelihood to result in splicing aberration with pathogenic consequences.
	A variant that encodes the same amino acid change as a previously established Class 5 pathogenic **missense** variant with a different underlying nucleotide change, is located in a known clinically important functional protein domain, with no evidence of mRNA aberration (splicing or expression) from in vitro mRNA assays on patient RNA, and the variant is absent from outbred control reference groups.	Having excluded possible mRNA defects caused by a nucleotide change, the clinical consequences of a rare missense variant should be equivalent irrespective of the underlying nucleotide change. Absence in controls and location in a functional domain provides additional support for evidence of pathogenicity.	Variant allele encodes same protein change as a proven pathogenic **missense** allele, and does not alter mRNA splicing.
	A small in-frame deletion variant that removes a codon for which a missense substitution Class 5 pathogenic variant has been described, is located in a known clinically important functional protein domain, unlikely to result in an alternative aberration via mRNA splicing, and is absent from outbred control reference groups.	For a given amino acid, the clinical consequences are equivalent for an in-frame deletion and the most severe missense substitution. Absence in controls and location in a functional domain provides additional support for evidence of pathogenicity.	Variant allele deletes an amino acid critical to function and proven to be associated with disease when altered.
Class 3: uncertain	Posterior probability of pathogenicity 0.05-0.949 from multifactorial likelihood analysis.	IARC recommendation for Class 3 Uncertain (Plon et al., 2008)	IARC class based on posterior probability from multifactorial likelihood analysis, thresholds for class as per Plon et al. 2008 (PMID: 18951446). Class 3 Uncertain based on posterior probability = [insert posterior].
	In the absence of clinical evidence to assign an alternative classification, variant allele tested for mRNA aberrations using *in vitro* assays of patient RNA that assess allele-specific transcript expression, and is found to produce mRNA transcript(s) predicted to encode intact full-length protein and/or isoforms that do not disrupt known clinically important functional residue(s).	Variant leads to transcript profile of equivocal clinical significance.	Insufficient evidence to determine clinical significance. Variant allele produces [insert "full-length transcript" AND/OR "in-frame r.#_#del transcript (encoding potentially functional protein)"]
	Insufficient evidence to classify variant.	Does not fit prescribed criteria for other classes	Insufficient evidence to determine clinical significance.
	Variant located at position listed in Table 6, unless proven to fall in another class based on additional evidence.	Variant has potential to lead to in-frame (naturally-occurring) transcripts that may rescue gene functionality.	Variant may result in mRNA transcript(s) that encode functional proteins.
	Variant with conflicting evidence for pathogenicity.	Variant with modest effect on gene/protein function and modest/intermediate effect on risk may demonstrate	Conflicting evidence for pathogenicity; potential intermediate risk variant.

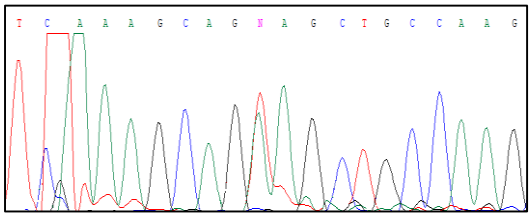
		some but not all features of a high-risk pathogenic variant, and should be highlighted for assessment of risk using alternative approaches.	
Class 2: Likely not pathogenic or of little clinical significance	Posterior probability of pathogenicity 0.001-0.049 from multifactorial likelihood analysis.	IARC recommendation for Class 2 Likely Not Pathogenic (Plon et al., 2008)	IARC class based on posterior probability from multifactorial likelihood analysis, thresholds for class as per Plon et al. 2008 (PMID: 18951446). Class 2 Likely Not Pathogenic based on posterior probability = [insert posterior].
	An exonic variant, that encodes the same amino acid change as a previously established Class 1 not pathogenic missense variant with a different underlying nucleotide change, and for which there is low bioinformatic likelihood to disrupt normal splicing.	With low likelihood of splicing aberration, the clinical consequences of a missense variant should be equivalent irrespective of the underlying nucleotide change.	Variant allele is predicted to not alter splicing and to encode the same protein change as a **missense** allele already proven to be not pathogenic or of little clinical significance.
	Synonymous substitution with low bioinformatic likelihood to disrupt normal splicing, determined to have combined prior probability of pathogenicity ≤0.02 from clinically calibrated bioinformatic analyses.	A silent substitution variant that is not bioinformatically predicted to effect mRNA function is extremely unlikely to result in clinical consequences equivalent to a high-risk pathogenic variant, as indicated by prior probability of ≤0.02 for variants in this stratum from analysis calibrating synonymous changes (Tavtigian, unpublished data, 2008) and bioinformatic predictions of variant effect on splicing against clinical information (Vallee et al., 2016).	Synonymous substitution variant, with low bioinformatic likelihood to result in a splicing aberration (Splicing prior probability 0.02; (http://priors.hci.utah.edu/PRIORS/)
Class 1: not pathogenic or of no clinical significance	Posterior probability of pathogenicity <0.001 from multifactorial likelihood analysis	IARC recommendation for Class 1 Not Pathogenic (Plon et al., 2008)	IARC class based on posterior probability from multifactorial likelihood analysis, thresholds for class as per Plon et al. 2008 (PMID: 18951446). Class 1 Not Pathogenic based on posterior probability = [insert posterior].
	Variant with reported frequency ≥1% in large outbred control reference groups	High-risk variants are not common in the general population, and outbred reference groups exclude the possibility that a selected variant is an undetected founder "mutation"	Not pathogenic based on frequency >1% in an outbred sampleset. Frequency [insert MAF (Population)], derived from [insert dataset (yyyy-mm-dd)].
	Exonic variant that encodes the same amino acid change as a previously established Class 1 Not Pathogenic **missense** variant with a different underlying nucleotide change, and for which there is no evidence of mRNA aberration from in vitro mRNA assays.	Having excluded possible mRNA defects caused by a nucleotide change, the clinical consequences of a missense variant should be equivalent irrespective of the underlying nucleotide change.	Variant allele does not alter splicing and is predicted only to encode the same protein change as a **missense** allele already proven to be not pathogenic or of little clinical significance
	Exonic variant encoding a missense substitution or a small in-frame insertion/deletion with prior probability≤0.02 from clinically calibrated bioinformatic analyses OR intronic variant AND	Multiple points of evidence indicate the variant is unlikely to be associated with high risk: variant is unlikely to affect protein function (from bioinformatic predictions); variant is predicted bioinformatically or	Variant allele has low bioinformatic likelihood to encode a missense alteration affecting protein function (Missense prior probability 0.02;

	Low bioinformatic likelihood to disrupt normal splicing OR Increased bioinformatic likelihood to disrupt normal splicing **but** no associated mRNA aberration (splicing or allelic imbalance) as determined using *in vitro* laboratory assays AND Co-occurrence *in trans* with a known pathogenic sequence variant in the same gene in an individual with no obvious additional clinical phenotype other than BRCA-associated cancer OR Allele frequency ≥0.001 and <0.01 in large outbred control reference groups.	shown from *in vitro* mRNA assays to result in normal splicing (mRNA assays); *in vivo* evidence for proficient function is indicated by co-occurrence *in trans* with a known pathogenic variant in the same gene and with no unusual clinical features, or observation that allele frequency in reference groups is greater than would be expected for a single non-founder variant leading to a dominant disorder	http://priors.hci.utah.edu/PRIORS/), AND low bioinformatic likelihood to alter mRNA splicing (splicing prior 0.xx; http://priors.hci.utah.edu/PRIORS/) OR does not alter splicing *in vitro* [insert reference] AND co-occurrence *in trans* with pathogenic variant [insert pathogenic variant] in patient [insert clinical features, age etc], reported in [insert reference]]. OR Frequency [insert MAF (Population)], derived from [insert dataset (yyyy-mm-dd)].

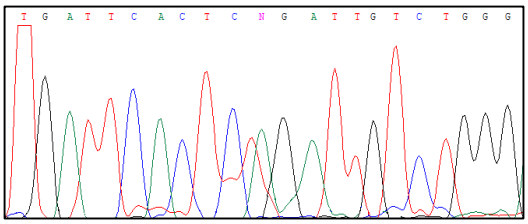
Appendix I



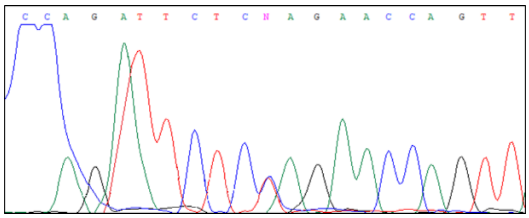
2*: PALB2* c.84C>T, p.Tyr28Tyr (rs761533286)



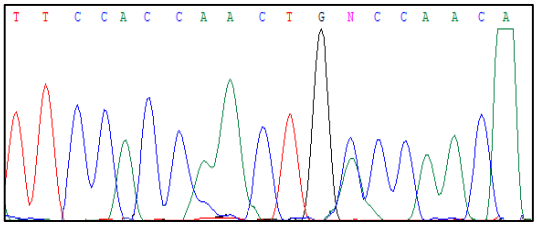
4: *PALB2* c.424A>T, p.Lys142Ter* (rs587782005)



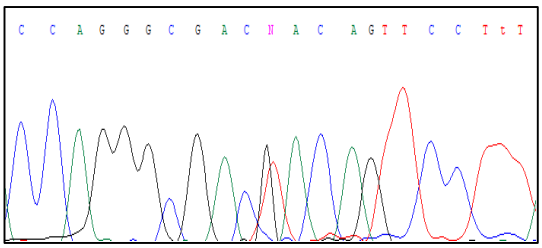
4: *PALB2* c.508A>T, p.Arg170Ter*



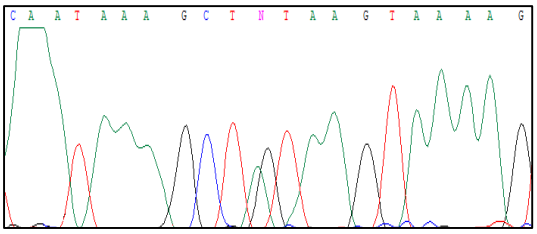
4: *PALB2* c.629C>T, p.Pro210Leu (rs57605939)



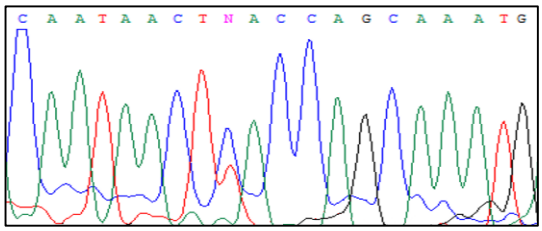
4*: PALB2* c.680C>A, p.Ala227Asp



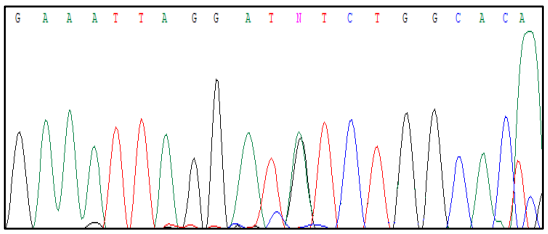
4: *PALB2* c.738T>G, p.Thr246Thr



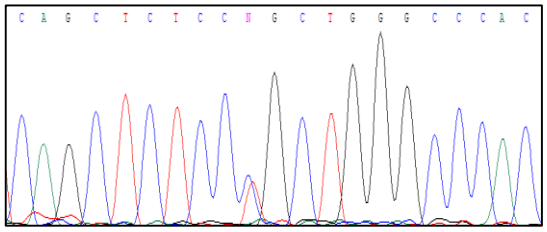
4: *PALB2* c.925A>G, p.Ile309Val (rs3809683)



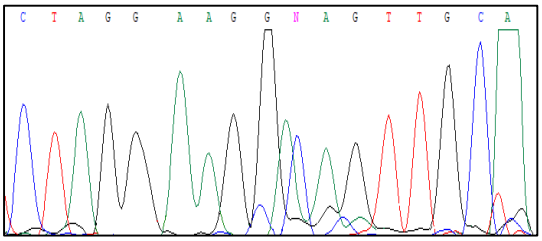
4: *PALB2* c.1010T>C, p.Leu337Ser (rs45494092)



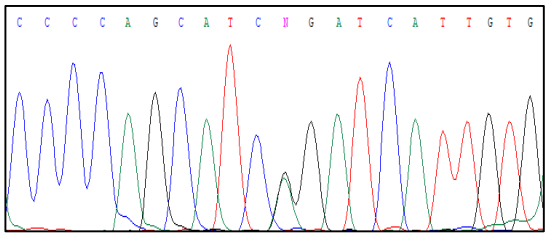
4: *PALB2* c.1395G>A, p.Met465Ile (rs746805049)



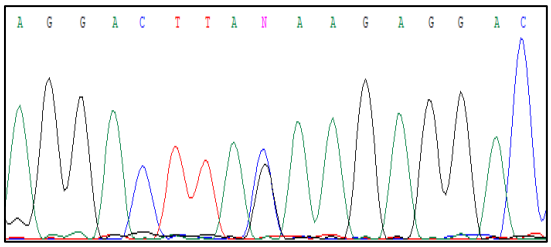
4: *PALB2* c.1470C>T, p.Pro490Pro (rs45612837)



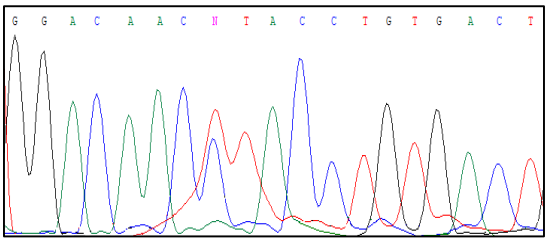
4: *PALB2* c.1508C>A, p.Ala503Glu



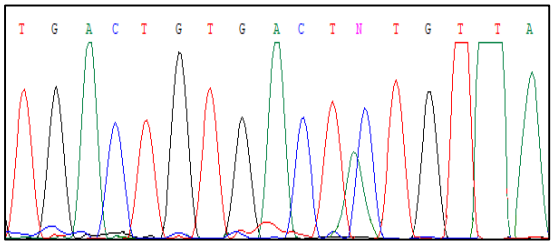
4: *PALB2* c.1572A>G, pSer524Ser (rs45472400)



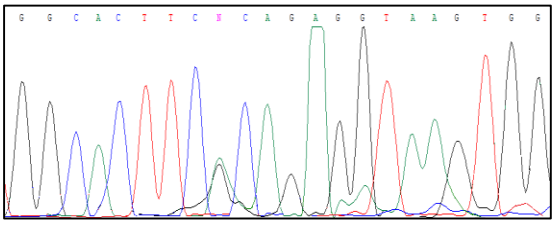
5: *PALB2* c.2014G>C, p.Glu672Gly (rs45532440)



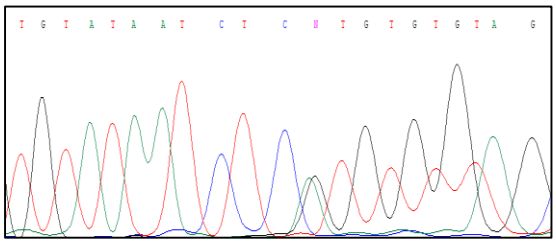
5: *PALB2* c.2393C>T, p.Pro798Leu



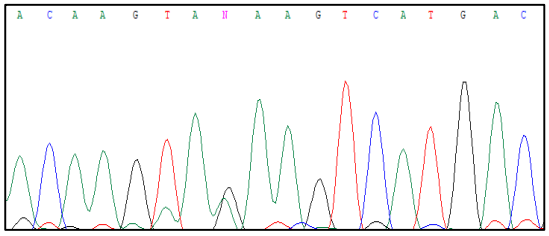
5: *PALB2* c.2411C>A, p.Ser804Tyr (rs149836639)



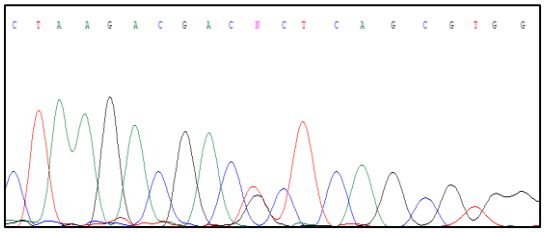
7: *PALB2* c.2743G>A, p.Ala915Thr (rs374736398)



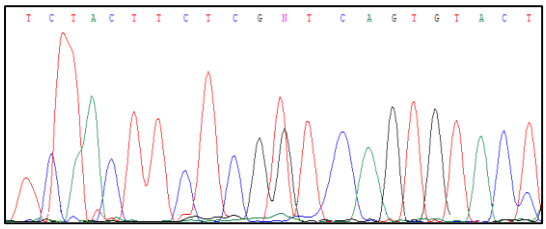
8: *PALB2* c.2794G>A, p.Val932Met (rs45624036)



9: *PALB2* c.2968G>A, p.Glu990Lys



12: *PALB2* c.3300T>G, p.Thr1100Thr (rs45516100)



13: *PALB2* c.3434G>T, p.Gly1145Val